Framework for Risk Assessment of Mixture Designs Incorporating Alkali-Silica Reactive Aggregates

ANA BERGMANN

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN CIVIL ENGINEERING

Department of Civil Engineering
Faculty of Engineering
University of Ottawa

To my family, whose love and support have been my constant guide.

To my husband, Guilherme, for your patience and unwavering belief in me. You are my greatest partner, and this journey would not have been possible without you.

To the little soul I carry in my heart, and to my rainbow baby, whose presence brings light after the storm. This work is a reflection of the strength you inspire in me.

With deepest gratitude and love, this thesis is for all of you.

Abstract

Abstract: Alkali aggregate reaction (AAR) affected structures show reduced serviceability and premature distress in over 50 countries worldwide. Once triggered, remediation is challenging and uncertain; therefore, prevention remains the most effective strategy against AAR. Several laboratory test protocols have been developed over the years to evaluate the potential reactivity of aggregates under varying conditions. Among them, the Accelerated Mortar Bar Test (AMBT) and Concrete Prism Test (CPT) are the most widely used. However, exposure site data has revealed notable discrepancies between laboratory outcomes and field performance. Therefore, this PhD project proposes a probabilistic framework integrating laboratory and field data with logistic regression to predict ASR risk. First, a systematic review and comparative analysis of laboratory methodologies and field behaviour is conducted, supported by a comprehensive bibliometric analysis. In addition, a robust database integrating laboratory and field data has been established to support the research's future steps. Next, the study focuses on the reliability of laboratory tests, indicating moderate accuracy in predicting field performance for the AMBT and the CPT. Then, by incorporating a multifactorial analysis integrating laboratory and field data with statistical and probabilistic modelling to account for variability in the test outcomes, this research assesses the risk of AAR occurrence in the field. More specifically traditional statistical methods, Bayesian analysis, and Beta distribution model are employed resulting in the probability distribution of AAR occurrence given laboratory test outcomes, environment and alkali loading. The findings highlight an opportunity for recalibration of these methods through advanced analytical models that account for environmental conditions, alkali content, and the presence of SCMs to improve predictive accuracy. Therefore, machine learning (ML) techniques, including decision tree classifiers and logistic regression are employed to appraise and adjust the thresholds of AMBT and CPT. Finally, the study proposes that thresholds should be observed dynamically, through a flexible coupled threshold-time (FCTT) approach, taking into account expansion levels, test duration, environmental factors, and alkali content to more accurately predict AAR occurrence. This dynamic risk assessment framework enables more informed decisions regarding mitigation

strategies based on a structure's specific information, improving alignment between laboratory outcomes and real-world durability.

Keywords: Alkali-aggregate reaction (AAR), Alkali-silica reaction (ASR), performance testing, laboratory/field correlation, probabilistic modelling, Bayesian analysis, logistic regression, machine learning, concrete durability, field prediction, Accelerated Mortar Bar Test (AMBT), Concrete Prism Test (CPT), threshold calibration.

Acknowledgements

This thesis would not have been possible without the support, guidance, and encouragement of so many wonderful individuals, to whom I am deeply grateful.

To my husband, Guilherme Leme, your unwavering patience and belief in me have been my greatest source of strength. You stood by me through every challenge, offering constant encouragement and understanding. I cannot express how much your presence and support have meant to me during this journey. This accomplishment is as much yours as it is mine.

To my parents, Clarice A.M. Bergmann and Luiz C. Bergmann, for your unconditional love and for always believing in my potential. Your guidance and support throughout my life have shaped who I am today. To my sister Izabella Bergmann, my brother Luiz E.M. Bergmann, and my brother-in-law Renan Zonta, thank you for your constant love and encouragement and for being my pillars of strength. Your belief in me has kept me motivated, especially during the toughest times.

I would like to extend my deepest gratitude to my supervisor, Dr. Leandro Sanchez, whose mentorship and guidance have been instrumental in the completion of this thesis. Your expertise and support throughout this process have been invaluable. You pushed me to challenge myself and think critically, I am truly fortunate to have had you as my mentor. I am especially grateful for the innumerous conferences and travels we had the chance to participate in together, which broadened my horizons and provided invaluable learning experiences.

To the members of my committee, Dr. Jason Ideker, Dr. Martin Noël, Dr. Hamzeh Hajiloo, Dr. Di Wang, and Dr. Elena Dragomirescu, thank you for kindly offering your time to improve this research. A special acknowledgment to the laboratory technical officers, Dr. Gamal Elnabelsya and Dr. Muslim Majeed. Thank you for always being available to answer my questions and support. Additional thanks to Dr. Fabio Biondini and his research team at Politecnico di Milano for the support during my internship. I would also like to sincerely thank my former supervisor, Dr. Gustavo Savaris, for being present throughout the process and offering valuable support.

To my dear friends, who have been my sounding boards, my confidantes, and my cheerleaders throughout this process. Your friendship, laughter, and support have carried me through the most challenging moments, and I am forever grateful for your presence in my life. Thank you also for understanding the times I was away, focused on this journey, it means a lot to have friends as patient and supportive as you.

To my colleagues at uOttawa and the uStructure team, especially Hian de Freitas, Sérgio Dantas, Olusola Olajide, Mayra de Grazia, and Cassandra Trottier, working alongside you has been an incredible experience. Your collaboration, shared knowledge, and friendship made this journey so much more rewarding. I am grateful for the conversations, the brainstorming sessions, and the camaraderie that we shared.

Finally, my heartfelt thanks to the RILEM Technical Committee 301 members, where I had the privilege of forming lasting connections with brilliant minds in the field. The opportunities I have had through this committee to engage with international colleagues have not only enriched my professional development but have also expanded my network in meaningful ways.

This thesis is a culmination of all the support and encouragement I've received from so many, and I am deeply grateful for each and every one of you.

Table of Contents

List of Tables

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

List of Abbreviations and Symbols

Term	Definition
α'	Parameter for the Beta distribution representing reactive cases
AAR	Alkali-Aggregate Reaction
ACPT	Accelerated Concrete Prism Test
ACR	Alkali-Carbonate Reaction
AMBT	Accelerated Mortar Bar Test
ASR	Alkali-Silica Reaction
ASTM	American Society for Testing and Materials
β_0	Intercept in logistic regression
β_ALK	Coefficient for alkali content
β_EXP	Coefficient for expansion
β_RH	Coefficient for relative humidity
β_TEMP	Coefficient for temperature
β'	Parameter for the Beta distribution representing non-reactive cases
CCT	Concrete Cylinder Test
CDF	Cumulative Distribution Function
CI	Confidence Interval
CPT	Concrete Prism Test
CSA	Canadian Standards Association
EXP	Expansion
F1 score	Harmonic mean of precision and recall
FN	False Negative
FP	False Positive
GCM	German Concrete Method
K coefficient	Ratio of long-term field performance expansion to accelerated laboratory test expansion
Logit(P)	Logit function
MCPT	Miniature Concrete Prism Test
MLE	Maximum Likelihood Estimation
Mol	Mole (used for concentration, e.g., 1 Mol NaOH solution)
Na_2O_{eq}	Sodium Oxide Equivalent
NaOH	Sodium Hydroxide
NCPT	Norwegian Concrete Prism Test
NR	Non-reactive
P(L_i | R)	Likelihood function, probability of observing data given reactivity
P(L_i)	Marginal likelihood, overall probability of observing a specific condition
P(R | L_i)	Posterior probability, the probability of reactivity given the observed condition

P(R)	Prior probability of reactivity
P(x)	Probability of occurrence
PC	Portland Cement
PDF	Probability Density Function
PPF	Percentile Point Function (Quantile function)
R	Reactive
R_at	Accelerated laboratory test expansion
R_lt	Long-term field performance expansion
R0, R1, R2, R3	Reactivity levels in ASTM C1778 (R0 = non-reactive, R1 = moderately reactive, R2 = highly reactive, R3 = very highly reactive)
SCMs	Supplementary Cementitious Materials
TN	True Negative
TP	True Positive
w/c	Water to Cement Ratio
α	Parameter for the Beta distribution representing reactive cases

Foreword

This Ph.D. thesis develops a probabilistic framework that integrates laboratory and field data with logistic regression to enhance the accuracy of predicting alkali-aggregate reaction (AAR) risk. The key objective is to address the discrepancies often observed between laboratory test results—such as the Accelerated Mortar Bar Test (AMBT) and Concrete Prism Test (CPT)—and the actual field performance of structures affected by AAR. By incorporating factors such as expansion levels, environmental conditions, and alkali content, the framework allows for dynamic risk assessments, thereby offering a more dependable prediction model. This can greatly assist in developing effective mitigation strategies for AAR-affected structures.

Organized into nine chapters, the thesis includes four scientific papers focused on key aspects of AAR testing and risk prediction. Chapter One introduces the global significance of AAR and presents the research structure. Chapter Two presents a detailed literature review, including accelerated laboratory testing methods, field exposure history, and the concepts of statistical and machine learning approaches in assessing AAR risks. Chapter Three outlines the research methodology used to meet the project's goals, while also summarizing the scope of the included scientific papers.

Chapters Four through Seven present the four scientific papers. Chapter Four (Paper 1) conducts a comprehensive review and comparative analysis of laboratory methodologies versus field behavior, establishing a database that integrates both sets of data as a foundation for the research. Chapter Five (Paper 2) assesses the reliability of AMBT and CPT, exploring their capacity to predict real-world field performance. In Chapter Six (Paper 3), a multifactorial analysis employs statistical and probabilistic modelling techniques, such as Bayesian analysis and Beta distribution models, to assess AAR risk. The model assesses laboratory results, environmental conditions, and alkali loading to calculate the probability of AAR occurrence in the field. Chapter Seven (Paper 4) explores machine learning techniques, using decision trees and logistic regression to recalibrate AMBT and CPT thresholds dynamically, accounting for factors like laboratory expansion, duration, environmental conditions, and alkali content. The dynamic risk assessment framework, discussed in Chapter Eight (Conclusion), builds on insights from these papers to propose

more accurate AAR prediction models, bridging the gap between laboratory outcomes and field performance. Finally, Chapter Nine outlines future work and recommendations related to this research.

Chapter 1: Introduction

1.1 Synopsis

Alkali-aggregate reaction (AAR), a major durability concern, affects concrete structures globally, with significant implications for the integrity and serviceability of infrastructures in over 50 countries [1–7]. AAR encompasses two main types: alkali-silica reaction (ASR) and alkali-carbonate reaction (ACR), both triggered when dissolved alkali hydroxides present in the concrete pore solution react with unstable mineral phases within the aggregates [8,9]. ASR, a more predominantly occurrence of AAR, generates a secondary product (ASR-gel) that swells upon water uptake, leading to internal stresses and cracking, reducing the mechanical properties of affected concrete [9,10].

Prevention is widely known as the most effective strategy against AAR, given the challenges associated with remediation once the reaction is triggered. To assess aggregate reactivity and the effectiveness of preventive measures, a variety of laboratory test methods have been developed, with the accelerated mortar bar test (AMBT) and concrete prism test (CPT) being among the most widely used [8,11]. However, the discrepancies between laboratory results and actual field performance have raised concerns regarding the reliability of these tests in accurately predicting long-term AAR risks [12–18]. Laboratory tests tend to accelerate the reaction process, which can prevent them from fully capturing critical real-world variables, such as environmental conditions; hence, field-exposed elements are the most reliable source for selecting a ground truth reference for comparison regarding ASR development [19]. Moreover, enhanced data availability from outdoor exposure sites underscores these differences, emphasizing the need for a thorough analysis of the reliability of the testing methodologies [14,15,20,21].

To address these limitations, this study undertakes a systematic bibliometric analysis to map trends, contributors, and collaborations in AAR research, consolidating laboratory and field data into a comprehensive database. This database enables an in-depth evaluation of discrepancies between laboratory tests (i.e., AMBT and CPT), and field performance, incorporating advanced data analysis methods and probabilistic modelling. By employing logistic regression, Bayesian analysis, and machine learning approaches, the study develops a flexible, probabilistic framework that dynamically accounts for environmental factors (i.e., temperature and relative humidity), alkali content, and laboratory expansion. This framework aims to refine AAR prediction models,

establishing a flexible coupled threshold-time (FCTT) approach that integrates laboratory and field data, allowing for dynamic risk assessments and informed decisions on AAR mitigation strategies. Finally, this research contributes a structured and data-driven methodology to assess and mitigate AAR risk in concrete structures, allowing for more informed decisions on aggregate selection, mixture design, and preventive strategies. This approach enhances the predictive capacity of laboratory methods, aligning them more closely with real-world durability insights, and supports the development of more effective guidelines for managing the long-term durability of critical infrastructure affected by AAR.

1.2 Research objectives

The main objectives of this project are:

- Conduct a comprehensive bibliometric analysis to identify key contributors, collaborations, and trends in AAR research, creating a robust database that integrates laboratory and field performance data. This will support a systematic examination of the relationship between laboratory testing (such as AMBT and CPT) and field performance in concrete structures, allowing for a thorough evaluation of the reliability of these test methods in predicting AAR risks (Phase 1 - Papers 1 and 2).
- Apply multifactorial statistical analysis, probabilistic methods, and machine learning techniques (i.e., decision trees and logistic regression) to address uncertainties in laboratory test outcomes. By enhancing predictive models, this objective aims to dynamically link laboratory and field data, accounting for the complex interactions between environmental factors and concrete composition (Phase 2 - Papers 3 and 4).
- Present the flexible coupled threshold-time approach as a probabilistic framework that integrates laboratory and field data. By incorporating environmental factors, alkali content, structural classification, and AAR expansion thresholds, this framework aims to provide a risk-based approach for preventing AAR-induced damage in varied exposure conditions of concrete structures (Phase 3 - Paper 5).

1.3 Core of the PhD Thesis – Scientific Papers

The research is organized into three distinct phases, each comprising related topics, as illustrated in Figure 1. 1.

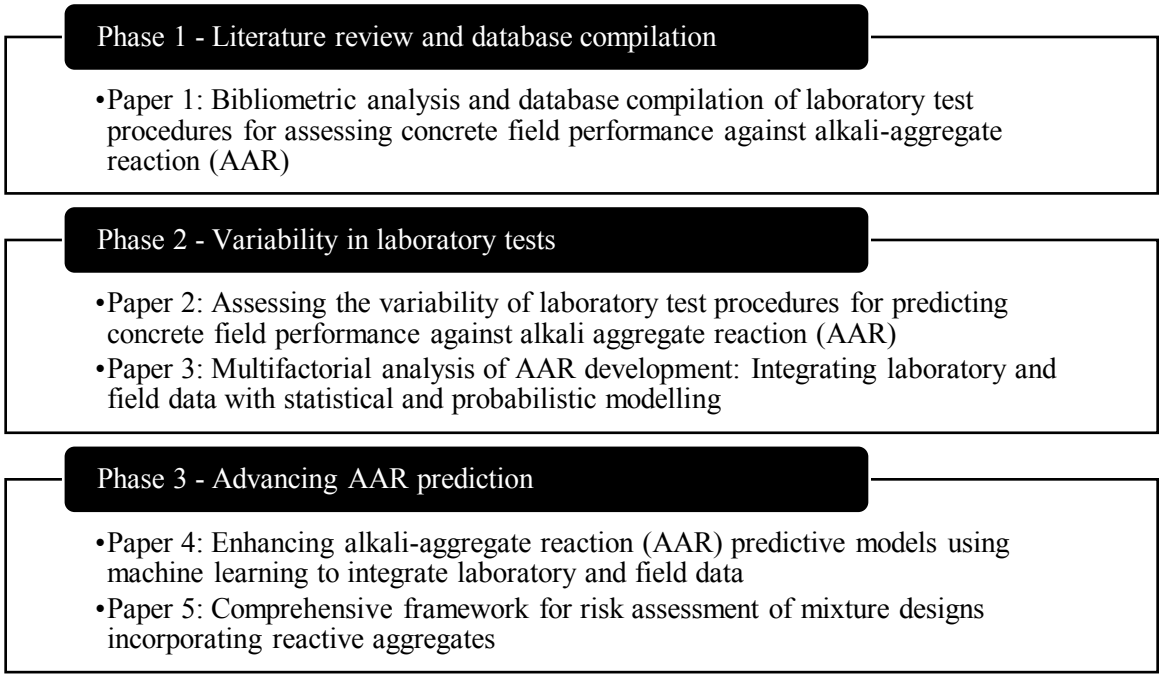


Figure 1. 1– Structure of the PhD project.

Phase 1 addresses Paper 1, which involves a comprehensive bibliometric analysis and the creation of a robust database that integrates laboratory test results with field performance data. This phase aims to systematically review the topic, mapping out key contributors, collaborations, and trends in AAR testing and field studies. By compiling data from both laboratory and field studies, this phase establishes a foundation for future research steps.

In phase 2, Paper 2 evaluates the variability and discrepancies between laboratory testing methods, such as AMBT and CPT, and assesses their accuracy in predicting field performance. This phase critically examines the limitations of these testing methods and seeks to quantify the differences between laboratory results and field behavior. Additionally, Paper 3 employes probabilistic modelling to address uncertainties in laboratory test outcomes. By quantifying these uncertainties, the study aims to enhance the reliability of predictive models, ultimately providing a clearer link between laboratory and field performance.

Phase 3 focuses on Papers 4 and 5, where advanced analytical techniques are employed to refine the predictive capabilities of AAR testing methods. In Paper 4, machine learning techniques,

including decision trees and logistic regression models, propose a flexible coupled threshold-time (FCTT) approach that evaluates thresholds probabilistically and dynamically, integrating expansion levels, test duration, environmental factors, and alkali content to improve AAR occurrence prediction. As a summary and conclusion of the findings, Paper 5 presents a probabilistic framework integrating laboratory and field data with logistic regression to predict ASR risk. The framework incorporates environmental conditions, alkali content, and structural classification to define the risk of AAR field occurrence.

In conclusion, this research provides a structured and dynamic approach to improving the prediction of AAR in concrete structures. By integrating advanced data analysis techniques with a deep understanding of laboratory and field discrepancies, this study contributes to informed decisions on mitigation strategies based on a structure's specific information, improving alignment between laboratory outcomes and real-world durability.

1.4 References

[1] M.D.A. Thomas, K.J. Folliard, J.H. Ideker, North America (USA and Canada), in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[2] J. Lindgård, B. Grelk, B.J. Wigum, J. Trägårdh, K. Appelqvist, E. Hol, M. Ferreira, M. Leivo, Nordic Europe, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[3] E.M.R. Fairbairn, South and Central America, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[4] K. Yamada, T. Miyagawa, Japan, China and South-East Asia, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[5] A. Shayan, S. Freitag, Australia and New Zealand, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[6] A.K. Mullick, Indian Sub-Continent, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[7] T. Kay, A.B. Poole, I. Sims, Middle East & North Africa, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[8] B. Fournier, M. Berube, Alkali-aggregate reaction in concrete: a review of basic concepts and engineering implications, CANADIAN JOURNAL OF CIVIL ENGINEERING 27 (2000) 167–191. https://doi.org/10.1139/l99-072.

[9] P.J. Nixon, I. Sims, eds., RILEM Recommendations for the Prevention of Damage by Alkali-Aggregate Reactions in New Concrete Structures, Springer Netherlands, Dordrecht, 2016. https://doi.org/10.1007/978-94-017-7252-5.

[10] B. Fournier, M.-A. Bérubé, Alkali–aggregate reaction in concrete: a review of basic concepts and engineering implications, 27 (2000).

[11] I. Sims, P. Nixon, RILEM Recommended Test Method AAR-0: Detection of Alkali-Reactivity Potential in Concrete—Outline guide to the use of RILEM methods in assessments of aggregates for potential alkali-reactivity, Mat. Struct. 36 (2003) 472–479. https://doi.org/10.1007/BF02481527.

[12] M. Thomas, A. Dunster, P. Nixon, B. Blackwell, Effect of fly ash on the expansion of concrete due to alkali-silica reaction – Exposure site studies, Cement and Concrete Composites 33 (2011) 359–367. https://doi.org/10.1016/j.cemconcomp.2010.11.006.

[13] B. Fournier, A. Bilodeau, N. Bouzoubaa, P.-C. Nkinamubanzi, Field and Laboratory Investigations on the Use of Fly Ash and Li-Based Admixtures to Prevent ASR in Concrete, in: United Kingdom, 2018.

[14] I. Borchers, J. Lindgård, C. Müller, Evaluation of laboratory test methods for assessing the alkali-reactivity potential of aggregates by field site tests, Materconstrucc 72 (2022) e286. https://doi.org/10.3989/mc.2022.17221.

[15] J. Custodio, D. Costa, A.B. Ribeiro, A.S. Silva, Assessment of potential alkali-silica reactivity of aggregates for concrete, 4TH INTERNATIONAL CONFERENCE ON STRUCTURAL INTEGRITY (ICSI 2021) 37 (2022) 590–597. https://doi.org/10.1016/j.prostr.2022.01.127.

[16] J. Lindgård, P.J. Nixon, I. Borchers, B. Schouenborg, B.J. Wigum, M. Haugen, U. Åkesson, The EU "PARTNER" Project — European standard tests to prevent alkali reactions in aggregates: Final results and recommendations, Cement and Concrete Research 40 (2010) 611–635. https://doi.org/10.1016/j.cemconres.2009.09.004.

[17] T. Drimalas, K.J. Folliard, J.H. Ideker, Findings from the University of Texas at Austin ASR Exposure Site after 20 Years, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th

International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 516–523. https://doi.org/10.1007/978-3-031-59349-9_59.

[18] S. Hayman, M. Thomas, N. Beaman, P. Gilks, Selection of an effective ASR-prevention strategy for use with a highly reactive aggregate for the reconstruction of concrete structures at Mactaquac generating station, Cement and Concrete Research 40 (2010) 605–610. https://doi.org/10.1016/j.cemconres.2009.08.015.

[19] B. Fournier, J.H. Ideker, K.J. Folliard, M.D.A. Thomas, P.-C. Nkinamubanzi, R. Chevrier, Effect of environmental conditions on expansion in concrete due to alkali–silica reaction (ASR), Materials Characterization 60 (2009) 669–679. https://doi.org/10.1016/j.matchar.2008.12.018.

[20] B. Fournier, R. Chevrier, A. Bilodeau, P.-C. Nkinamubanzi, N. Bouzoubaa, COMPARATIVE FIELD AND LABORATORY INVESTIGATIONS ON THE USE OF SUPPLEMENTARY CEMENTING MATERIALS (SCMs) TO CONTROL ALKALI-SILICA REACTION (ASR) IN CONCRETE, in: Brazil, 2016.

[21] C.A. MacDonald, Kingston Outdoor Exposure Site for ASR –29 Year Update, (2020).

Chapter 2: Literature Review

2.1 Alkali aggregate reaction

Alkali-aggregate reaction (AAR) is a chemical reaction between the alkalis present in the pore solution of concrete and reactive minerals within the aggregates [1]. First identified in the 1920s and thoroughly researched by Stanton in the 1940s [2], AAR is now recognized as a global phenomenon, affecting concrete structures across various environmental and operational conditions [3–13]. Today, AAR is considered one of the most detrimental mechanisms that can compromise the long-term durability and structural integrity of concrete infrastructure. The progression of AAR leads to the development of microcracks within the concrete matrix, significantly reducing its mechanical properties and overall durability [14].

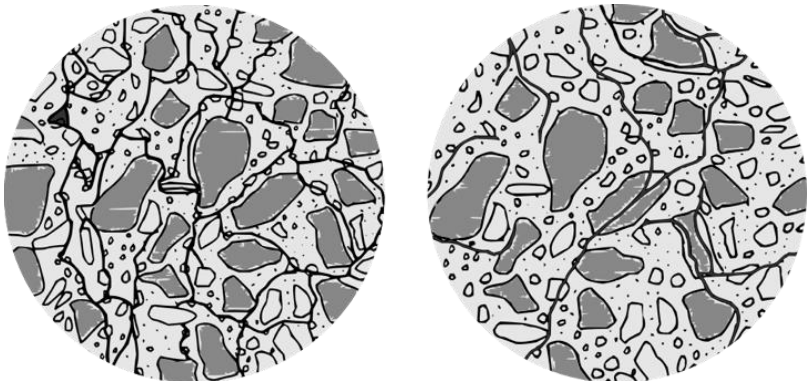
The AAR kinetics are influenced by several factors, including the type of reactive minerals within the aggregates, the concentration of alkali hydroxides in the concrete's pore solution, and environmental factors such as temperature, moisture, and humidity [15–19]. While AAR can be categorized broadly into two distinct types based on the nature of the reactive mineral—alkali-silica reaction (ASR) and alkali-carbonate reaction (ACR)—ASR is by far the most prevalent and well-documented form.

(a) Alkali-silica reaction (ASR)

ASR primarily occurs in concrete containing aggregates with reactive siliceous materials, such as poorly crystallized silica minerals, volcanic glass, or quartz-bearing rocks [20–22]. The ASR process begins when reactive silica (i.e, SiO_2) interacts with the alkaline pore solution in concrete. This pore solution typically contains sodium (i.e, Na^+) and potassium (i.e, K^+) ions, which originate from the cement's alkali content. The high pH of the pore solution, often around 13-14 due to the presence of hydroxide ions (i.e, OH^-), facilitates the dissolution of the reactive silica [23].

Once the silica is dissolved, it reacts with the alkali hydroxides (i.e., NaOH and KOH) present in the pore solution. This reaction leads to the formation of a secondary product, often referred to as ASR gel, a hygroscopic material that absorbs water and expands [20]. As the gel swells, it expands generating internal pressures within the concrete, particularly in and around the aggregate particles where the gel forms. This swelling exerts tensile stresses on the concrete, leading to the formation of cracks [22]. As the internal pressure increases, the cracks propagate from the aggregate particles

into the surrounding cement paste. This cracking pattern can lead to significant structural damage, as it compromises the mechanical properties of the concrete, including its compressive strength, tensile strength, and modulus of elasticity [14]. Figure 2. 1 illustrates the typical cracking patterns associated with ASR, highlighting scenarios involving both reactive fine and coarse aggregates.



(a) (b)

Figure 2. 1 – Alkali-aggregate reaction cracking pattern: (a) ASR from the reactive sand (fine aggregate); (b) ASR from the reactive coarse aggregate. Adapted from [14].

ASR is associated with specific silica minerals, including opal, tridymite, cristobalite, volcanic glass, chert, cryptocrystalline (or microcrystalline) quartz, and strained quartz. These minerals can be found in various rock types such as shale, sandstone, silicified carbonate rocks, chert, flint, quartzite, quartz-arenite, gneiss, argillite, granite, greywacke, siltstone, arenite, arkose, and hornfels [2].

The extent of mechanical property loss due to ASR is closely related to the degree of concrete expansion, as shown in Table 2. 1 [14]. With minor expansions (0.0%-0.05%), noticeable reductions in modulus of elasticity and tensile strength can range up to 30% and 30-70%, respectively, while compressive strength experiences a minimal reduction of about 5%. Moderate expansion (0.12%) exacerbates cracking within the aggregates and extends into the cement paste, with the rate of decrease in tensile strength and modulus of elasticity slowing down, and compressive strength diminishing by about 10%. Higher expansions (0.20%) extend the cracking further into the cement paste, leading to a 25% reduction in compressive strength. At an expansion rate of 0.30%, a network of interconnected cracks forms, potentially resulting in a 40% loss in compressive strength.

Table 2. 1 – Mechanical property losses (damage results) as per ASR damage degree [14].

Distress Mechanism	Classification of ASR Damage Degree (%)	Reference Expansion Level (%)	Damage Results		
			Stiffness Loss (%)	Compressive Strength Loss (%)	Tensile Strength Loss (%)
ASR	Negligible	0.00–0.03	–	–	–
	Marginal	0.04 ± 0.01	5–37	(−)10–15	15–60
	Moderate	0.11 ± 0.01	20–50	0–20	40–65
	High	0.20 ± 0.01	35–60	13–25	45–80
	Very high	0.30 to 0.50 ± 0.01	40–67	20–35	
	Ultra-high	0.50 to 1.00 ± 0.01	–	–	–
		≥1.00 ± 0.01	–	–	–

(b) Alkali-carbonate reaction (ACR)

ACR is a distinct form of AAR that occurs between alkali hydroxides in the concrete's pore solution and certain types of carbonate rocks, such as argillaceous dolomitic limestone [2,24]. Unlike ASR, which involves the reaction with silica, ACR specifically targets carbonate minerals, leading to a different set of chemical and physical changes in the concrete.

This reaction primarily involves the dedolomitization process, where dolomite ($CaMg(CO_3)_2$) breaks down into calcite ($CaCO_3$) and brucite ($Mg(OH)_2$) [20]. The formation of these new minerals can alter the volume and texture of the aggregate, potentially leading to expansion. The expansion in ACR may not solely result from the dedolomitization reaction but may also involve other mechanisms. For example, mineral restructuring due to the newly formed minerals, including crystal lattices and the precipitation of expansive reaction products, and, in some cases, ACR can occur alongside ASR, especially in aggregates containing both reactive silica and carbonate minerals [24].

The microscopic characteristics of ACR-induced damage are distinct from those of ASR. ACR-related cracks are often observed near the cement paste and adjacent to the aggregate particles [25,26]. These cracks typically show minimal amounts of reaction products, unlike ASR, where the presence of gel is more common. The cracking pattern in ACR-affected concrete tends to be more localized around the reactive carbonate aggregates and is generally less pervasive than that seen in ASR-affected concrete.

ACR-related damage can become evident relatively quickly, sometimes within five years of construction [20]. This rapid onset is due to the relatively fast kinetics of the dedolomitization reaction under favorable environmental conditions (e.g., high alkali content, sufficient moisture, and suitable temperature). However, the extent of damage may vary depending on factors such as

the composition of the aggregate, the specific environmental conditions, and the presence of other reactive phases in the concrete.

2.2 Laboratory test methods for assessing aggregate reactivity potential

The evaluation of the aggregates' reactivity in concrete has been the subject of extensive research and development due to the critical need to prevent AAR-related deterioration in concrete. Over the decades, various laboratory test methods have been developed to assess the reactivity potential of aggregates, each with its unique advantages and limitations, reflecting the evolving understanding of AAR-induced deterioration mechanisms and mitigation strategies.

2.2.1 Petrographic analysis

Petrographic examination, as per ASTM C295 [27], is a detailed microscopic assessment that has been traditionally used as the initial step to evaluate the mineralogical composition of aggregates. The petrographic examination process serves multiple critical purposes: to determine the physical and chemical characteristics of the aggregate that may affect its performance in concrete, describe and classify the constituents of the sample, and evaluate the relative amounts of these constituents, especially when their properties significantly impact material performance [32]. This process also allows for the comparison of new aggregate sources with established ones, using performance records as a benchmark. Therefore, this method appraises the presence of unstable mineral phases in the aggregates, such as opal, chalcedony, and strained quartz, which are known to contribute to ASR in alkaline environments [19,28–30].

Historically, petrographic analysis has been essential to understanding the mineralogical composition of aggregates used in concrete, serving as the first step in assessing new aggregates against ASR [31]. The procedure for conducting a petrographic examination as per ASTM C295 [27] is meticulous and involves several key steps. Initially, a representative sample of the aggregate is collected and prepared by embedding it in a suitable media, such as epoxy, and then slicing it into thin sections. These sections are polished to ensure the surface is smooth and free of artifacts that might interfere with microscopic observation. The polished sections are then examined under a polarizing microscope, which allows the petrographer to identify the different minerals present based on their optical properties, such as birefringence, pleochroism, and refractive indices [32].

Regarding ASR, the method emphasizes the identification of reactive forms of silica, which can vary significantly in reactivity depending on their crystallinity and morphology. The petrographic examination is not only concerned with identifying these minerals but also with assessing their abundance, distribution, and the degree of strain or deformation they might exhibit, as these factors influence their reactivity [33]. Furthermore, the examination aims to classify aggregates regarding their content of potentially reactive forms of silica into three distinct classes: Class I—Very unlikely to be alkali-reactive; Class II—Alkali-reactivity uncertain; and, Class III—Very likely to be alkali-reactive.

Despite its importance, petrographic analysis has certain limitations. One of the primary criticisms is that it is a qualitative rather than a quantitative method. While it can identify the presence of reactive minerals, it does not provide a measure of the potential reactivity of the aggregate in concrete. Furthermore, petrographic analysis alone is insufficient for predicting AAR in the field since it is not a performance test and thus does not account for the dynamic interactions between aggregates and the cement paste under varying environmental conditions [20,21,34]. Therefore, while ASTM C295 is invaluable as a first step in aggregate evaluation, it is often necessary to complement it with performance-based tests, such as the Accelerated Mortar Bar Test or the Concrete Prism Test, to obtain a comprehensive assessment of the aggregate's suitability for use in concrete [32,33]. According to the guidelines provided by RILEM AAR-0 [33], if it indicates the presence of potentially reactive silica (Class II or III), further testing should be undertaken. These tests typically involve laboratory expansion tests that are critical for confirming the reactivity potential of the aggregates considering the interaction with the cement past under environmental conditions.

In conclusion, petrographic examination remains a reference for aggregate evaluation in the construction industry, providing essential information about the mineralogical characteristics of aggregates. However, its limitations mean that it should be used in conjunction with other methods to ensure a thorough understanding of the aggregate's potential behavior in concrete.

2.2.2 Chemical method

The chemical method, standardized by ASTM C289 [35], was designed to assess the potential reactivity of aggregates in combination with alkalis from Portland cement. This method involves a chemical determination of the potential for ASR by measuring the dissolution of silica and the

reduction in alkalinity of sodium hydroxide (NaOH) solution when reacted with a prepared aggregate sample.

In the procedure, an aggregate sample is first crushed and sieved to obtain particles between 150 μm and 300 μm in size, a critical step as it maximizes the surface area for the chemical reaction. The prepared aggregate is mixed with 1M NaOH in reaction containers resistant to corrosion, sealed and placed in a constant-temperature bath maintained at 80°C for 24 hours. After the reaction period, the solution is filtered to measure the amount of dissolved silica, which provides an indication of the aggregate's reactivity. This can be measured using either a gravimetric method or a photometric method. Additionally, the reduction in the alkalinity of the NaOH solution, resulting from the reaction with silica, is assessed. These two measurements (i.e., dissolved silica concentration and alkalinity reduction) are then used to evaluate the aggregate's potential reactivity. The results are interpreted using a classification plot provided in ASTM C289-07 [35], which categorizes the aggregate as innocuous, potentially deleterious, or deleterious.

While this method offers a quick assessment of potential reactivity, it is not fully reliable for all aggregate types, particularly those containing slowly reactive forms of silica, such as strained quartz, or aggregates containing carbonates. The test may not accurately predict the long-term behavior of aggregates in concrete, especially in cases of slow reactivity. Due to these limitations, ASTM C289 [35] has been removed from the current standards and is no longer used as a standalone method for determining the suitability of aggregates for concrete.

2.2.3 Mortar Bar Test (MBT)

In the 1940s, the Mortar Bar Test (MBT) was introduced as the first standardized method, as per ASTM C227 [36], for assessing aggregate reactivity [2]. This test method was used to determine the susceptibility of cement-aggregate combinations to expansive reactions involving hydroxyl ions (OH^-) associated with alkalis, such as sodium and potassium. Specifically, it measures the variation in the length of mortar bars containing the cement-aggregate combination during storage under prescribed test conditions. The procedure involves immersing the mortar bars in a highly alkaline solution, typically 1M NaOH, and maintaining the temperature at 38°C to accelerate the reaction [21]. Expansions exceeding 0.05% at three months or 0.10% at six months are considered indicative of potential reactivity in the cement-aggregate combination, suggesting that the aggregate could be potentially reactive when used in concrete mixtures [37].

However, despite its widespread initial adoption, the MBT faced considerable criticism and challenges over time. One of the primary issues was the excessive leaching of alkalis from the mortar bars, which often led to unreliable results. Additionally, the aggregate grading used in the test was often unrepresentative of the aggregates typically used in concrete mixtures, further compromising the test's accuracy [16,38]. Consequently, the MBT's limitations prompted the development of alternative methods that could provide quicker and more reliable results [18].

2.2.4 Accelerated Mortar Bar Test (AMBT)

The development of the Accelerated Mortar Bar Test (AMBT) in the 1980s, standardized as ASTM C1260 [39], provided faster results by immersing mortar bars in a 1M NaOH solution at 80°C [18]. These conditions were chosen to induce maximum expansion in reactive aggregates within a 12-day timeframe, covering the induction, main expansion and posterior expansion reaction phases [18]. To establish the test thresholds, a linear correlation between the AMBT and MBT outcomes was performed, indicating that although the correlation was moderate (r=0.67), it was determined that an expansion of 0.05% at 3 months in MBT corresponded to 0.11% expansion at 12 days in AMBT. Over time, studies were developed to evaluate these thresholds aiming to better reflect field conditions [19,28,33,40], and currently, they are established as per ASTM C1778 [41]: expansions below 0.1% at 14 days indicate non-reactive (i.e., innocuous) behaviour, 0.1% to 0.3% indicate moderately reactive, 0.3% to 0.45% indicate highly reactive, and expansions above 0.45% are classified as very highly reactive aggregates. Despite its progress, the AMBT has faced criticism for using mortar instead of concrete, which may not accurately represent field conditions, and for potentially yielding false positives due to the harsh conditions, especially the high temperature [42,43].

Various factors intrinsic to the test can influence the outcome, including alkali content and cement fineness, aggregate grading and crushing, mortar bar characteristics (i.e., w/c ratio), and storage conditions (i.e., relative humidity, curing, and leaching) [40,44]. For instance, specific particle sizes, such as finer aggregates, can lead to higher expansion [19,45]. Regarding the cement, its particles' fineness increases the surface area available for reaction leading to greater expansion, while its alkali content, particularly higher concentrations of sodium and potassium ions, increases the pore solution alkalinity collaborating with the expansion potential [46]. Mortar bar characteristics used in AMBT, including mix proportions and water-to-cement ratio, impact the

results. Higher w/c ratios, which result in more porous mortars, can either facilitate the ingress of alkali ions, increasing the potential for expansion or allow for the leaching of alkalis, thereby reducing the potential for expansion [47]. Additionally, storage conditions affect the test results, with higher concentrations of NaOH solution yielding greater expansions [23,48].

2.2.5 Concrete Prism Test (CPT)

The development of the Concrete Prism Test (CPT), standardized as ASTM C1293 [49], was driven by the need to address the limitations of earlier methods offering a more realistic assessment in terms of storage conditions (i.e., temperature and relative humidity) [2]. The CPT was initially conducted using concrete prism specimens with a cement content of 310 kg/m³, stored in humid chambers at 23°C, with a reactivity threshold set at 0.02% expansion after 84 days [43]. However, this setup was later refined in 1994 to better simulate field conditions, increasing the cement content to 420 kg/m³ and an alkali content to 1.25% Na_2O_{eq}, achieved by adding NaOH to the mixing water, leading to an alkali content of 5.25 kg/m^3 [49]. This modification aimed to induce a more accurate representation of field performance, particularly adopting moderate temperature conditions (i.e., 38°C), similar to MBT, for storage conditions. Therefore, the new temperature is high enough to promote the dissolution of silica from reactive aggregates and the formation of ASR gel avoiding excessive solubility and related issues seen at temperatures like 60°C, which can induce artificial reactions (i.e., impact on ASR-gel viscosity) or cause excessive expansion that would not normally occur in the field [19]. Currently, the CPT uses a 1-year expansion threshold of 0.04% to classify aggregates as non-reactive; expansions equal to or exceeding this threshold indicate a potential ASR development [41].

Despite the CPT's higher reliability and closer approximation to field conditions, some aggregates that react slowly may still present classification challenges [41]. Various parameters influence the test results, including alkali leaching, the equivalent alkali content of the system, the use of different fine aggregates, and storage conditions [19,50]. Alkali leaching, for example, can distort test responses, requiring measures such as covering specimens with plastic or accounting for the leached content to reduce the effect [51]. The higher the equivalent alkali content of the system, the greater the expansion; therefore, assessing the PC Na_2O_{eq} (i.e., % Na_2O + 0.658 K_2O) and introducing the remaining alkalis (i.e., 1.25% Na_2O_{eq} - % PC Na_2O_{eq}) as NaOH is essential to guarantee a comparison among the outcomes. These considerations are critical in maintaining the

test's accuracy, as CPT is designed to reduce the likelihood of false positives and provides a better correlation with field performance [19]. However, the test's long duration remains a significant challenge, especially when timely project planning and decision-making are required.

Besides evaluating the susceptibility of an aggregate to developing ASR, the CPT also assesses the combination of aggregate with cementitious materials as a common mitigation strategy for ASR, such as the employment of slag [49,52]. In this context, the test is conducted similarly to the standard CPT but the cement is partially replaced with the desired amount of SCM on a percent-by-mass basis. Additionally, the alkali content is proportionally reduced from the standard 5.25 kg/m³ Na_2O_{eq}. Furthermore, the evaluation period is extended to 2 years, although the 0.04% expansion threshold remains the critical limit for determining reactivity.

2.2.6 Accelerated Concrete Prism Test (ACPT)

The development of the Accelerated Concrete Prism Test (ACPT) was driven by the need to shorten the lengthy testing period associated with the CPT while maintaining a realistic assessment of ASR potential. Initially proposed in 1992, the ACPT aimed to accelerate the rate of expansion by increasing the storage temperature to 60°C, thereby reducing the test duration from a year to as little as 13 weeks [50]. Additionally, expansion in the ACPT has been correlated with CPT identifying potential thresholds, in which a 0.03% expansion at 3 months at 60°C corresponds to the 0.04% expansion at 1 year at 38°C observed in CPT. Currently, ACPT is a recommended test method from RILEM, AAR-4.1, following the high temperature of 60°C, but with a test duration suggested for at least 20 weeks [21]. Despite the faster results, the ACPT faces challenges, particularly with alkali leaching, which has shown increased leaching from test prisms [50,51]. Consequently, careful consideration of leaching effects is essential when interpreting ACPT outcomes, as higher temperatures tend to intensify this issue [21].

2.2.7 Miniature Concrete Prism Test (MCPT)

The Miniature Concrete Prism Test (MCPT), standardized as AASHTO T380 [53], was developed aiming to provide an accelerated and reliable assessment of aggregate reactivity. The MCPT utilizes smaller concrete prisms (i.e., 50x50x285 mm) with a cement content of 420 kg/m³ and an alkali content boosted to 1.25% Na_2O_{eq}. The MCPT addresses the aggregate gradation issue, by employing coarse aggregate sizes ranging from 4.75 mm to 12.5 mm, eliminating the need for

crushing and offering a more field-realistic evaluation. The prisms are immersed in a 1M NaOH solution at 60°C for a primary test period of 56 days, a lower temperature compared to AMBT to avoid excessive reactivity. Although the high temperature, which could lead to excessive expansion, studies have shown that the MCPT correlated well with field data when compared to CPT and AMBT [54,55].

2.2.8 Norwegian Concrete Prism Test (NCPT)

The Norwegian Concrete Prism Test (NCPT), also recommended by RILEM as AAR-10, is similar to CPT but employs larger prisms (i.e., 100x100x450mm) to reduce leaching effects [56]. This means that relative to their volume, they have less exposed surface area from which alkalis can leach out into the surrounding environment [57]. Similar to the CPT, the NCPT involves storing concrete prisms at 38°C in a humid environment to accelerate ASR. This test is particularly valuable for assessing the potential reactivity of aggregate combinations with binder compositions, including both pure PC and composite cement containing SCMs [56]. The NCPT has two primary applications: first, it evaluates how the content of SCMs can mitigate ASR susceptibility in specific aggregate combinations, and second, it determines the required binder composition to produce non-reactive concrete. The acceptance criteria differ depending on the purpose of the test. When evaluating the reactivity of aggregates combined with a high alkali cement, the threshold is set at 0.04% expansion after 52 weeks, same as CPT. However, for performance testing, where the focus is on assessing the effectiveness of different binder combinations, the criteria changes to a threshold of <0.030% expansion applied after 52 weeks and <0.060% after 104 weeks [58]. Concretes incorporating PC, fly ash, and/or silica fume are only tested for 52 weeks, while all other binder combinations require testing for 104 weeks.

2.2.9 Recent test methods

In recent years, new test methods have been developed aiming to provide rapid assessment of the ACR potential of aggregates, including the Concrete Microbar Test (CMBT). The CMBT, similar to the AMBT, utilizes smaller prisms (i.e., 40 x 40 x 160 mm) while maintaining high temperatures (i.e., 80°C) and immersion in 1M NaOH for 30 days [59]. Since the method focuses on ACR, and some carbonate aggregates produce expansion only when used in larger particles, the aggregate size is increased to a range of 5 to 10 mm. When incorporated as a method recommended by

RILEM (AAR-5), additional adjustments were made to the test period and aggregate distribution [21]. For instance, final readings are taken at 14 days, and aggregate particles exceeding 8 mm must be crushed.

The recent methods offer various approaches depending on specific project requirements and aggregate characteristics [16,28]. They address the limitations of earlier tests by incorporating variable alkali levels and exposure conditions, aiming to enhance the correlation between laboratory results and field performance [15]. For instance, The German Concrete Method (GCM) and the Danish Mortar Bar Test (TI-B51) are examples of such methods that aim to simulate field conditions more precisely. The GCM uses concrete prisms stored in fog chambers at 40°C for nine months, while the TI-B51 immerses samples in a NaCl solution at 50°C for 52 weeks.

Given the variety of methods to assess AAR potential, Table 2.2 summarizes the key features of current the tests, highlighting differences in sample dimensions, shape, storage conditions, and test duration. Some tests focus on the aggregates (AAR-3, AAR-4.1), some on the mix design performance (AAR-10, AAR-11), and others on both the mix and the exposure performance (AAR-12) [56]. They are usually benchmarked against expansion thresholds at specific times (i.e., 0.1 % for AMBT at 14 days and 0.04 % for CPT at 1 year).

Table 2. 2 – Laboratory procedures for evaluating the AAR potential of aggregates.

Procedure	Sample		Test Conditions		Test duration
	Shape	Size	Temperature	Storage	
Accelerated Mortar Bar Test – AMBT – (RILEM AAR-2.1)	Prism	25x25x285 mm^3	80° C	Samples immersed in a 1 M NaOH solution	14 days
Accelerated Mortar Bar Test – AMBT – (RILEM AAR-2.2 / ASTM C1260-22 / AS 1141.60.1)	Prism	40x40x160 mm^3	80° C	Samples immersed in a 1 M NaOH solution	14 days
Concrete Prism Test – CPT – (RILEM AAR-3 / CSA A23.2-14A / AS 1141.60.2)	Prism	75x75x250 mm^3	38° C	RH > 95 %	52 weeks
Accelerated Concrete Prism Test – ACPT – (RILEM AAR-4.1)	Prism	75x75x250 mm^3	60° C	RH > 95 %	20 weeks
Concrete Microbar Test – CMBT – (RILEM AAR-5)	Prism	40x40x160 mm^3	80° C	Samples immersed in a 1 M NaOH solution	14 days
Miniature concrete prism test – MCPT (AASHTO T380)	Prism	50x50x285 mm^3	60° C	Samples immersed in a 1 M NaOH solution	56 days
Danish Mortar Bar Test – TI-B51	Prism	40x40x160 mm^3	50° C	Samples immersed in a 1 M NaCl solution	52 weeks
Norwegian concrete prism test – NCPT – (RILEM AAR-10)	Prism	100×100×450 mm^3	38° C	RH > 95 %	52 weeks
Concrete Cylinder Test – CCT	Cylinder	f = 100 mm h = 200 mm	38° C, 50° C	RH > 95 %	15 weeks
German Concrete Method – GCM	Prism Cube	100x100x450 mm^3 300x300x300 mm^3	40° C	Samples storage in fog chambers	9 months
Alkali-Wrapped Concrete Prism Test – AW-CPT – (RILEM AAR-13)	This procedure can be combined with any of the above methods			Samples wrapped with water-holding material with alkali hydroxide solution (same as concrete pore solution)	

Future directions for research could involve even the refinement of current methodologies by reflecting real-world conditions. These insights would facilitate the translation of laboratory results into actionable guidelines for practical applications. The continuous advance in understanding such relationships ensures the production of durable and long-lasting concrete structures under diverse environmental conditions, contributing even to more sustainable infrastructures.

2.3 Field studies on AAR development

Field-exposed structures (i.e., blocks) are widely acknowledged as the most reliable method for investigating the long-term behaviour against AAR [43]. These structures, exposed to varying environmental conditions over extended periods, provide valuable insights regarding the kinetics

of AAR. As such, field performance is considered a suitable benchmark for evaluating outcomes from laboratory testing methods [19].

Due to the AAR development rate, this approach can be time-consuming and contrasts with the rapid characteristic of an ideal test method [43]. Additionally, it may be limited to the mix variations (i.e., SCM type, alkali content, aggregate nature and type) due to restricted resources. However, the available results must be explored in full to fill the gap between laboratory and field performance, thereby enhancing the predictive accuracy for AAR risk assessment in concrete.

Currently, most ongoing research sites probing long-term AAR performance lie in the northern hemisphere [42,55,58,60–74]. Figure 2. 2 demarcates regions with reported field investigations against AAR: Argentina, Canada, China, France, Germany, Iceland, Italy, Japan, Norway, Portugal, Spain, UK, USA.

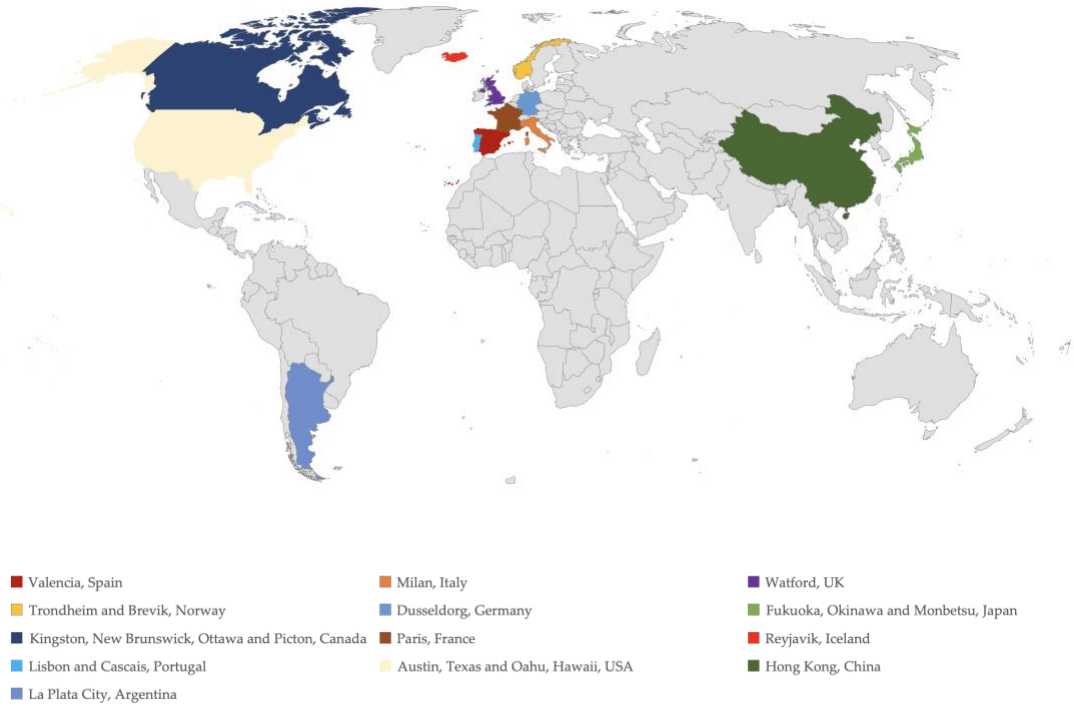


Figure 2. 2 – Location of exposed sites evaluating AAR development on field structures.

2.3.1 Ontario Hydro and Ontario Ministry of Transportation (MTO) Sites - Kingston, Ontario, Canada (1985)

Since the 1960s, Ontario Hydro has conducted extensive studies to validate laboratory tests against field performance, particularly focusing on the effects of freezing, thawing, and deleterious aggregates on concrete durability [16,66]. In 1985, the Ontario Ministry of Transportation (MTO) built six air-entrained concrete sidewalk sections in Kingston, Canada, using argillaceous dolomitic limestone aggregates known for causing ACR. These sections were tested under different conditions, including high and low-alkali concretes, as well as mixtures with slag. All sections exhibited significant expansion, confirming the high reactivity of the aggregates.

In 1991, MTO expanded the study with an additional site using the highly reactive Spratt limestone to create unreinforced and reinforced concrete beams and pavement slabs. Six concrete mixtures with varying alkali content and supplementary cementitious materials (SCMs) like slag, fly ash, and silica fume were tested. Long-term monitoring revealed that the 14-day AMBT correlated better with the 20-year expansion data from the concrete blocks, for those incorporating SCMs compared to the expansion measured at 28 days [75]. Additionally, high-alkali cement concrete cracked at much early ages when compared to low-alkali cement, 5 years versus 12 years respectively [66,76].

After 32 years, all but two mixtures in the unreinforced beams displayed unacceptable field expansions. Additionally, the AMBT provided a better prediction of field performance than the CPT, which tended to underpredict the actual field expansions [77]. The findings underscore the importance of long-term field data in validating ASR mitigation strategies.

2.3.2 Iceland (1987)

In 1987, six air-entrained concrete walls were constructed at the Icelandic Building Research Institute using sea-dredged coarse aggregates from Hvalfjörður and Saltvík, along with a mixture of fine aggregates from different locations [16]. The study aimed to assess the reactivity potential of these aggregates. Therefore, visual surveys and petrographic examinations confirmed the reactivity of the Hvalfjörður aggregates, with cracking and gel formation observed in some walls. Subsequent field and laboratory investigations aimed to provide inputs regarding the reliability of performance-based, such as ACPT, focusing on different types of cement, alkali level, silica fume and reactive aggregate particle sizes [78]. The research confirmed that for concrete mixes,

particularly those with very high or very low alkali levels, the expansion results from concrete prism tests generally matched the expansion observed in cubes at the outdoor exposure site. However, other mixes, especially those using cements with inter-milled silica fume or those with boosted alkali levels, tend to exhibit significantly higher expansion in outdoor conditions than what the concrete prism tests would suggest. Additionally, there are conflicting findings regarding the impact of reactive aggregate particle sizes; while high amounts of reactive sand lead to greater expansion in the prism tests, the outdoor exposure site results indicate that reactive gravel is more influential in driving expansion.

2.3.3 Building Research Establishment (BRE), United Kingdom (Late 1980s – 1990s)

The BRE in the UK established exposure sites, in the late 1980s to early 1990s, evaluate the effect of fly ash, slag, metakaolin, and lithium-based admixtures on ASR [16]. The study included concrete blocks and cubes made with five types of reactive aggregates and various levels of high-alkali cement [79].

The study found that concrete blocks with flint and greywacke aggregates showed significant expansion and cracking, particularly in control mixes without fly ash [80]. Fly ash, when used at 25% and 40% replacement levels, was highly effective in reducing or eliminating expansion and cracking, even in high-alkali environments. Lithium-based admixtures were also tested, indicating that as alkali content increases more lithium is required and the dosage also varies based on the aggregate used [16].

However, the study revealed some inconsistencies between laboratory and field test results [80]. Specifically, field-exposed blocks exhibited more significant expansion than the laboratory CPT had predicted, indicating that the CPT might not fully capture the long-term behavior of ASR under natural conditions. On the other hand, the AMBT provided a reasonable prediction of the amount of fly ash needed to prevent expansion in concrete with flint aggregates. However, the AMBT was unable to determine the influence of PC alkalis on the required fly ash dosage, suggesting a limitation in its predictive accuracy for real-world conditions.

2.3.4 CANMET, Ottawa, Canada (1991)

In 1991, CANMET (Canada Centre for Mineral Energy Technology) initiated a field and laboratory investigation to evaluate the reliability of laboratory tests in predicting ASR and the efficacy of SCMs and lithium-based admixtures in preventing ASR [28,52,81,82]. The study involved over 250 concrete mixtures with reactive aggregates, varying alkali contents, and SCMs like fly ash, slag, and silica fume. Test specimens, including prisms, blocks, and slabs, were exposed outdoors and monitored for ASR development over time.

Initially, accelerated tests, such as AMBT and CPT, were identified as reliable in predicting the long-term performance of SCMs in controlling expansion due to ASR [28]. Additionally, the AMBT generally indicated the effectiveness of SCMs in controlling expansion when evaluated at 14 days with expansion limit of 0.1% [82]. However, later studies indicated that CPT, as outlined in the CSA A23.2-28A [83] standard, was only effective in predicting the short-term efficacy (i.e., up to 10 years) of SCMs in controlling ASR expansion [16,52]. As the project progressed and the test specimens were exposed to outdoor conditions for extended periods, the correlation between CPT results and field performance began to diminish, as CPT underestimated the amount of SCMs required to control ASR [81]. This decline in correlation was partly attributed to factors such as the leaching of alkalis from the specimens and the influence of local environmental conditions at the exposure site.

Additionally, valuable insights emerged from the study regarding the influence of environmental conditions on ASR development [42]. As the concrete specimens were exposed to varying outdoor environments (i.e., Ottawa and Austin), it became evident that climatic conditions, such as temperature and humidity, played a crucial role in ASR progression. For instance, warmer climates were observed to significantly speed up the reaction, leading to deleterious expansion in concrete specimens up to 4 to 5 times faster than in cooler climates. This variation in reaction speed under different environmental conditions highlighted that a level of ASR mitigation that might be sufficient in cooler regions could fall short in warmer areas. Therefore, the study underlined the importance of long-term field exposure tests in different climatic regions to better predict the actual performance of ASR.

2.3.5 Picton Cement Plant, Ontario, Canada (1998)

In 1998, seven outdoor pavement slabs were constructed at the Picton cement plant, in Picton, Ontario, Canada, utilizing the highly reactive Spratt limestone [16,60]. The concrete mixtures included variations with high-alkali cement, silica fume, and slag. After two years, it was attested that the combination of silica fume and blast-furnace slag in ternary blends provides enhanced resistance to ASR expansion compared to using either material alone [60]. Additionally, ternary blend concretes demonstrate greater durability against ASR and other forms of deterioration than OPC mixtures and concretes with a single SCM replacement.

Yet, after six years of exposure, significant ASR cracking was observed only in the control mixture that used high-alkali cement without any supplementary cementitious materials (SCMs) [16]. While the mixtures containing silica fume did not exhibit visible cracking, petrographic analysis revealed the presence of micro-level ASR features [84]. This site has provided valuable insights into the role of SCMs, particularly silica fume and slag, in mitigating ASR.

2.3.6 PARTNER Project (Early 2000s)

The PARTNER project (2002-2006), funded by the European Community, aimed to establish a unified test procedure for evaluating the alkali-reactivity of aggregates across Europe [62]. A total of 100 cubes were produced for 50 tests and it involved 24 partners from 14 countries, with both laboratory and field testing of 13 different aggregate types across eight field locations in Norway, Sweden, Denmark, Germany, France, Spain, Italy, and Portugal [85,86]. The findings from this extensive research for new European standards emphasized the significant influence of environmental conditions on ASR development. Additionally, the project produced a petrographic atlas of potentially reactive rocks in Europe, providing a valuable resource for future assessments [32,87].

The conclusions of the PARTNER project highlight that the RILEM test methods were successful in identifying the potential alkali-reactivity of most aggregates, particularly those that react within "normal" time frames of 5 to 20 years [86]. However, the methods were less reliable in detecting "slowly" reactive aggregates, which exhibit reactivity over longer periods, typically exceeding 15-20 years. Discrepancies between lab and field results were generally linked to uncertainties in field data, aggregate variability, or pessimum effects, indicating the need for further localized studies.

Moreover, the field tests revealed that ASR occurs similarly across different European climate zones, though the reaction tends to manifest earlier in southern Europe due to higher temperatures.

2.3.7 University of Texas at Austin, Texas, USA (2001)

The University of Texas at Austin, funded by TxDOT, established an exposure site in 2001 aimed at understanding the mechanisms of ASR and DEF, developing effective test methods and preventive measures, and transferring the knowledge gained to improve the durability of local transportation infrastructure [16,88]. The study involved extensive laboratory testing, field evaluations, and the development of new specifications and guidelines to mitigate ASR and DEF in both new and existing concrete structures.

A significant aspect of the project was the evaluation of variations in test methods, such as AMBT and CPT, for their ability to predict ASR-related expansion and cracking in concrete. Initially, the CPT was identified as the most reliable for predicting field performance, particularly when run for extended periods [89]. However, challenges were noted in its ability to assess the effects of cement alkalinity on expansion. The AMBT was found to be a reasonable indicator of aggregate reactivity but tended to be exceeded for some aggregates, leading to potential false positives. The research also highlighted the importance of real-world field testing, with outdoor exposure sites providing insights into the performance of mitigation measures like SCMs and lithium compounds.

A recent study of the outdoor exposure site over a 20-year period indicated that correlations between laboratory tests like the AMBT and the CPT with field performance have revealed discrepancies [88]. Specifically, these laboratory methods often underestimate the amount of SCMs needed to prevent ASR in high alkali loading scenarios, and they fail to predict the inefficacy of lithium nitrate in such cases. The exposure blocks continue to be monitored annually, providing valuable long-term data to inform future specifications and best practices regarding ASR prediction.

2.3.8 University of New Brunswick Campus, Fredericton, Canada (2005)

The University of New Brunswick (UNB) operates an outdoor exposure site on its Fredericton campus, focusing on different types of concrete, including high-alkali control and fly ash mixtures [16]. In 2005, a research program was initiated to evaluate concrete mixtures for the potential

reconstruction of the Mactaquac Generating Station, which is heavily affected by ASR [65]. Blocks made with highly reactive Springhill greywacke and different fly ash contents were installed near the generating station.

After seven years of monitoring, significant expansion was only observed in the high-alkali control block, demonstrating the effectiveness of fly ash in controlling ASR [16]. It was observed that the use of low-calcium fly ash can significantly reduce the damaging effects of ASR in concrete [65]. However, due to the highly reactive nature of the aggregates, a fly ash content exceeding 30% may be necessary to limit expansion to acceptable levels. Additionally, it has been suggested that the existing AMBT and CPT, even when applying the 28-day and 2-year expansion benchmarks, might not adequately determine the necessary prevention measures for aggregates that contribute substantial amounts of alkali to concrete.

2.3.9 COIN Project (2007-2014)

The COIN project (2007-2014), conducted in Norway and linked with the RILEM TC 219-ACS initiative, aimed to develop reliable ASR performance testing methods by examining the impact of various parameters, such as curing and storage conditions, on concrete's internal moisture state, alkali leaching, and ASR expansion [90]. The project was divided into two parts: a study (Part I) that highlighted the critical role of alkali leaching in controlling ASR expansion, and a follow-up study (Part II) that tested 20 different concrete mixtures across two field exposure sites in Norway and Portugal. The project tested concrete mixtures with different cement types and alkali contents using control reactive aggregates from Norway (Ottersbo) and Canada (Spratt) [16].

The findings showed that increasing prism cross-sections significantly reduced alkali leaching and improved the reliability of CPT results [91]. However, challenges remain with the use of "alkali wrapping" procedures, particularly when testing low-alkali systems. The results contributed to refining ASR testing protocols and informed recommendations for future ASR performance testing standards.

2.3.10 Oregon State University, USA (2011)

At Oregon State University, an exposure site was established in 2011 to investigate the potential of fine lightweight aggregates in mitigating ASR in concrete [16]. This site focuses on evaluating

the performance of concrete blocks containing lightweight aggregates, with a particular emphasis on understanding how these materials can reduce or prevent ASR-related expansion.

2.3.11 University of Toronto (UofT) Leaside Site, Toronto, Canada (early 2010s)

The University of Toronto (UofT) Leaside outdoor exposure site was established in the early 2010s. Specifically, the site was active around 2015 and has been operated at the St. Marys Cement head office in Leaside, Toronto [16,92]. This site, which focuses on evaluating the effectiveness of SCMs in low-alkali concrete mixes, includes air-entrained concrete cubes made with reactive Sudbury and Spratt aggregates. After three years of exposure, the blocks showed minimal expansion (i.e., ≤0.02%), indicating effective ASR mitigation.te

2.3.12 University of Hawaii, Manoa, USA & DOT Facility, Lawrence, Massachusetts, USA (early 2010s)

The exposure site at the University of Hawaii in Manoa, established in June 2011, serves as part of the Federal Highway Administration's ASR Development and Deployment Program [93]. At this site, 40 concrete mixtures were produced using local basaltic aggregates from various Hawaiian quarries, as well as imported sand from British Columbia (Canada). The mixtures incorporated varying alkali contents, fly ash, and lithium-nitrate admixtures to assess their effectiveness in mitigating ASR. The blocks, placed at the Magoon Research and Instruction Facilities, are being closely monitored for length changes and the onset of cracking over time. The aim is to compare the data gathered with laboratory tests to evaluate the long-term predictive accuracy of ASR susceptibility under Hawaii's specific environmental conditions.

Similarly, at the Lawrence, Massachusetts site, established in June 2012, concrete blocks from 73 different mixtures, including those with highly reactive aggregates, were produced [93]. By maintaining identical blocks across multiple locations, such as the Hawaii site and other sites in Texas, Ontario and New Brunswick, researchers aimed to assess the impact of environmental exposure on ASR progression, along with the effectiveness of mitigation strategies.

2.3.13 RILEM TC 258-AAA Project (2015)

The RILEM TC 258-AAA project, launched in 2015, is part of a broader European effort to evaluate ASR under various environmental conditions [16]. Concrete mixtures with highly reactive aggregates (i.e., New Mexico – NM, Texas; Ottesbo – Ott, Norway) and different levels of Class F fly ash (i.e. 20% and 30%) were cast into cubes at LNEC (Portugal) and distributed across 10 exposure sites in Europe and North America. These sites include Trondheim (Norway), Brevik (Norway), Düsseldorf (Germany), Paris (France), Lisbon (Portugal), Cascais (Portugal), Reykjavik (Iceland), Austin (USA), Treat Island (USA), and Ottawa (Canada).

Preliminary results revealed that the control mixes with NM aggregates exhibited expansions between 0.28% and 0.42% after two years, with the Texas and LNEC sites recording the most significant expansion for the Ott control cubes. In contrast, the cubes incorporating 20% or 30% Class F fly ash did not show any noticeable expansion after two years of outdoor exposure, demonstrating the effectiveness of fly ash in mitigating alkali-silica reaction (ASR) under the tested conditions.

2.3.14 ODOBA International Project (2016)

The ODOBA (*Observatoire de la Durabilité des Ouvrages en Béton Armé*) project, initiated in 2016 by IRSN (*Institut de Radioprotection et de Sûreté Nucléaire*), focuses on studying internal swelling reactions like DEF and AAR in large-scale concrete structures, specifically relevant to nuclear power plant containment [94]. Located at the ODE (*Observatoire de la Durabilité des Enceintes*) outdoor experimental platform in Cadarache, France, the project involves the construction of massive 8 m³ concrete blocks that mimic the composition and structure of those used in French nuclear reactors.

The blocks are heavily instrumented to monitor the effects of ISR under both natural and accelerated aging conditions. The study has shown that blocks with moderate alkali levels and potentially reactive aggregates have so far exhibited minimal AAR-related expansion, prompting a shift in focus towards developing more reactive concrete mixes for future experiments. The insights gained from ODOBA aim to improve the understanding of ISR in large structures and guide the design of durable concrete for nuclear facilities.

2.3.15 Technical University of Denmark (DTU) (2018)

The exposure site at the Technical University of Denmark (DTU) in Kongens Lyngby, Denmark, is dedicated to studying ASR in concrete blocks containing different reactive aggregates [95]. Established as part of a broader research effort from 2018 to 2023, the site is used to assess the long-term effects of ASR, comparing natural exposure with accelerated test methods. Additionally, the site evaluates the effectiveness of hydrophobic silane impregnation in reducing or delaying ASR, with continuous monitoring of moisture content and structural integrity.

2.3.16 Switzerland exposure sites, Midlands and Alps (2022)

In Switzerland, a project launched in 2022 involves establishing two ASR exposure sites to study the long-term behavior of concrete with reactive aggregates under natural conditions [96]. One site is located in the Swiss Midlands, representing a moderate climate, and the other is in the Alps at an altitude of 2200 meters, exposing the concrete to harsher, alpine conditions. The goal is to validate CPT and Residual Expansion Test (RET), commonly used to assess the durability of structural concrete. Forty different concrete mixtures, reflecting those used in Swiss engineering and dam construction, have been produced and tested. The concrete cubes will be monitored over 15 years, with periodic core sampling to evaluate the tests' effectiveness. Initial findings show that temperature fluctuations significantly influence the expansion rates, with warmer conditions potentially accelerating ASR development. Additionally, the study aims to enhance the understanding of how laboratory test results correlate with real-world conditions.

2.3.17 Summary of exposed fields

Table 2. 3 uncovers information regarding some exposed field specimens, such as block shape, sizes, surface area (m^2) and volume (m^3). As observed, field studies predominantly utilize non-standardized block sizes, ranging from 100 mm to 3000 mm, nor structural elements, such as prims and cubes. Among the goals of the current studies on field structures, the following is observed:

- AAR mitigation solutions, emphasizing the use of SCMs in real-world environments [60,64,65,73];
- Microstructure of AAR-affected concrete [62];
- New laboratory test procedures [61,68];

- Evaluating the long-term performance of exposed blocks to different environmental conditions [42,69].

Table 2. 3 –Occurrence of field exposed structures by their characteristics.

Shape	Size	Surface area (m2)	Volume (m3)	Ref
Prism	400x400x700 mm^3	1.440	0.112	[42,63,65,71–73]
Prism	380x380x710 mm^3	1.368	0.103	[55,61,67]
Prism	400x400x600 mm^3	1.280	0.096	[68,69]
Prism	600x600x2000 mm^3	1.200	0.072	[66,73]
Prism	150x700x700 mm^3	1.400	0.074	[72,73]
Cube	300x300x300 mm^3	0.540	0.027	[58,62]
Prism	200x1200x4000 mm^3	11.680	0.960	[66]
Prism	286x910x910 mm^3	2.697	0.237	[64]
Cube	3000x3000x3000 mm^3	54.000	27.000	[65]
Prism	60x100x200 mm^3	0.076	0.001	[74]
Prism	1200x4000x2000 mm^3	30.400	9.600	[73]
Cube	100x100x100 mm^3	0.060	0.001	[58]
Cube	150x150x150 mm^3	0.135	0.003	[58]

Additionally, it is observed the absence of a universally accepted AAR assessment protocol for field evaluation, with over 20 evaluation methodologies being identified. They include simply measuring expansion/length measurement (E), crack width (CW), porosity (P), and mechanical properties: compressive strength (CS), modulus of elasticity (ME), flexural strength (FS) to chemical and microscopic evaluations: chemical analysis (CA), pore solution analysis (PS), chloride diffusion (CD), rapid chloride permeability test (RCPT), alkali-leaching analysis (AL), salt scaling (SS), thermogravimetric analysis (TGA), concrete air-void spacing factor (ASF), Raman microscopy (RM), scanning electron microscope (SEM). In addition, AAR-affected concrete has also been assessed using more elaborate techniques, including petrographic analysis (PA), damage rate index (DRI), stiffness damage test (SDT), and qualitative damage assessment (QDA). Finally, it is worth mentioning that only two non-destructive approaches have been observed: ultrasonic pulse velocity (UPV) and bulk electrical resistivity (BER).

Finally, the current field structure results are being compared with seven different laboratory testing methods to assess aggregates' reactivity. It was observed that that ACPT, AMBT, AW-CPT, CCT, CPT, MCPT, and NCTP were among the tests performed. In fact, the CPT test was the most common, followed by the AMBT test. These observations restate that the ABMT and CPT

are among the most commonly used and trusted tests to determine the aggregate reactivity potential.

2.4 Statistical Methods

Uncertainty is an unavoidable aspect of engineering, arising from the inherent randomness in natural processes (i.e., aleatory uncertainty) and the limitations in knowledge or predictive models (i.e., epistemic uncertainty) [97]. Aleatory uncertainty is tied to the variability in data that is naturally present and often cannot be reduced, reflecting the unpredictable nature of real-world phenomena. On the other hand, epistemic uncertainty results from incomplete or imperfect knowledge, which can potentially be reduced through better models or improved data collection. Despite their differences, both types of uncertainty play a critical role in engineering and must be accounted for when making decisions about design and planning.

Therefore, statistical and probabilistic methods are essential tools for managing uncertainty in engineering decision-making. These approaches allow engineers to quantify the likelihood of various outcomes and assess the risks associated with them, leading to more informed and balanced decisions. For instance, when designing structures, engineers can use probabilistic models to estimate the probability of failure under different conditions and to evaluate the trade-offs between cost, safety, and performance. By integrating these methods into the decision-making process, designs that are both reliable and optimized for efficiency can be created, even in the face of uncertainty.

In summary, statistics can be understood as the science of data. This involves collecting, summarizing, organizing, analyzing, interpreting data, and building models [98,99]. Therefore, statistical methods are mathematical tools used to analyze, interpret, and make inferences from data. These methods are essential in research for identifying patterns, testing hypotheses, and making decisions based on data. Statistical methods can be broadly categorized into two general branches, including descriptive statistics and inferential statistics [99,100].

Descriptive statistics focuses on data collection, organization and presentation. However, it doesn't allow making conclusions beyond the analyzed data. In descriptive statistics, three general methods are used, including measures of central tendency: mean, median, and mode; measures of spread: range, quartiles and percentiles, variance, and standard deviation; and, graphical displays,

such as dot plots, frequency distribution, histograms, frequency polygons, bar graphs, pie chart, box and whiskers plot [100].

In contrast, inferential statistics makes use of the features of a sample to make assumptions about the unknown parameters in a certain population [101]. As a result, it is concerned with drawing conclusions based on a dataset by extending beyond the available information. In inferential statistics, probability theory is utilized to make decisions regarding the distributions of the data employed in descriptive statistics. Several methods of inferential statistics include sampling theory, estimation theory, hypothesis testing, and regression analysis.

In sampling theory, the goal based on sample mean, variance and distribution is to select samples large enough to be examined completely [101]. In estimation theory, the prediction or estimation of a population parameter based on the available data is the target relying on point estimate, interval estimate and confidence interval techniques. Additionally, it accounts for techniques like maximum likelihood estimation and minimum mean squared error estimation to make assumptions. In hypothesis testing, a hypothesis is assumed in order to be accepted or rejected given a specific confidence limit. Finally, regression analysis reveals the mathematical relationship between variables finding an expression to best represent the data.

Bayesian interference is generally employed to fit a probability model to a set of data, summarizing results as probability distributions on parameters [102,103]. The Bayesian data analysis process involves three key steps: (i) setting up a full probability model (i.e., prior distribution); (ii) conditioning on observed data (i.e., likelihood function acting on the posterior distribution); and, (iii) evaluating the fit of the model [102]. For instance, Bayes' Theorem [103] can be employed to identify the probability of exposed blocks being reactive given the laboratory test outcomes (i.e., AMBT and CPT) and thresholds [41].

Figure 2. 3 describes the stages of Bayesian inference, including the prior distribution as a non-informative beta distribution; the likelihood function based on the observed data as a binomial distribution; and finally, the posterior distribution, reflecting the updated belief after considering the new data.

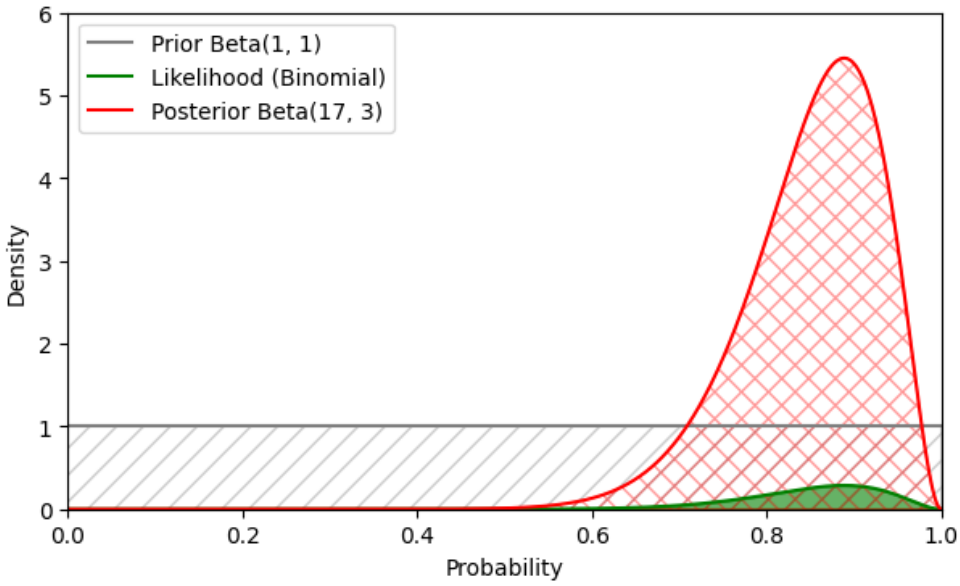


Figure 2. 3 – Stages of Bayesian inference: Prior, Likelihood and Posterior functions.

The prior probability $P(R)$ represents the initial belief about the probability that the dependent variable, i.e., whether the block is reactive (yes or no), given the conditional variables (i.e., level of reactivity, environment and alkali content). As seen in (Equation 2. 1), this probability is calculated as the total number of reactive blocks ($N_{reactive}$) divided by the total number of blocks (N_{total}).

$$P(R) = \frac{N_{reactive}}{N_{total}} \quad \text{(Equation 2. 1)}$$

In the case of continuous random variables ranging from 0 to 1, a Beta distribution is often used as a prior function for the parameter p of the binomial distribution [104]. It provides a way to incorporate uncertainty around the calculated probabilities. The Beta distribution's probability density function (PDF) is parameterized by two positive parameters, α and β, and is given by (Equation 2. 2).

$$f(t;\, \alpha, \beta) = \frac{1}{B(\alpha, \beta)} t^{\alpha-1}(1-t)^{\beta-1} \quad \text{(Equation 2. 2)}$$

Where and $B(\alpha, \beta)$ is the beta function defined as (Equation 2. 3)

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt \quad \text{(Equation 2. 3)}$$

For this study, a non-informative prior was assumed $\alpha \; and \; \beta = 1$, corresponding to a uniform distribution over [0, 1]. This choice avoids introducing bias into the model.

The likelihood function $P(L_i|R)$ incorporates the probability of the observed data given the conditional parameters. In the current case, the likelihood is expressed as the number of reactive blocks at a specific condition ($N_{L_i}^{reactive}$) divided by the total number of reactive blocks (Equation 2. 4).

$$P(L_i|R) = \frac{N_{L_i}^{reactive}}{N_{reactive}} \quad \text{(Equation 2. 4)}$$

In the Bayesian analysis, the likelihood of observing *x* successes in *n* trials is given by the binomial distribution, as presented in (Equation 2. 5).

$$Binomial\ (x|n,p) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{(Equation 2. 5)}$$

The marginal likelihood $P(L_i)$, representing the overall probability of observing a specific condition (L_i) is assessed as indicated in (Equation 2. 6), by the division of the total number of blocks at a specific condition ($N_{L_i}^{total}$) over the total number of blocks (N_{total}). The marginal likelihood normalizes the posterior distribution, ensuring total probability is equal to 1.

$$P(L_i) = \frac{N_{L_i}^{total}}{N_{total}} \quad \text{(Equation 2. 6)}$$

Therefore, by combining the prior and likelihood using Bayes' Theorem, the posterior probability ($(P(R|L_i)$). This posterior represents the probability of a block being reactive given the observed condition (Equation 2. 7).

$$P(R|L_i) = \frac{P(L_i|R) \cdot P(R)}{P(L_i)} = \frac{N_{L_i}^{reactive}}{N_{L_i}^{total}} \quad \text{(Equation 2. 7)}$$

Expressing the posterior probability as a Beta distribution of α' and β' parameters, the posterior can be represented as per (Equation 2. 8) [105].

$$Beta\ (\alpha', \beta') = Beta\ (N_{L_i}^{reactive} + 1, N_{L_i}^{non-reactive} + 1) \quad \text{(Equation 2. 8)}$$

Finally, the cumulative distribution function (CDF) of the Beta distribution gives the probability that a Beta-distributed random variable is less than or equal to *x*. The inverse of CDF is known as the percentile point function (PPF) or the quantile function, allows for the computation of confidence intervals. For a 95% confidence interval by beta distribution, the values required are at 2.5th and 97.5% percentile of the distribution. Therefore, the lower and upper bounds are given as per (Equation 2. 9) and (Equation 2. 10), respectively.

$$CI_{lower} = Beta^{-1}(0.025; \alpha, \beta) \quad \text{(Equation 2. 9)}$$

$$CI_{upper} = Beta^{-1}(0.975; \alpha, \beta) \quad \text{(Equation 2. 10)}$$

In summary, this Bayesian approach, incorporating the Beta distribution allows for probabilistic modelling that relies on observed data and includes uncertainty quantification.

2.5 Machine Learning

Machine learning (ML) is based in a multidisciplinary concept, combining artificial intelligence, probability and statistics, computational complexity theory, control theory, information theory, philosophy, psychology, neurobiology, and so on [106]. Its relevance lies in its ability to uncover patterns and insights from vast amounts of data, enabling more accurate predictions and decision-making. ML ca ben categorized into three groups: supervised learning, unsupervised learning, and reinforcement learning [107]. Supervised learning relies on labeled data, where the training involves input-output pairs to learn a mapping function. In unsupervised learning, the approach

involves analyzing and identifying patterns within unlabeled data to extract meaningful features. On the other hand, reinforcement learning is based on the interaction between states, actions, and rewards to develop an optimal policy.

A few models are popular in ML, among them Support Vector Machine (SVM), Decision Trees (DT), Regression Analysis, and k-Nearest Neighbors (k-NN) [107]. For instance, regression analysis predicts outcomes by modeling relationships between variables, with linear models for straightforward relationships and nonlinear methods like logistic regression for more complex scenarios. Additionally, DT are predictive models that recursively split data into subsets, with classification trees handling discrete inputs and regression trees processing continuous data, often using pruning to prevent overfitting. Therefore, for classification problems, such as predicting the likelihood of field reactivity, two main models can be highlighted: logistic regression and DT.

The logistic regression model is based on the logistic function (Equation 2. 11), which transforms input data into a probability value ranging between 0 and 1, suitable for modeling binary outcomes.

$$Logistic\ function = \frac{1}{1 + e^{-x}} \quad \text{(Equation 2. 11)}$$

In this context, a weight factor is used to predict the log odds of the probability that a given input belongs to a particular class (0 or 1). The log odds, or the logit function (Equation 2. 12) is defined as the natural logarithm of the odds ratio, in which P represents the probability of the event occurring.

$$odds = \frac{P}{P-1} \ \rightarrow logit(P) = ln\left(\frac{P}{1-P}\right) \quad \text{(Equation 2. 12)}$$

Considering the logit function, the probability occurrence can be estimated as per (Equation 2. 13).

$$logit(P) = \ mx + b \ \rightarrow P(x) = \frac{1}{1 + e^{-(mx+b)}} \quad \text{(Equation 2. 13)}$$

The maximum likelihood estimation (MLE) is the loss function in the logistic regression. It aims to find the parameter values that maximize the likelihood of the observed data. In this process,

MLE optimizes the model parameters by interactively adjusting them to ensure the best fit between the model and the data. MLE is expressed as in (Equation 2. 14), where y_i is the actual output and $\hat{y}_i$ is the predicted probability.

$$L(m) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i) \quad \text{(Equation 2. 14)}$$

In logistic regression, each coefficient represents the change in the log odds of the outcome for a one-unit change in the predictor variable, while keeping the other variables constant. In that sense, this approach was employed to train the model to predict the occurrence of reactivity in the field (Equation 2. 15) based on the accelerated laboratory test outcomes (i.e., aggregate reactivity levels or expansion, EXP), environmental conditions of exposed elements (i.e., temperature, TEMP and relative humidity, RH) and mix-design information (i.e., alkali content, ALK).

$$x=\beta_0+\beta_{TEMP}\cdot\text{TEMP}+\beta_{RH}\cdot\text{RH}+\beta_{EXP}\cdot\text{EXP}+\beta_{ALK}\cdot\text{ALK} \quad \text{(Equation 2. 15)}$$

The probability of reactivity can, therefore, be calculate as per (Equation 2. 16).

$$P(x) = \frac{1}{1 + e^{-(\beta_0+\beta_{TEMP}\cdot\text{TEMP}+\beta_{RH}\cdot\text{RH}+\beta_{EXP}\cdot\text{EXP}+\beta_{ALK}\cdot\text{ALK})}} \quad \text{(Equation 2. 16)}$$

DT, a type of supervised machine learning model, is particularly well-suited for tasks where the goal is to establish decision rules based on a set of input variables [108]. Following the structure presented in Figure 2. 4, DT can effectively incorporate threshold rules, making them ideal for determining the optimal expansion levels and time points that predict reactivity. Each split in the tree aims to maximize the separation between reactive and non-reactive outcomes, thereby refining the predictive accuracy.

Along its structure, the root node represents the entire dataset, being the starting point of the tree and the most significant decision is made. Next, internal nodes represent the points where the data is further split based on additional characteristics, leading to further branching. Leaf nodes are then the terminal points providing the output or classification where no further splits occur, representing a final classification made by the model.

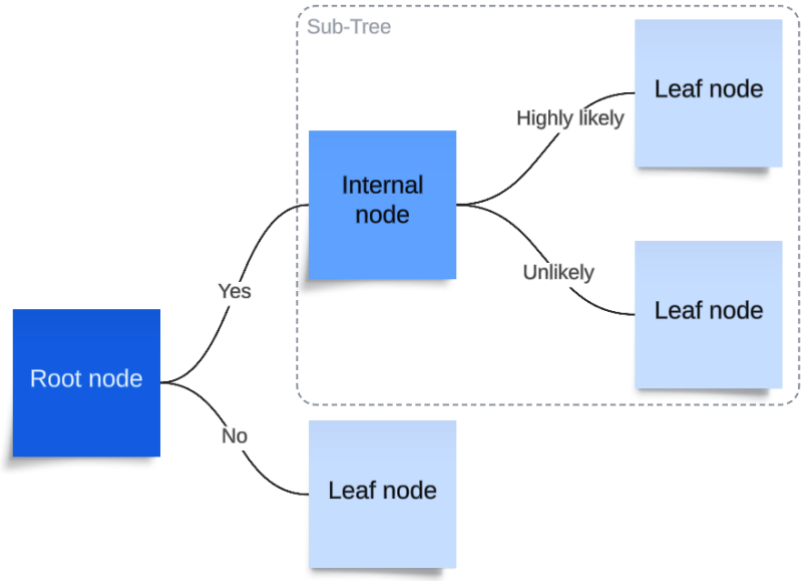


Figure 2. 4 – Decision tree structure, including root, internal and leaf nodes.

In this sense, the dataset was split into training and test sets usually using a 70-30 ratio, ensuring that the model's performance could be robustly evaluated on unseen data. Moreover, the model is trained using a recursive partitioning approach, where data is split based on expansion percentages at key time intervals. To prevent overfitting and ensure the model remains interpretable, the DT is constrained to a maximum depth of three. Additionally, a minimum amount of samples can be defined to split an internal node or a leaf node.

Evaluating the models involves evaluating their performance, typically conducted through a sensitivity analysis using a confusion matrix. This matrix quantifies the proportions of true positives (TP), true negatives (TN), and their false counterparts: false positives (FP) and false negatives (FN) [109]. By breaking down the model's predictions into these categories, it becomes possible to assess the aspects of the model's performance, leading to a more comprehensive understanding of its accuracy and reliability.

The metrics include sensitivity (i.e., TP/(TP+FN)), specificity (i.e., also known as recall=TN/(TN+FP)), precision (i.e., positive predictive value or TP/(TP+FP)), and the F-1 score (i.e., the harmonic mean of precision and specificity). Sensitivity is particularly important in scenarios where missing a positive case could have serious consequences. On the other hand, specificity is vital in contexts where false positives can lead to preventable actions. Precision indicates the proportion of correct positive predictions, making it crucial in situations where the

cost of a false positive is high. Finally, the F1 score is useful when both precision and recall are needed, ensuring that the model performs well across different aspects.

2.6 Conclusion

AAR is a critical factor in the deterioration of concrete, particularly concerning the long-term durability of infrastructure. Encompassing both ASR and ACR, AAR leads to internal cracking, expansion, and subsequent degradation of concrete structures. The global scope of research on AAR underscores its significant impact, making it one of the foremost concerns in ensuring concrete performance over time.

Laboratory testing methods, such as the Accelerated Mortar Bar Test (AMBT) and Concrete Prism Test (CPT), play a vital role in evaluating the reactivity potential of aggregates. However, numerous field studies conducted under varying environmental conditions have consistently revealed discrepancies between laboratory test outcomes and actual field performance. These findings highlight the need for improved correlation between laboratory results and real-world behavior, as field exposure tests often uncover long-term effects of AAR that accelerated laboratory tests fail to predict accurately. This gap underscores the importance of leveraging available data more effectively through the integration of statistical methods and machine learning. By applying probabilistic models and ML algorithms, it is possible to explore the likelihood of AAR-related damage under diverse conditions, thereby enhancing the predictive power of laboratory tests. These tools enable more accurate risk assessments by quantifying uncertainties, analyzing patterns in extensive datasets, and developing models that better predict the potential for field reactivity. Consequently, incorporating these methods into a comprehensive framework allows for a more robust assessment of AAR risks in the field, ultimately improving the durability and safety of infrastructure.

2.7 Reference

[1] G.E. Blight, M.G. Alexander, Alkali-Aggregate Reaction and Structural Damage to Concrete, 2011.

[2] M.D.A. Thomas, B. Fournier, K.J. Folliard, Alkali-Aggregate Reactivity (AAR) Facts Book, Book (2013) 224.

[3] M. Alexander, G. Blight, Southern and Central Africa, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[4] E.M.R. Fairbairn, South and Central America, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[5] V.R. Falikman, N.K. Rozentahl, Russian Federation, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[6] I. Fernandes, Ö. Andiç-Çakir, C. Giebson, K. Seyfarth, Mainland Europe, Turkey and Cyprus, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[7] T. Kay, A.B. Poole, I. Sims, Middle East & North Africa, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[8] J. Lindgård, B. Grelk, B.J. Wigum, J. Trägårdh, K. Appelqvist, E. Hol, M. Ferreira, M. Leivo, Nordic Europe, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[9] A.K. Mullick, Indian Sub-Continent, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[10] A. Shayan, S. Freitag, Australia and New Zealand, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[11] I. Sims, United Kingdom and Ireland, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[12] M.D.A. Thomas, K.J. Folliard, J.H. Ideker, North America (USA and Canada), in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[13] K. Yamada, T. Miyagawa, Japan, China and South-East Asia, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[14] L.F.M. Sanchez, T. Drimalas, B. Fournier, D. Mitchell, J. Bastien, Comprehensive damage assessment in concrete affected by different internal swelling reaction (ISR) mechanisms, Cement and Concrete Research 107 (2018) 284–303. https://doi.org/10.1016/j.cemconres.2018.02.017.

[15] I. Borchers, J. Lindgård, C. Müller, Evaluation of laboratory test methods for assessing the alkali-reactivity potential of aggregates by field site tests, Materconstrucc 72 (2022) e286. https://doi.org/10.3989/mc.2022.17221.

[16] B. Fournier, J. Lindgård, B.J. Wigum, I. Borchers, Outdoor exposure site testing for preventing Alkali-Aggregate Reactivity in concrete – a review., MATEC Web Conf. 199 (2018) 03002. https://doi.org/10.1051/matecconf/201819903002.

[17] Y. Kawabata, K. Yamada, The mechanism of limited inhibition by fly ash on expansion due to alkali–silica reaction at the pessimum proportion, Cement and Concrete Research 92 (2017) 1–15. https://doi.org/10.1016/j.cemconres.2016.11.002.

[18] R.E. Oberholster, G. Davies, An accelerated method for testing the potential alkali reactivity of siliceous aggregates, Cement and Concrete Research 16 (1986) 181–189. https://doi.org/10.1016/0008-8846(86)90134-1.

[19] J. Lindgård, Ö. Andiç-Çakır, I. Fernandes, T.F. Rønning, M.D.A. Thomas, Alkali–silica reactions (ASR): Literature review on parameters influencing laboratory performance testing, Cement and Concrete Research 42 (2012) 223–243. https://doi.org/10.1016/j.cemconres.2011.10.004.

[20] B. Fournier, M.-A. Bérubé, Alkali–aggregate reaction in concrete: a review of basic concepts and engineering implications, 27 (2000).

[21] P.J. Nixon, I. Sims, eds., RILEM Recommendations for the Prevention of Damage by Alkali-Aggregate Reactions in New Concrete Structures, Springer Netherlands, Dordrecht, 2016. https://doi.org/10.1007/978-94-017-7252-5.

[22] A. Leemann, M. Góra, B. Lothenbach, M. Heuberger, Alkali silica reaction in concrete - Revealing the expansion mechanism by surface force measurements, Cement and Concrete Research 176 (2024) 107392. https://doi.org/10.1016/j.cemconres.2023.107392.

[23] M.G. Alexander, Alkali–aggregate reaction, in: Developments in the Formulation and Reinforcement of Concrete, Elsevier, 2019: pp. 87–113. https://doi.org/10.1016/B978-0-08-102616-8.00004-6.

[24] P.E. Grattan-Bellew, T. Katayama, So-Called Alkali-Carbonate Reaction (ACR), in: Alkali-Aggregate Reaction in Concrete, CRC Press, 2017: p. 26. https://doi.org/10.1201/9781315708959.

[25] M. Alexander, Alkali-Aggregate Reaction and Structural Damage to Concrete: Engineering Assessment, Repair and Management, CRC Press, 2011. https://doi.org/10.1201/b10773.

[26] L.F.M. Sanchez, B. Fournier, M. Jolin, J. Duchesne, Reliable quantification of AAR damage through assessment of the Damage Rating Index (DRI), Cement and Concrete Research 67 (2015) 74–92. https://doi.org/10.1016/j.cemconres.2014.08.002.

[27] ASTM C295/C295M, Guide for Petrographic Examination of Aggregates for Concrete, ASTM International, 2019. https://doi.org/10.1520/C0295_C0295M-19.

[28] B. Fournier, V. Malhotra, Evaluation of Laboratory Test Methods for Alkali-Silica Reactivity, Cement, Concrete, Aggr. 21 (1999) 173. https://doi.org/10.1520/CCA10431J.

[29] A.B. Poole, I. Sims, D.A. St John, Concrete petrography: a handbook of investigative techniques, Second edition, CRC Press, Boca Raton, 2019.

[30] C09 Committee, ASTM C856-20 Practice for Petrographic Examination of Hardened Concrete, (2020). https://doi.org/10.1520/C0856_C0856M-20.

[31] I. Fernandes, M. dos A. Ribeiro, H. Martins, M. Broekmans, I. Sims, P. Nixon, F. Noronha, Assessment of Concrete Aggregate for ASR Potential by Petrography. The Work Developed by RILEM TC-ACS (2007-2013), ENGINEERING GEOLOGY FOR SOCIETY AND TERRITORY, VOL 5: URBAN GEOLOGY, SUSTAINABLE PLANNING AND LANDSCAPE EXPLOITATION (2015) 37–40. https://doi.org/10.1007/978-3-319-09048-1_6.

[32] I. Fernandes, M.D.A. Ribeiro, M.A.T.M. Broekmans, I. Sims, eds., Petrographic Atlas: Characterisation of Aggregates Regarding Potential Reactivity to Alkalis, Springer Netherlands, Dordrecht, 2016. https://doi.org/10.1007/978-94-017-7383-6.

[33] I. Sims, P. Nixon, RILEM Recommended Test Method AAR-0: Detection of Alkali-Reactivity Potential in Concrete—Outline guide to the use of RILEM methods in assessments of aggregates for potential alkali-reactivity, Mat. Struct. 36 (2003) 472–479. https://doi.org/10.1007/BF02481527.

[34] CSA A23.2-27A, Standard Practices to identify degree of alkali-reactivity of aggregates and to identify measures to avoid deleterious expansion in concrete., in: CSA Standards A23.1–09/A23.2–09 Concrete Materials and Methods of Concrete Construction/Test Methods and Standard Practices for Concrete, 13th edition, CSA Group, Ontario, Canada, 2019: pp. 594–610.

[35] ASTM C289-07 Standard Test Method for Potential Alkali-Silica Reactivity of Aggregates (Chemical Method) (Withdrawn 2016), (2007).

[36] C09 Committee, ASTM C227-10 Test Method for Potential Alkali Reactivity of Cement-Aggregate Combinations (Mortar-Bar Method), (2010). https://doi.org/10.1520/C0227-10.

[37] ASTM C33-23 Specification for Concrete Aggregates, (2023). http://www.astm.org/cgi-bin/resolver.cgi?C33-03 (accessed August 6, 2024).

[38] M.-A. Bérubé, J. Frenette, Testing Concrete for AAR in NaOH and NaCl solutions at 38°C and 80°C, Cement and Concrete Composites 16 (1994) 189–198. https://doi.org/10.1016/0958-9465(94)90016-7.

[39] ASTM C1260, Test Method for Potential Alkali Reactivity of Aggregates (Mortar-Bar Method), ASTM International, 2022. https://doi.org/10.1520/C1260-22.

[40] R.D. Hooton, C.A. Rogers, Development of the NBRI rapid mortar bar test leading to its use in North America, Construction and Building Materials 7 (1993) 145–148. https://doi.org/10.1016/0950-0618(93)90051-D.

[41] ASTM C1778, Guide for Reducing the Risk of Deleterious Alkali-Aggregate Reaction in Concrete, ASTM International, 2022. https://doi.org/10.1520/C1778-22.

[42] B. Fournier, J.H. Ideker, K.J. Folliard, M.D.A. Thomas, P.-C. Nkinamubanzi, R. Chevrier, Effect of environmental conditions on expansion in concrete due to alkali–silica reaction (ASR), Materials Characterization 60 (2009) 669–679. https://doi.org/10.1016/j.matchar.2008.12.018.

[43] M. Thomas, B. Fournier, K. Folliard, J. Ideker, M. Shehata, Test methods for evaluating preventive measures for controlling expansion due to alkali–silica reaction in concrete, Cement and Concrete Research 36 (2006) 1842–1856. https://doi.org/10.1016/j.cemconres.2006.01.014.

[44] A. Shayan, H. Morris, A comparison of RTA T363 and ASTM C1260 accelerated mortar bar test methods for detecting reactive aggregates, Cement and Concrete Research 31 (2001) 655–663. https://doi.org/10.1016/S0008-8846(00)00491-9.

[45] S. Multon, M. Cyr, A. Sellier, P. Diederich, L. Petit, Effects of aggregate size and alkali content on ASR expansion, Cement and Concrete Research 40 (2010) 508–516. https://doi.org/10.1016/j.cemconres.2009.08.002.

[46] B. Lothenbach, K. De Weerdt, D. Hooton, J. Duchesne, A. Leemann, Can We Relate ASR Expansion to the Pore Solution Composition?, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 104–112. https://doi.org/10.1007/978-3-031-59419-9_13.

[47] B. Fournier, M.A. Bérubé, Application of the NBRI accelerated mortar bar test to siliceous carbonate aggregates produced in the St. Lawrence Lowlands (Quebec, Canada) part 2: Proposed limits, rates of expansion, and microstructure of reaction products, Cement and Concrete Research 21 (1991) 1069–1082. https://doi.org/10.1016/0008-8846(91)90067-R.

[48] J. Lindgård, B. Fournier, T.F. Rønning, M.D.A. Thomas, Alkali–aggregate reaction: performance testing, exposure sites and regulations, Proceedings of the Institution of Civil Engineers - Construction Materials 169 (2016) 189–196. https://doi.org/10.1680/jcoma.15.00077.

[49] ASTM C1293, Test Method for Determination of Length Change of Concrete Due to Alkali-Silica Reaction, (2024). https://doi.org/10.1520/C1293_C1293M-23A.

[50] J.H. Ideker, B.L. East, K.J. Folliard, M.D.A. Thomas, B. Fournier, The current state of the accelerated concrete prism test, Cement and Concrete Research 40 (2010) 550–555. https://doi.org/10.1016/j.cemconres.2009.08.030.

[51] J. Lindgård, T. Østnor, B. Fournier, Ø. Lindgård, T. Danner, G. Plusquellec, K. De Weerdt, Determining alkali leaching during accelerated ASR performance testing and in field exposed cubes using cold water extraction (CWE) and μXRF, MATEC Web Conf. 199 (2018) 03004. https://doi.org/10.1051/matecconf/201819903004.

[52] B. Fournier, R. Chevrier, A. Bilodeau, P.-C. Nkinamubanzi, N. Bouzoubaa, COMPARATIVE FIELD AND LABORATORY INVESTIGATIONS ON THE USE OF SUPPLEMENTARY CEMENTING MATERIALS (SCMs) TO CONTROL ALKALI-SILICA REACTION (ASR) IN CONCRETE, in: Brazil, 2016.

[53] AASHTO T380. (2019) Standard Method of Test for Potential Alkali Reactivity of Aggregates and Effectiveness of ASR Mitigation Measures (Miniature Concrete Prism Test, MCPT)., (2019).

[54] A. Ghanizadeh, M. Thomas, T. Drimalas, K.S. Teja, A. Parashar, J.H. Ideker, R. Lute, K. Folliard, Using the Miniature Concrete Prism Test (MCPT) to evaluate ASR preventive measures, (2022).

[55] J. Tanesi, T. Drimalas, K.S.T. Chopperla, M. Beyene, J.H. Ideker, H. Kim, L. Montanari, A. Ardani, Divergence between Performance in the Field and Laboratory Test Results for Alkali-Silica Reaction, Transportation Research Record 2674 (2020) 120–134. https://doi.org/10.1177/0361198120913288.

[56] T.F. Rønning, J. Lindgård, I. Borchers, ASR performance testing concepts – RILEM AAR-10, AAR-11 and AAR-12, (2022).

[57] T.F. Ronning, B.J. Wigum, J. Lindgard, Recommendation of RILEM TC 258-AAA: RILEM AAR-10: determination of binder combinations for non-reactive mix design using concrete prisms-38 degrees C test method, MATERIALS AND STRUCTURES 54 (2021). https://doi.org/10.1617/s11527-021-01679-w.

[58] J. Custódio, J. Lindgård, B. Fournier, A. Santos Silva, M.D.A. Thomas, T. Drimalas, J.H. Ideker, R.-P. Martin, I. Borchers, B. Johannes Wigum, T.F. Rønning, Correlating field and laboratory investigations for preventing ASR in concrete – The LNEC cube study (Part I – Project plan and laboratory results), Construction and Building Materials 343 (2022) 128131. https://doi.org/10.1016/j.conbuildmat.2022.128131.

[59] Z. Xu, X. Lan, M. Deng, M. Tang, A new accelerated method for determining the potential alkali-carbonate reactivity, Cement and Concrete Research 32 (2002) 851–857. https://doi.org/10.1016/S0008-8846(01)00758-X.

[60] R. Bleszynski, R.D. Hooton, M.D.A. Thomas, C. Rogers, Durability of Ternary Blend Concrete with Silica Fume and Blast-Furnace Slag: Laboratory and Outdoor Exposure Site Studies, MJ 99 (2002). https://doi.org/10.14359/12329.

[61] K.S.T. Chopperla, T. Drimalas, M. Beyene, J. Tanesi, K. Folliard, A. Ardani, J.H. Ideker, Combining reliable performance testing and binder properties to determine preventive measures for alkali-silica reaction, Cement and Concrete Research 151 (2022) 106641. https://doi.org/10.1016/j.cemconres.2021.106641.

[62] I. Fernandes, A. Leemann, B. Fournier, E. Menendez, J. Lindgård, I. Borchers, J. Custódio, PARTNER project post-documentation study. Condition assessment of field exposure site

cubes. Results of microstructural analyses, Cement and Concrete Research 162 (2022) 107006. https://doi.org/10.1016/j.cemconres.2022.107006.

[63] G. Giaccio, M.C. Torrijos, C. Milanesi, R. Zerbino, Alkali–silica reaction in plain and fibre concretes in field conditions, Mater Struct 52 (2019) 31. https://doi.org/10.1617/s11527-019-1332-2.

[64] E.R. Giannini, A.F. Bentivegna, K.J. Folliard, Coatings and overlays for concrete affected by alkali-silica reaction, in: Dresden, Germany, 2011.

[65] S. Hayman, M. Thomas, N. Beaman, P. Gilks, Selection of an effective ASR-prevention strategy for use with a highly reactive aggregate for the reconstruction of concrete structures at Mactaquac generating station, Cement and Concrete Research 40 (2010) 605–610. https://doi.org/10.1016/j.cemconres.2009.08.015.

[66] R.D. Hooton, C. Rogers, C.A. MacDonald, T. Ramlochan, Twenty-Year Field Evaluation of Alkali-Silica Reaction Mitigation, MJ 110 (2013). https://doi.org/10.14359/51685905.

[67] J.H. Ideker, A.F. Bentivegna, K.J. Folliard, M.C.G. Juenger, Do Current Laboratory Test Methods Accurately Predict Alkali-Silica Reactivity?, MJ 109 (2012). https://doi.org/10.14359/51683914.

[68] Y. Kawabata, K. Yamada, Y. Sagawa, S. Ogawa, Alkali-Wrapped Concrete Prism Test (AW-CPT) – New Testing Protocol Toward a Performance Test against Alkali-Silica Reaction–, ACT 16 (2018) 441–460. https://doi.org/10.3151/jact.16.441.

[69] Y. Kawabata, K. Yamada, T. Kawakami, Y. Sagawa, Environmental impacts on expansion of concrete due to alkali–silica reaction, Magazine of Concrete Research (2022) 1–11. https://doi.org/10.1680/jmacr.22.00158.

[70] I. Robertson, L. Shen, Field Evaluation of Concrete using Hawaiian Aggregates for Alkali Silica Reaction, MATEC Web Conf. 199 (2018) 03005. https://doi.org/10.1051/matecconf/201819903005.

[71] N. Smaoui, M.-A. Bérubé, B. Fournier, B. Bissonnette, B. Durand, Evaluation of the expansion attained to date by concrete affected by alkali–silica reaction. Part I: Experimental study, Can. J. Civ. Eng. 31 (2004) 826–845. https://doi.org/10.1139/l04-051.

[72] N. Smaoui, B. Fournier, M.-A. Bérubé, B. Bissonnette, B. Durand, Evaluation of the expansion attained to date by concrete affected by alkalisilica reaction. Part II: Application

to nonreinforced concrete specimens exposed outside, Can. J. Civ. Eng. 31 (2004) 997–1011. https://doi.org/10.1139/l04-074.

[73] M.D.A. Thomas, B. Fournier, K.J. Folliard, M.H. Shehata, J.H. Ideker, C. Rogers, Performance Limits for Evaluating Supplementary Cementing Materials Using Accelerated Mortar Bar Test, MJ 104 (2007). https://doi.org/10.14359/18573.

[74] S. Yang, J.-X. Lu, C.S. Poon, Recycling of waste glass in dry-mixed concrete blocks: Evaluation of alkali-silica reaction (ASR) by accelerated laboratory tests and long-term field monitoring, Construction and Building Materials 262 (2020) 120865. https://doi.org/10.1016/j.conbuildmat.2020.120865.

[75] C.A. MacDonald, C. Rogers, R.D. Hooton, THE RELATIONSHIP BETWEEN LABORATORY AND FIELD EXPANSION – OBSERVATIONS AT THE KINGSTON OUTDOOR EXPOSURE SITE FOR ASR AFTER TWENTY YEARS, in: Austin, Texas, USA, 2012.

[76] C.A. MacDonald, Kingston Outdoor Exposure Site for ASR –29 Year Update, (2020).

[77] C.A. MacDonald, The Kingston Exposure Site for ASR at 32 Years, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 540–549. https://doi.org/10.1007/978-3-031-59349-9_62.

[78] B.J. Wigum, G.J. Einarsson, ALKALI AGGREGATE REACTION IN ICELAND RESULTS FROM LABORATORY TESTING COMPARED TO FIELD EXPOSURE SITE, (n.d.).

[79] M.D.A. Thomas, SUMMARY OF BRE RESEARCH ON THE EFFECT OF FLY ASH ON AIKALI-SILICA REACTION IN CONCRETE, in: Melbourne, Australia, 1996.

[80] M. Thomas, A. Dunster, P. Nixon, B. Blackwell, Effect of fly ash on the expansion of concrete due to alkali-silica reaction – Exposure site studies, Cement and Concrete Composites 33 (2011) 359–367. https://doi.org/10.1016/j.cemconcomp.2010.11.006.

[81] B. Fournier, A. Bilodeau, N. Bouzoubaa, P.-C. Nkinamubanzi, Field and Laboratory Investigations on the Use of Fly Ash and Li-Based Admixtures to Prevent ASR in Concrete, in: United Kingdom, 2018.

[82] B. Fournier, A. Ferro, V. Sivasundaram, CANMET/Industry Research Consortium on Alkali-Aggregate Reactivity in Concrete, EPRI and CANMET, Ontario, Canada, 2001. https://www.epri.com/research/products/1004031.

[83] CSA A23.2-28A, Standard Practice for laboratory testing to demonstrate the effectiveness of supplementary cementitious materials and lithium-based admixtures to prevent alkali-silica reaction in concrete.pdf, in: CSA Standards A23.1–09/A23.2–09 Concrete Materials and Methods of Concrete Construction/Test Methods and Standard Practices for Concrete, 13th edition, CSA Group, Ontario, Canada, 2019: pp. 611–618.

[84] D. Hooton, M. Thomas, T. Ramlochan, R. Bleszynski, Durability of Ternary Blend Concrete with Silica Fume and Blast-Furnace Slag: Laboratory and Outdoor Exposure Site Studies, in: Trondheim, Norway, 2008. https://doi.org/10.14359/12329.

[85] I. Borchers, C. Müller, Field Site Tests Established in the Partner Project for Evaluating the Correlation Between Laboratory Tests and Field Performance.PDF, in: Trondheim, Norway, 2008.

[86] J. Lindgård, P.J. Nixon, I. Borchers, B. Schouenborg, B.J. Wigum, M. Haugen, U. Åkesson, The EU "PARTNER" Project — European standard tests to prevent alkali reactions in aggregates: Final results and recommendations, Cement and Concrete Research 40 (2010) 611–635. https://doi.org/10.1016/j.cemconres.2009.09.004.

[87] I. Fernandes, M.A.T.M. Broekmans, P. Nixon, I. Sims, M. dos A. Ribeiro, F. Noronha, B. Wigum, Alkali-silica reactivity of some common rock types. A global petrographic atlas, QUARTERLY JOURNAL OF ENGINEERING GEOLOGY AND HYDROGEOLOGY 46 (2013) 215–220. https://doi.org/10.1144/qjegh2012-065.

[88] T. Drimalas, K.J. Folliard, J.H. Ideker, Findings from the University of Texas at Austin ASR Exposure Site after 20 Years, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 516–523. https://doi.org/10.1007/978-3-031-59349-9_59.

[89] K.J. Folliard, R. Barborak, T. Drimalas, L. Du, S. Garber, J. Ideker, T. Ley, S. Williams, M. Juenger, B. Fournier, M.D.A. Thomas, Preventing ASR/DEF in New Concrete: Final Report (0-4085-5), Texas Department of Transportation and the Federal Highway Administration, 2006.

[90] J. Lindgård, T.F. Rønning, M.D.A. Thomas, B. Fournier, A.S. Silva, ASR - PERFORMANCE TESTING: MAIN FINDINGS IN THE NORWEGIAN COIN PROJECT, in: Sao Paulo – Brazil, 2016.

[91] J. Lindgard, M.D.A. Thomas, E.J. Sellevold, B. Pedersen, O. Andic-Cakir, H. Justnes, T.F. Ronning, Alkali-silica reaction (ASR)-performance testing: Influence of specimen pre-treatment, exposure conditions and prism size on alkali leaching and prism expansion, CEMENT AND CONCRETE RESEARCH 53 (2013) 68–90. https://doi.org/10.1016/j.cemconres.2013.05.017.

[92] S.U. Einarsdóttir, Modifications to Laboratory Test Methods to Evaluate the Beneficial Effects of Low-Alkali Binders on the Alkali-Silica Reaction, University of Toronto, 2017.

[93] M.D.A. Thomas, K.J. Folliard, F. Benoit, T. Drimalas, S. Garber, Methods for Preventing ASR in New Construction: Results of Field Exposure Sites, FHWA Office of Pavement Technology, 2013.

[94] C. Pelissou, B. Durville, S. Morin, Main Outcomes for AAR Studies on Large Scale Experiments in the ODOBA International Project, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 566–576. https://doi.org/10.1007/978-3-031-59349-9_65.

[95] K.K. Hansen, B. Grelk, R. Kofoed, J.B. Mørk, Field Investigation of Concrete Blocks with Different ASR Reactive Aggregate Types with and Without Surface Hydrophobic Impregnation, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 550–557. https://doi.org/10.1007/978-3-031-59419-9_64.

[96] A. Leemann, D. Sirtoli, J. Justs, M. Wyrzykowski, Establishing Two ASR Exposure Sites in Switzerland – Approach and Goals, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 532–539. https://doi.org/10.1007/978-3-031-59349-9_61.

[97] A.H.-S. Ang, W.H. Tang, A.H.-S. Ang, Probability concepts in engineering: emphasis on applications in civil & environmental engineering, 2nd ed, Wiley, New York, 2007.

[98] K.M. Ramachandran, C.P. Tsokos, Descriptive Statistics, in: Mathematical Statistics with Applications in R, Elsevier, 2015: pp. 1–52. https://doi.org/10.1016/B978-0-12-417113-8.00001-1.

[99] O.C. Ibe, Introduction to Descriptive Statistics, in: Fundamentals of Applied Probability and Random Processes, Elsevier, 2014: pp. 253–274. https://doi.org/10.1016/B978-0-12-800852-2.00008-0.

[100] B.R. Martin, Statistics, Experiments, and Data, in: Statistics for Physical Science, Elsevier, 2012: pp. 1–20. https://doi.org/10.1016/B978-0-12-387760-4.00001-9.

[101] O.C. Ibe, Introduction to Inferential Statistics, in: Fundamentals of Applied Probability and Random Processes, Elsevier, 2014: pp. 275–305. https://doi.org/10.1016/B978-0-12-800852-2.00009-2.

[102] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, Bayesian Data Analysis Third edition (with errors fixed as of 13 February 2020), (n.d.).

[103] LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S, Phil. Trans. R. Soc. 53 (1763) 370–418. https://doi.org/10.1098/rstl.1763.0053.

[104] S. Sinharay, Continuous Probability Distributions, (n.d.).

[105] M. Shaw, Use of Bayes' theorem and the beta distribution for reliability estimation purposes, Reliability Engineering & System Safety 31 (1991) 145–153. https://doi.org/10.1016/0951-8320(91)90115-N.

[106] T.M. Mitchell, Machine Learning, McGraw-Hill, New York, 1997.

[107] L. Zou, Meta-learning basics and background, in: Meta-Learning, Elsevier, 2023: pp. 1–22. https://doi.org/10.1016/B978-0-323-89931-4.00010-9.

[108] K. Yeturu, Machine learning algorithms, applications, and practices in data science, in: Handbook of Statistics, Elsevier, 2020: pp. 81–206. https://doi.org/10.1016/bs.host.2020.01.002.

[109] H. Belyadi, A. Haghighat, Supervised learning, in: Machine Learning Guide for Oil and Gas Using Python, Elsevier, 2021: pp. 169–295. https://doi.org/10.1016/B978-0-12-821929-4.00004-4.

Chapter 3: Research program

Once AAR is triggered, remediation becomes complex and uncertain, making prevention the most effective strategy to avoid such deterioration. Although accelerated laboratory tests like the Accelerated Mortar Bar Test (AMBT) and Concrete Prism Test (CPT) are widely used, their predictive accuracy has been questioned due to discrepancies between laboratory results and actual field performance. Therefore, by critically evaluating current methods and incorporating advanced analytical techniques, this research offers a structured, dynamic approach to improving AAR prediction in concrete structures. By combining advanced data analysis techniques with a detailed understanding of laboratory-to-field discrepancies, this study contributes to more informed decision-making for AAR mitigation strategies.

The research is organized into three distinct phases. Phase 1 focuses on Paper 1, which involves conducting a detailed bibliometric analysis and establishing a comprehensive database that combines laboratory test results with field performance. This phase aims to systematically identify key contributors, collaborations, and trends in AAR testing and field studies.

In Phase 2, Paper 2 investigates the reliability of laboratory testing methods, such as the AMBT and CPT, in predicting AAR occurrence in the field. This phase critically examines the variability of these test methods over time by introducing the K coefficient, defined as the ratio of long-term field performance expansion (R_{lt}) to the accelerated laboratory test expansion (R_{at}) at a specific time. The K coefficient enables comparative analysis to evaluate how effectively these accelerated tests predict long-term behavior under actual field conditions. Building on this, Paper 3 applies probabilistic modelling to address uncertainties in laboratory test data, aiming to identify the risks associated with relying on established tests to predict field performance.

Phase 3 addresses Papers 4 and 5, which use advanced analytical techniques to enhance the predictive accuracy of AAR testing methods. In Paper 4, machine learning approaches, including decision trees and logistic regression models, propose a flexible coupled threshold-time (FCTT) approach that assesses AAR thresholds dynamically and probabilistically. This approach integrates variables such as expansion levels, test duration, environmental conditions, and alkali concentration to better predict AAR occurrence. Concluding the findings, Paper 5 presents a probabilistic framework that integrates lab and field data using logistic regression to assess ASR risk. This framework incorporates environmental factors, alkali levels, and structural classifications to determine the likelihood of AAR in field conditions.

Figure 3. 1 illustrates the core concepts and methodologies applied across the papers, providing a cohesive overview of the research program. This structure establishes a systematic progression of topics, clearly outlining the analytical steps and methodologies employed throughout the study.

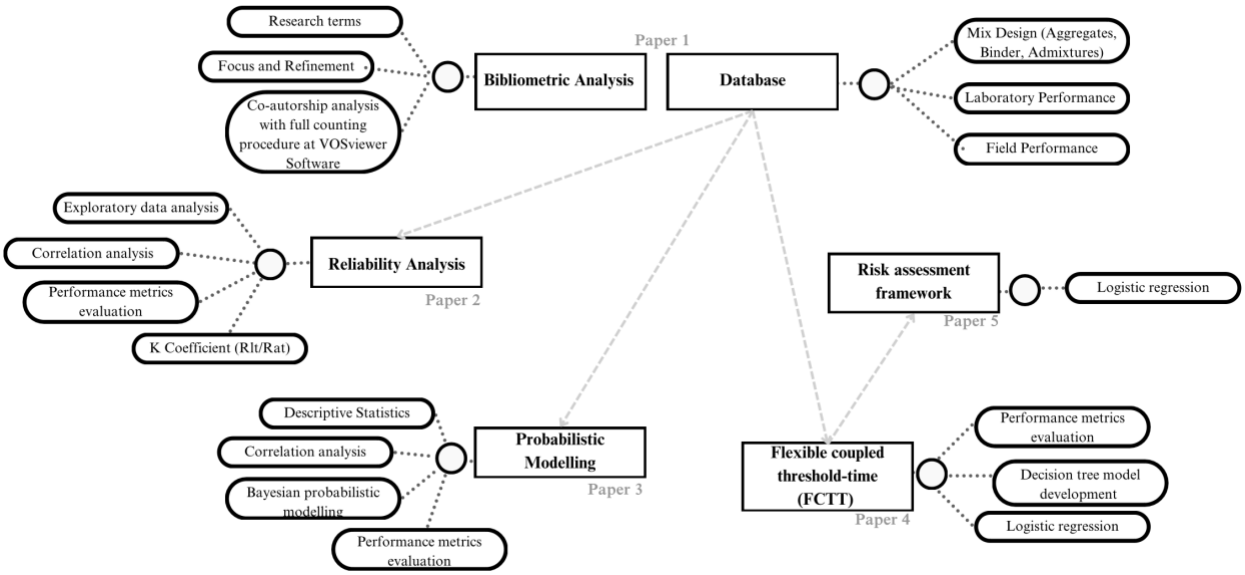


Figure 3. 1 – Overview of the research program.

3.1 Bibliometric Analysis: Research terms, focus, refinement and network analysis

To thoroughly evaluate discrepancies in laboratory testing for assessing AAR in field structures, a comprehensive bibliometric analysis was conducted. This methodology systematically collects, organizes, and analyzes existing literature to uncover research trends and assess their relationships [1].

The process begins by defining the purpose and scope of the bibliometric study, with a focus on key studies on AAR-induced deterioration in field structures, particularly at exposure sites. The analysis centers on literature correlating laboratory tests with field performance, emphasizing studies of newly constructed test structures (i.e., concrete blocks) over studies of established infrastructure like dams and bridges.

An extensive search strategy was developed to ensure relevance and accuracy. A set of keywords was selected, including “Alkali,” “silica,” “carbonate,” “aggregate,” “reaction,” “efficiency,” “combining,” “correlating,” “evaluation,” “divergence,” “prediction,” “monitoring,” “performance,” “comparing,” “comparison,” “field,” “in-situ,” “blocks,” “sites,” “exposed,”

"laboratory," "tests," and "results." Recognized databases such as Web of Science were utilized as primary sources. Optimization searching tools, such as Boolean operators (i.e., AND, OR, NOT), were applied to streamline search results.

After the initial search, a refinement step was implemented. Citation topics were limited to "concrete science," "mechanics," "testing and maintenance," "computer vision and graphics," "sustainability science," "management," and "geotechnical engineering," ensuring that irrelevant topics were excluded from the analysis. The final data collection step involved reviewing all abstracts to ensure only publications covering both laboratory and field performance were included, thereby eliminating information not within the research scope.

After gathering the bibliometric data, VOSviewer software was used to conduct a networking analysis through co-authorship clustering, visually mapping the connections among researchers and highlighting trends over time. This networking analysis enables an in-depth look at collaboration patterns, illustrating how research communities within AAR studies are organized and connected. A quantitative assessment was also carried out, analyzing the volume of co-authored documents per author and per country to gain insights into the geographic and institutional contexts in which current field studies are conducted. In this approach, all co-authorship links were assigned equal weight as part of a full counting procedure, ensuring a balanced representation of each collaboration within the network. This analysis ultimately provides a clearer picture of the global research landscape, including key contributors and potential gaps in collaboration across regions.

3.2 Database collection and structuring

To deepen the understanding of correlations between laboratory and field performance in AAR assessments, a comprehensive database was created [2]. This database, developed from a range of published sources, includes experimental data from five major projects involving accelerated laboratory tests conducted under various setups and field-exposed concrete members subjected to diverse environmental conditions [3–7]. Key information is organized into categories covering material selection (e.g., types of aggregates, cement, supplementary cementitious materials (SCMs), and admixtures) and mix proportions.

In addition to material and mix details, the database provides a detailed account of sample characteristics for both laboratory and field tests, such as surface area, volume, test setup, and

environmental conditions. Performance data, specifically expansion over time, is included for each sample, allowing for a direct comparison between lab and field outcomes. This comprehensive dataset integrates laboratory results with field performance data, creating a foundation for advanced analysis aimed at improving the predictive power of accelerated laboratory tests regarding AAR in new concrete structures.

3.3 Descriptive statistics

Understanding uncertainty is critical in engineering, as it arises both from natural variability (i.e., aleatory uncertainty) and from limitations in knowledge or models (i.e., epistemic uncertainty) [8]. Descriptive statistics help quantify these uncertainties, allowing researchers to more accurately assess the likelihood of various outcomes in AAR-related field performance. By systematically summarizing and visualizing data, descriptive statistics support further analytical steps, guiding inferential statistical methods that extend beyond simple data description to draw conclusions, predict trends, and model behavior under real-world conditions.

In this study, descriptive statistics, including mean, median, standard deviation, variance, and range, were calculated to provide insights into the central tendency, spread, and distribution shape (i.e., skewness) of the data [9,10]. These measures give an overview of how data points, such as expansion rates, are distributed across different experimental setups and environmental conditions. Histograms were used to visually represent each variable's distribution within subsets of the dataset, allowing easy identification of data patterns or outliers [11].

3.4 Correlation analysis

The comparative analysis method employs linear correlation, specifically using Pearson correlation coefficients, to evaluate the strength and direction of relationships between key variables in laboratory and field test results. This approach starts by constructing a correlation matrix, which visually organizes the relationships between different variables (i.e., laboratory expansion, temperature, alkali content, and relative humidity) across various testing setups. By doing so, the matrix identifies which factors are most strongly associated with field reactivity outcomes, enabling an understanding of the primary drivers of AAR in different scenarios.

To deepen the analysis, scatter plots provide a visual comparison of laboratory test results against field expansions, offering insight into the spread and alignment of data points. Linear regression further supports this analysis by quantifying the strength of the association between laboratory and field test outcomes, producing correlation coefficients that indicate the degree of predictive alignment.

3.5 Performance metrics evaluation

When evaluating the performance of laboratory tests like AMBT and CPT in predicting AAR outcomes, performance metrics provide essential insights. Accuracy, for instance, assesses the overall correctness of the test, calculated as the proportion of correct predictions (i.e., true positives and true negatives) relative to the total number of predictions. This metric indicates how often lab test results align with actual field outcomes, making it valuable measure of test reliability.

Precision, which focuses specifically on the accuracy of positive predictions, is calculated as the ratio of true positives (i.e., correctly identified reactive cases) to the sum of true positives and false positives. High precision is crucial in contexts where minimizing false positives is important, as it indicates that most positive predictions are correct. Sensitivity, or recall, evaluates the test's ability to identify reactive cases correctly. It compares the number of true positives to the total of true positives and false negatives, ensuring that the test detects reactivity effectively, which is especially important when missing reactive cases could lead to structural risks. Finally, the F1-score, the harmonic mean of precision and recall, provides a balanced view, accounting for both false positives and false negatives. This score is particularly useful when there is an imbalance in class distribution or when both precision and sensitivity are priorities.

Improving test predictive performance requires reducing false negatives, thereby increasing sensitivity, while also accurately identifying non-reactive cases to enhance specificity. This balanced approach helps to reliably detect reactivity without excessively raising false positives, resulting in a more effective predictive model for AAR risk.

3.6 K coefficient

The K coefficient serves as a valuable metric for analyzing the equivalency between accelerated laboratory test outcomes and field performance over time. Defined as the ratio of long-term field

expansion (R_{lt}) to the accelerated laboratory test expansion (R_{at}) at a specific point in time, the K coefficient provides insight into how well accelerated tests like AMBT and CPT predict actual long-term behavior in field conditions:

$$K = \frac{R_{lt}}{R_{at}}$$

The primary purpose of the coefficient K is to establish a relationship between the accelerated test results and the actual field performance. By calculating K, a first indication of the validity of accelerated tests as predictors of long-term behaviour can be concluded.

- K>1: Indicates that the field expansion is greater than the laboratory expansion. This could suggest that the laboratory test underpredicts the actual field performance.
- K<1: Indicates that the field expansion is less than the laboratory expansion. This could suggest that the laboratory test overpredicts the actual field performance.
- K=1: Indicates a perfect correlation where the accelerated laboratory expansion matches the field performance expansion.

3.7 Bayesian probabilistic modelling

Bayesian inference is generally employed to fit a probability model to a set of data, summarizing results as probability distributions on parameters [12,13]. The Bayesian data analysis process involves three key steps: (i) setting up a full probability model (i.e., prior distribution); (ii) conditioning on observed data (i.e., likelihood function acting on the posterior distribution); and, (iii) evaluating the fit of the model [12]. In the current context, Bayes' Theorem [13] was utilized to identify the probability of the exposed concrete members being reactive giving the current test outcomes (i.e., AMBT and CPT) and thresholds [14].

In the current study, Bayesian inference was based on a data-driven prior derived from the observed data. The likelihood function was calculated by dividing the number of reactive samples by the total number of reactive samples, while the marginal likelihood represented the ratio of total samples in each category to the total number of samples. The posterior probability was then computed using Bayes' theorem, incorporating the prior, likelihood, and marginal likelihood. To account for uncertainty, confidence intervals around the posterior probability were estimated using

a neutral or non-informative beta distribution (i.e., Beta(1,1)) based on the observed counts of reactive and non-reactive blocks, as seen in Figure 3. 2.

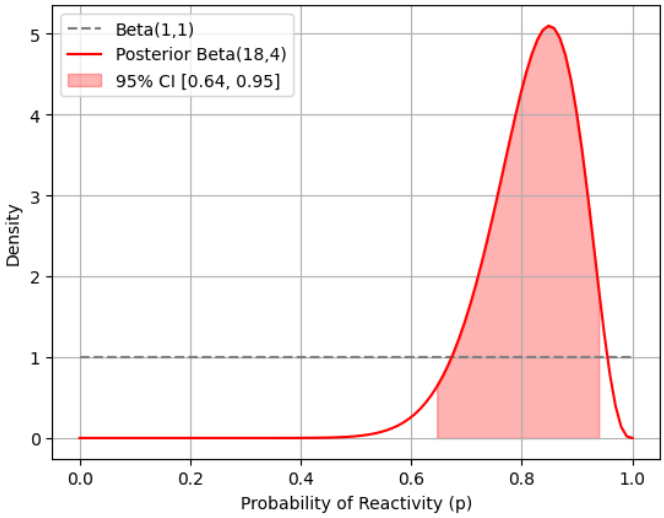


Figure 3. 2 – Beta and posterior distribution with CI.

Prior probability

The prior probability $P(R)$ represents the initial belief about the probability that the dependent variable, i.e., whether the block is reactive (i.e., YES | NO), given the conditional variables (i.e., level of reactivity, environment and alkali content of the binder). As seen in Equation 3.1, this probability is calculated as the total number of reactive blocks ($N_{reactive}$) divided by the total number of blocks (N_{total}).

$$P(R) = \frac{N_{reactive}}{N_{total}} \quad \text{(Equation 3. 1)}$$

Likelihood function

The likelihood function $P(L_i|R)$ incorporates the probability of the observed data given the conditional parameters. In the current case, the likelihood is expressed as the number of reactive blocks at a specific condition ($N_{L_i}^{reactive}$) divided by the total number of reactive blocks Equation 3.2. This approach provides a simple estimation of the likelihood based on the collected data.

$$P(L_i|R) = \frac{N_{L_i}^{reactive}}{N_{reactive}}$$ (Equation 3. 2)

Marginal likelihood and posterior probability

The marginal likelihood $P(L_i)$, representing the overall probability of observing a specific condition (L_i) is assessed as indicated in Equation 3.3, by the division of the total number of blocks at a specific condition ($N_{L_i}^{total}$) over the total number of blocks (N_{total}). The marginal likelihood normalizes the posterior distribution, ensuring that the total probability is equal to 1. By combining the prior and likelihood using Bayes' Theorem, the posterior probability (($P(R|L_i)$) is then computed, the posterior probability represents the probability of a block being reactive given the observed condition as per Equation 3.4.

$$P(L_i) = \frac{N_{L_i}^{total}}{N_{total}}$$ (Equation 3. 3)

$$P(R|L_i) = \frac{P(L_i|R) \cdot P(R)}{P(L_i)} = \frac{N_{L_i}^{reactive}}{N_{L_i}^{total}}$$ (Equation 3. 4)

Beta distribution and Confidence intervals

After observing the data, the parameters of a neutral and uniform Beta (i.e., Beta(1,1)) distribution are updated to model the posterior distribution [15]. Then, parameters α' and β', derived from the observed data are equivalent to the number of reactive and non-reactive blocks at a specific condition ($N_{L_i}^{reactive} + 1;\ N_{L_i}^{non-reactive} + 1$) as represented per Equation 3.5 [16]. This posterior Beta distribution now models the probability of reactivity after observing the data. The probability density function (PDF) of the Beta distribution is represented in Equation 3.6, and the Beta function is defined in Equation 6.7.

$$Beta\ (\alpha', \beta') = Beta\ (N_{L_i}^{reactive} + 1, N_{L_i}^{non-reactive} + 1)$$ (Equation 3. 5)

$$f(t;\ \alpha, \beta) = \frac{1}{B(\alpha, \beta)} t^{\alpha-1}(1-t)^{\beta-1}$$ (Equation 3. 6)

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt \quad \text{(Equation 3. 7)}$$

Finally, the cumulative distribution function (CDF) of the Beta distribution gives the probability that a Beta-distributed random variable is less than or equal to a given value *x*. The inverse of CDF, known as the percentile point function (PPF) or the quantile function, allows for the computation of confidence intervals. For a 95% confidence interval, the values required are at 2.5th and 97.5% percentile of the distribution. Therefore, the lower and upper bounds are given as per Equation 3.8 and Equation 3.9, respectively.

$$CI_{lower} = Beta^{-1}(0.025; \alpha, \beta) \quad \text{(Equation 3. 8)}$$

$$CI_{upper} = Beta^{-1}(0.975; \alpha, \beta) \quad \text{(Equation 3. 9)}$$

In summary, this Bayesian approach, incorporating the Beta distribution allows for probabilistic modelling that relies on observed data and includes uncertainty quantification.

Model performance evaluation

Brier score measures how close the predicted probabilities are to the actual binary outcomes [17]. It ranges from 0 to 1, with 0 being the ideal score. The Brier score penalizes both the magnitude of prediction errors and how well-calibrated the model is. Therefore, considering p_i as the predicted probability and y_i the actual binary outcome (i.e., 1 for field reactive and 0 for non-reactive), Equation 3.10 indicates the calculations for Brier score.

$$Brier\ Score = \frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2 \quad \text{(Equation 3. 10)}$$

In this sense, when evaluating the reliability of laboratory tests in predicting field reactivity, two key parameters are considered: posterior probability and Brier score. For reactive cases, the most reliable test will show high posterior probability, indicating strong predictive power for reactive

outcomes, and low Brier score, reflecting accurate and well-calibrated predictions. Conversely, for non-reactive cases, the most reliable test will display low posterior probability, suggesting the test correctly identifies non-reactive outcomes, while still maintaining low Brier score to ensure the accuracy and calibration of the predictions.

3.8 Decision tree model development

DT, a type of supervised machine learning model, is particularly well-suited for tasks where the goal is to establish decision rules based on a set of input variables [18]. Following the structure presented in Figure 3. 3, DT can effectively define threshold rules, making them ideal for determining the optimal expansion levels and time points that predict reactivity. Each split in the tree aims to maximize the separation between reactive and non-reactive outcomes, thereby refining the predictive accuracy.

Along its structure, the root node represents the entire dataset, being the starting point of the tree and the most significant decision is made. Next, internal nodes represent the points where the data is further split based on additional characteristics, leading to further branching. Leaf nodes are then the terminal points providing the output or classification where no further splits occur, representing a final classification made by the model.

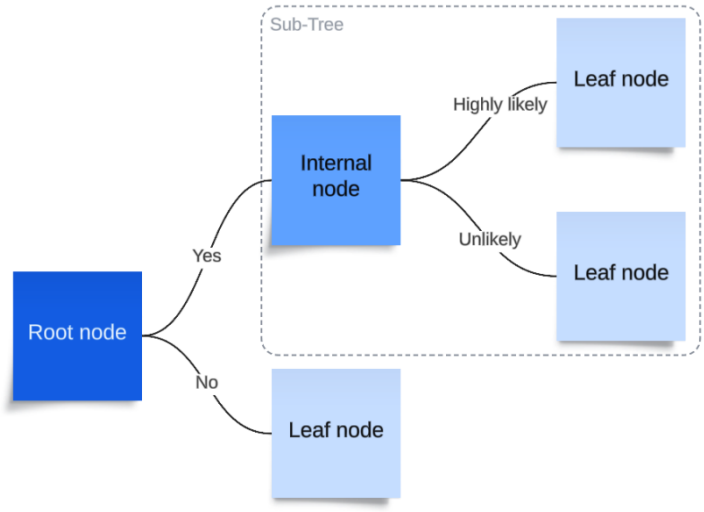


Figure 3. 3 – Decision tree structure, including root, internal and leaf nodes.

In this sense, the dataset was split into training and test sets using a 70-30 ratio, ensuring that the model's performance could be robustly evaluated on unseen data. Moreover, the model was trained

using a recursive partitioning approach, where data was split based on expansion percentages at key time intervals. To prevent overfitting and ensure the model remained interpretable, the DT was constrained to a maximum depth of four. Additionally, a minimum of ten samples were required to split an internal node, and at least five samples were needed at a leaf node.

The decision tree was initially employed to identify and interpret the critical thresholds for predicting field reactivity. Its visualization helped extract decision rules that represent practical thresholds for laboratory test outcomes. To further refine these thresholds, a second decision tree model was developed, incorporating additional parameters such as alkali content (i.e., cement alkali content plus boosting alkali) and environmental factors (i.e., temperature and relative humidity).

This enhanced model allowed for a more comprehensive analysis of the factors influencing AAR risk, enabling the identification of specific thresholds and rules. The second decision tree provided deeper insights into the interactions between variables, aiming to improve the accelerated laboratory outcomes' predictive accuracy in estimating field reactivity based on laboratory results.

3.9 Logistic regression

Logistic regression is a classification machine learning model used to predict the probability of a binary outcome, such as the likelihood of field reactivity [18,19]. The logistic regression model is based on the logistic function Equation 3.11, which transforms input data into a probability value ranging between 0 and 1, suitable for modeling binary outcomes.

$$Logistic\ function = \frac{1}{1 + e^{-x}} \quad \text{(Equation 3. 11)}$$

In this context, a weight factor is used to predict the log odds of the probability that a given input belongs to a particular class (0 or 1). The log odds, or the logit function Equation 3.12 is defined as the natural logarithm of the odds ratio, in which P represents the probability of the event occurring.

$$odds = \frac{P}{P-1} \rightarrow logit(P) = ln\left(\frac{P}{1-P}\right) \quad \text{(Equation 3.12)}$$

Considering the logit function, the probability occurrence can be estimated as per Equation 3.13.

$$logit(P) = mx + b \rightarrow P(x) = \frac{1}{1 + e^{-(mx+b)}} \quad \text{(Equation 3.13)}$$

The maximum likelihood estimation (MLE) is the loss function in the logistic regression; it aims to find the parameter values that maximize the likelihood of the observed data. In this process, MLE optimizes the model parameters by interactively adjusting them to ensure the best fit between the model and the data. MLE is expressed by Equation 3.14, where y_i is the actual output and $\hat{y}_i$ is the predicted probability.

$$L(m) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i) \quad \text{(Equation 3.14)}$$

In logistic regression, each coefficient represents the change in the log odds of the outcome for a one-unit change in the predictor variable, while keeping the other variables constant. In that sense, this approach was employed to train the model to predict the occurrence of reactivity in the field (Equation 3.15) based on the accelerated laboratory test outcomes (i.e., aggregate reactivity levels or expansion, EXP), environmental conditions of exposed elements (i.e., temperature, TEMP and relative humidity, RH) and mix-design information (i.e., alkali content of the binder, ALK).

$$x=\beta_0+\beta_{TEMP}\cdot\text{TEMP}+\beta_{RH}\cdot\text{RH}+\beta_{EXP}\cdot\text{EXP}+\beta_{ALK}\cdot\text{ALK} \quad \text{(Equation 3.15)}$$

The probability of reactivity can, therefore, be calculate as per Equation 3.16.

$$P(x) = \frac{1}{1 + e^{-(\beta_0+\beta_{TEMP}\cdot\text{TEMP}+\beta_{RH}\cdot\text{RH}+\beta_{EXP}\cdot\text{EXP}+\beta_{ALK}\cdot\text{ALK})}} \quad \text{(Equation 3.16)}$$

3.10 References

[1] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, W.M. Lim, How to conduct a bibliometric analysis: An overview and guidelines, Journal of Business Research 133 (2021) 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070.

[2] A. Bergmann, L.F.M. Sanchez, Alkali-Aggregate Reaction (AAR) Comparative Data: Accelerated Laboratory Outcomes vs. Field Performance, (2024). https://doi.org/10.17632/5cscpggnhg.1.

[3] J. Custódio, J. Lindgård, B. Fournier, A. Santos Silva, M.D.A. Thomas, T. Drimalas, J.H. Ideker, R.-P. Martin, I. Borchers, B. Johannes Wigum, T.F. Rønning, Correlating field and laboratory investigations for preventing ASR in concrete – The LNEC cube study (Part I – Project plan and laboratory results), Construction and Building Materials 343 (2022) 128131. https://doi.org/10.1016/j.conbuildmat.2022.128131.

[4] B. Fournier, R. Chevrier, A. Bilodeau, P.-C. Nkinamubanzi, N. Bouzoubaa, COMPARATIVE FIELD AND LABORATORY INVESTIGATIONS ON THE USE OF SUPPLEMENTARY CEMENTING MATERIALS (SCMs) TO CONTROL ALKALI-SILICA REACTION (ASR) IN CONCRETE, in: Brazil, 2016.

[5] C.A. MacDonald, C. Rogers, R.D. Hooton, THE RELATIONSHIP BETWEEN LABORATORY AND FIELD EXPANSION – OBSERVATIONS AT THE KINGSTON OUTDOOR EXPOSURE SITE FOR ASR AFTER TWENTY YEARS, in: Austin, Texas, USA, 2012.

[6] J. Lindgård, P.J. Nixon, I. Borchers, B. Schouenborg, B.J. Wigum, M. Haugen, U. Åkesson, The EU "PARTNER" Project — European standard tests to prevent alkali reactions in aggregates: Final results and recommendations, Cement and Concrete Research 40 (2010) 611–635. https://doi.org/10.1016/j.cemconres.2009.09.004.

[7] J.H. Ideker, A.F. Bentivegna, K.J. Folliard, M.C.G. Juenger, Do Current Laboratory Test Methods Accurately Predict Alkali-Silica Reactivity?, MJ 109 (2012). https://doi.org/10.14359/51683914.

[8] A.H.-S. Ang, W.H. Tang, A.H.-S. Ang, Probability concepts in engineering: emphasis on applications in civil & environmental engineering, 2nd ed, Wiley, New York, 2007.

[9] K.M. Ramachandran, C.P. Tsokos, Descriptive Statistics, in: Mathematical Statistics with Applications in R, Elsevier, 2015: pp. 1–52. https://doi.org/10.1016/B978-0-12-417113-8.00001-1.

[10] O.C. Ibe, Introduction to Descriptive Statistics, in: Fundamentals of Applied Probability and Random Processes, Elsevier, 2014: pp. 253–274. https://doi.org/10.1016/B978-0-12-800852-2.00008-0.

[11] B.R. Martin, Statistics, Experiments, and Data, in: Statistics for Physical Science, Elsevier, 2012: pp. 1–20. https://doi.org/10.1016/B978-0-12-387760-4.00001-9.

[12] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, Bayesian Data Analysis Third edition (with errors fixed as of 13 February 2020), (n.d.).

[13] LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S, Phil. Trans. R. Soc. 53 (1763) 370–418. https://doi.org/10.1098/rstl.1763.0053.

[14] ASTM C1778, Guide for Reducing the Risk of Deleterious Alkali-Aggregate Reaction in Concrete, ASTM International, 2022. https://doi.org/10.1520/C1778-22.

[15] S. Sinharay, Continuous Probability Distributions, (n.d.).

[16] M. Shaw, Use of Bayes' theorem and the beta distribution for reliability estimation purposes, Reliability Engineering & System Safety 31 (1991) 145–153. https://doi.org/10.1016/0951-8320(91)90115-N.

[17] D.A. Redelmeier, D.A. Bloch, D.H. Hickam, Assessing predictive accuracy: How to compare brier scores, Journal of Clinical Epidemiology 44 (1991) 1141–1146. https://doi.org/10.1016/0895-4356(91)90146-Z.

[18] K. Yeturu, Machine learning algorithms, applications, and practices in data science, in: Handbook of Statistics, Elsevier, 2020: pp. 81–206. https://doi.org/10.1016/bs.host.2020.01.002.

[19] H. Belyadi, A. Haghighat, Supervised learning, in: Machine Learning Guide for Oil and Gas Using Python, Elsevier, 2021: pp. 169–295. https://doi.org/10.1016/B978-0-12-821929-4.00004-4.

Chapter 4: Bibliometric analysis and database compilation of laboratory test procedures for assessing concrete field performance against alkali-aggregate reaction (AAR)

Ana Bergmann and Leandro Sanchez

Abstract:

Alkali-aggregate reaction (AAR), a harmful deterioration mechanism, reduces serviceability and causes premature distress in concrete infrastructure worldwide. Once the reaction is triggered, remediation is challenging and uncertain; therefore, prevention remains the most effective strategy against AAR. Among the laboratory tests developed to assess the reactivity of aggregates and preventative measures against AAR, the accelerated mortar bar test (AMBT) and the concrete prism test (CPT) are the most widely used. Nevertheless, discrepancies between laboratory outcomes and field performance have been observed over the years, yet have never been thoroughly and systematically quantified. This study undertakes a critical review and compares currently used laboratory methodologies for assessing AAR-induced development in the field, employing a comprehensive bibliometric analysis. The findings reveal that most studies use visual representation and simple linear correlations to assess discrepancies between laboratory tests and field performance. To address these issues, a robust database is established, integrating laboratory and field performance data. This database will support future research employing advanced data analysis techniques, ultimately aiming to enhance the prediction of AAR occurrence in field structures.

Keywords: Alkali-aggregate reaction; Performance testing; Laboratory/field correlation; Alkali-silica reaction.

4.1 Introduction

Alkali-aggregate reaction (AAR) is a critical deterioration mechanism affecting concrete infrastructure worldwide. This reaction, comprising two distinct types: alkali-silica reaction (ASR) and the alkali-carbonate reaction (ACR), significantly reduces the durability and serviceability of affected structures. AAR is triggered and sustained by the combination of reactive aggregates, high alkali concentration in the concrete pore solution, and sufficient moisture content [1–3].

Once AAR is triggered, remediation is challenging and uncertain, therefore, prevention remains the best approach to avoid such a deterioration. Performance tests conducted in the laboratory, particularly the accelerated mortar bar test (AMBT) and concrete prism test (CPT), are essential for assessing aggregate's reactivity and the effectiveness of preventative measures [1,4]. However, discrepancies between laboratory test outcomes and actual field performance raise doubts about the reliability of these test methods. The increased availability of data from outdoor exposure sites further magnifies these discrepancies, instigating a reassessment of the methodologies [5].

The variations in testing setups and conditions play a significant role in these discrepancies. For example, alkali concentration and leaching can vary depending on the method. Some alkalis may be recycled during the test, while others are bound or leached out depending on moisture levels and test conditions [2]. The effect of temperature on alkali movement and concentration is also uncertain [6]. Thus, determining the effective alkali loading during tests and the real alkali threshold remains a challenge [7]. Sealed prisms may exhibit shrinkage due to lack of water, while submerged prisms may leach out alkalis, resulting in minimal expansion [5]. Moreover, wrapping methods can significantly influence expansion outcomes due to varying levels of alkali leaching [6]. In summary, modifying pretreatment, storage conditions, and prism size can yield different expansion results even with the same mortar or concrete composition [6–8].

The current study employs a comprehensive bibliometric analysis to deepen the understanding of current laboratory test conditions and field performance related to AAR. It highlights the main variables under study, such as alkali loading, storage conditions, sample characteristics, and the outcomes (i.e., expansion over time). By quantifying discrepancies, it becomes feasible to anticipate the correlation of current testing methods with field performance. Furthermore, this research enables the establishment of a robust database that integrates laboratory findings and field performance data. This database will serve as a foundation for advanced data analysis using stochastic approaches and artificial intelligence, ultimately enhancing the prediction of AAR-induced development in field structures.

4.2 Background

4.2.1 Laboratory test methods for assessing aggregate reactivity potential

The evaluation of the aggregates' reactivity in concrete has been the subject of extensive research and development due to the critical need to prevent AAR-related deterioration in concrete. Over the decades, various laboratory test methods have been developed to assess the reactivity potential of aggregates, each with its unique advantages and limitations, reflecting the evolving understanding of AAR-induced deterioration mechanisms and mitigation strategies.

4.2.1.1 Petrographic examination

Petrographic examination, as per ASTM C295 [9], is a microscopic assessment that has been traditionally used as the initial step to appraise the presence of unstable mineral phases in the aggregates, such as opal, chalcedony, and strained quartz, within alkaline environments [6,8,10,11]. The petrographic examination begins by preparing a representative aggregate sample, embedding it in a suitable media, such as epoxy, and slicing it into thin sections. The sections are polished to remove artifacts and ensure a smooth surface. Under a polarizing microscope, the minerals are identified based on their optical properties, such as refractive indices [12]. The goal is to classify aggregates based on their content of reactive forms of silica into three classes: Class I—Very unlikely to be alkali-reactive; Class II—Alkali-reactivity uncertain; and, Class III—Very likely to be alkali-reactive. Although essential, petrographic analysis alone is insufficient for predicting AAR in the field since it is not a performance test and thus does not account for the dynamic interactions between aggregates and the cement paste under varying environmental conditions [1,2,13].

4.2.1.2 Chemical methods

Chemical methods were proposed aiming to evaluate the potential reactivity of aggregates by measuring the silica dissolution and alkalinity reduction of the sodium hydroxide solution in which the aggregates are soaked in the test [14]. In this method, the aggregate sample is reduced to 150-300 μm particles, optimizing surface area for the chemical reaction, and immersed in 1M NaOH at 80°C for 24 hours. The measurements of dissolved silica and reduction in NaOH solution

alkalinity, after the filtering process, are then compared to a classification plot and aggregates can be categorized into three potential reactivity levels: innocuous, potentially deleterious, and deleterious. Although this test offers rapid results, it was found to be somewhat unreliable, particularly in appraising slowly reactive aggregates bearing strained quartz. This procedure, standardized by ASTM C289 [14], has been removed from the standards.

4.2.1.3 Mortar Bar Test (MBT)

In the 1940s, the Mortar Bar Test (MBT) was introduced as the first standardized method, as per ASTM C227 [15], for assessing aggregate's reactivity [16]. This test involves immersing mortar bars in a highly alkaline solution (i.e., 1M NaOH) at a temperature of 38ºC to accelerate ASR-induced reaction [2]. Expansions exceeding 0.05% at three months or 0.10% at six months suggested that the aggregate appraised displays the potential to generate ASR in concrete [17]. Despite its initial adoption, the MBT faced significant criticism and challenges. It was later found to be unreliable due to excessive alkali leaching along with the aggregate's preparation (i.e., crushing and sieving) required prior to testing, which could potentially modify some features of the aggregates (i.e., texture, mineralogical composition, etc.) and thus compromise (i.e., increase or decrease) the test results [5,18]. Additionally, the standard fails to specify the alkali content of the cement and the w/c ratio, two aspects known to affect the outcomes, with higher alkali content and lower w/c ratio leading to greater expansions [19]. The MBT's limitations prompted the development of alternative methods that could provide quicker and more reliable outcomes [20].

4.2.1.4 Accelerated Mortar Bar Test (AMBT)

The Accelerated Mortar Bar Test (AMBT) was then developed in the 1980s and standardized as ASTM C1260 [21], based on the method suggested by Oberholster and Davies from NBRI [20]; this test provides faster results by immersing mortar bars in a 1M NaOH solution at 80°C. These conditions were selected to detect the potential expansion from all reactive aggregate sources, from quickly and highly reactive to slowly and marginally reactive aggregates. To establish the test time and threshold, a linear correlation between the AMBT and MBT outcomes was performed, indicating that although the correlation was moderate (r=0.67), it was determined that an expansion of 0.05% at 3 months in MBT corresponded to 0.11% expansion at 12 days in AMBT.

However, discussions on the time and limits have been extensively explored among researchers aiming to better reflect field conditions, which are often checked against CPT outcomes [4,6,8,22]. In China, the expansions above 0.40% at 16 days indicate reactive behavior, 0.20% to 0.40% suggest potentially reactive behavior, and below 0.20% indicate innocuous behavior [23]. In contrast, 0.15% at 14 days and 0.33% at 28 days effectively classify the Canadian aggregates [22]. In Australia, expansions above 0.1% at 10 days indicate reactive aggregates, while expansions below 0.1% at 10 days but above 0.1% at 21 days indicate slowly reactive aggregates behavior [24]. Overall, the distinction between reactive and non-reactive aggregates falls between 0.08% and 0.20%, with 0.15% at 16 days being reasonable across different regions [25]. Yet, the limits and time are currently established as per ASTM C1778 [26]: expansions below 0.1% at 14 days indicate non-reactive (i.e., innocuous) behaviour, 0.1% to 0.3% indicate moderately reactive, 0.3% to 0.45% indicate highly reactive, and expansions above 0.45% are classified as very highly reactive aggregates.

Various factors intrinsic to the test can influence the outcome and have been extensively studied to calibrate the procedure. These factors, which are detailed and accounted for in the testing procedure, include alkali content and fineness of the cement, aggregate grading and crushing, mortar bar characteristics (i.e., w/c ratio), and storage conditions (i.e., relative humidity, curing, and leaching) [22,24]. For instance, finer aggregate particles can lead to higher expansion [6,27]. Increased cement fineness raises the surface area available for reaction, leading to greater expansion, while alkali content, particularly higher concentrations of sodium and potassium ions, increases the pore solution alkalinity enhancing expansion potential [28]. Mortar bar characteristics, such as mix proportions and w/c ratio, impact the results, with higher w/c ratios resulting in more porous mortars either facilitating the ingress of alkali ions, increasing the potential for expansion, or allowing for alkali leaching, reducing it [29]. Additionally, storage conditions, particularly NaOH concentration affect the test results, with higher concentrations yielding greater expansions [3,30].

Despite its progress and faster results, the AMBT has faced criticism for using mortar instead of concrete, and not accurately representing field conditions, potentially yielding false positives due to the harsh conditions, especially the high temperature [31,32]. In fact, it has been advised against using AMBT to reject aggregates, only to accept them [19].

4.2.1.5 Concrete Prism Test (CPT)

The development of the Concrete Prism Test (CPT), standardized as ASTM C1293 [33], was driven by the need to address the limitations of earlier methods offering a more realistic assessment in terms of storage conditions (i.e., temperature and relative humidity) and specimens' features (i.e., concrete vs. mortar, aggregate grading, etc.) [16]. In the 1980s, the CPT was initially conducted using concrete prism specimens with a cement content of 310 kg/m³, stored in humid chambers (i.e., 100% relative humidity) at 23°C, with a reactivity threshold set at 0.02% expansion after 84 days [32]. However, this setup was later refined in 1994 to better simulate field conditions, increasing the cement content to 420 kg/m³ and the alkali content of the ~~cement~~ mix to 1.25% Na_2O_{eq} of the mass of cement, achieved by adding NaOH to the mixing water, leading to an alkali content of 5.25 kg/m^3 instead of 3.9 kg/m^3 [33]. This modification aimed to induce a more accurate representation of field performance, particularly adopting moderate temperature conditions (i.e., 38°C), similar to MBT, for storage conditions. The new temperature is high enough to asses all types of aggregates, including slowly reactive aggregates [19]. Moreover, the evaluation time has extended to 1-year, which is still adopted along with a threshold of 0.04% to classify aggregates as non-reactive; expansions equal to or exceeding this threshold indicate a potential ASR development [26].

Besides evaluating the susceptibility of an aggregate to developing ASR, the CPT also assesses preventative measures (i.e., the combination of aggregate with cementitious materials), a common mitigation strategy for ASR, such as the employment of slag [33,34]. In this context, the test is conducted similarly to the standard CPT but the cement is partially replaced with the desired amount of SCM on a percent-by-mass basis. Additionally, the alkali content is proportionally reduced from the standard 5.25 kg/m³ Na_2O_{eq}. Furthermore, the evaluation period is extended to 2 years, although the 0.04% expansion threshold remains the critical limit for determining reactivity. Various intrinsic factors within the test can influence the outcome and have been extensively studied for procedure calibration. These factors, which are detailed and accounted for in the testing procedure, include aggregate grading, w/c, and the equivalent alkali content of the system. For instance, larger aggregates lead to reduced initial expansion, but it increases continuously during later ages [6]. In this sense, aggregate grading is addressed by limiting coarse fraction to 70% of the concrete volume and establishing mass fractions per sieve ranges of 4.75-9.5mm, 9.5-12.5mm, and 12.5-19mm. Additionally, w/c is standardized to a small range of 0.42 to 0.45 giving room to

adjust the mix fresh state, reducing the test variability [19]. Finally, the higher the equivalent alkali content of the system, the greater the expansion; therefore, assessing the PC Na_2O_{eq} (i.e., % Na_2O + 0.658 K_2O) and introducing the remaining alkalis (i.e., 1.25% Na_2O_{eq} - % PC Na_2O_{eq}) as NaOH is essential to guarantee a comparison among the outcomes. These considerations are critical in maintaining the test's accuracy, as CPT is designed to reduce the likelihood of false positives and provide a better correlation with field performance [6].

Despite the CPT's higher reliability and closer approximation to field conditions, some aggregates, especially those that react slowly, may still present classification challenges [26]. In this sense, research has shown other parameters than the ones considered in the standard may influence the test results, including alkali leaching, aggregate shape, the use of different fine aggregates as reference, and storage conditions [6,35]. Alkali leaching, for example, can distort test responses, requiring measures such as covering specimens with plastic or accounting for the leached content to reduce the effect [7]. Additionally, large amounts of elongated and flat particles, a consequence of the crushing system, may lower expansions, allowing marginally reactive aggregate to pass the test [19]. Moreover, the nature of the sand, even if non-reactive, can alter the expansion rate of the same coarse aggregate [25]. Along with these limitations, the test's long duration remains a significant challenge, especially when timely project planning is required.

4.2.1.6 Accelerated Concrete Prism Test (ACPT)

The development of the Accelerated Concrete Prism Test (ACPT) was driven by the need to shorten the lengthy testing period associated with the CPT while maintaining a realistic assessment of ASR potential. Initially proposed in 1992, the ACPT aimed to accelerate the rate of expansion observed in the CPT by increasing the storage temperature to 60°C, thereby reducing the test duration from a year to as little as 13 weeks [35].

When comparing the results of the ACPT and CPT, different time frames and thresholds are identified, ranging from 0.015% to 0.04%, over a 4 to 13-week assessment period. Yet, a high correlation (r=0.9) is reported when 0.03% expansion at 3 months at 60°C corresponds to the 0.04% expansion at 1 year at 38°C observed in CPT. Currently, ACPT is a recommended test method from RILEM, AAR-4.1, following the high temperature of 60°C with a test duration suggested for at least 20 weeks, but still not standardized [2]. Despite the faster results, the high temperature can induce artificial reactions (i.e., impact on ASR-gel viscosity) or cause excessive

expansion that would not normally occur in the field [6]. Moreover, the ACPT faces challenges regarding alkali leaching, which has shown increased leaching from test prisms [7,35]. Consequently, careful consideration of leaching effects is essential when interpreting ACPT outcomes, as higher temperatures tend to intensify this issue [2].

4.2.1.7 Miniature Concrete Prism Test (MCPT)

The Miniature Concrete Prism Test (MCPT), standardized as AASHTO T380 [36], was developed aiming to provide an accelerated assessment of aggregate reactivity addressing the drawbacks of the AMBT and CPT, but still incorporating some of their features [37]. For instance, similarly to CPT, it tests concrete, keeping a cement content of 420 kg/m³ and the alkali content of the mix to 1.25% Na_2O_{eq} of the cement mass. However, the MCPT utilizes smaller concrete prisms (i.e., 50x50x285 mm), and reduced volume of coarse aggregate (i.e. 65%). Similarly to AMBT, the prisms are immersed in a 1M NaOH solution, however at 60°C for a varying period of 56 to 84 days, being the longest to address slow reactive aggregates [37]. Moreover, the MCPT eliminates the need for aggregate crushing. Comparing MCPT outcomes to field data, although the high temperature could lead to excessive expansion, a good correlation has been reported [37,38].

When comparing MCPT outcomes with AMBT, a poor correlation (r=0.5) is observed for expansion limits of 0.04% at 56 days compared to 0.1% at 14 days in the AMBT [39]. However, a high correlation (r=0.968) is observed among the CPT and MCPT outcomes. Extensive research has suggested limits of 0.03% at 56 days to accommodate slow reactive aggregates, accounting for test variability [39]. Therefore, for the current standard [36], aggregates are considered nonreactive if 56-day expansion is ≤0.030% or between 0.031% and 0.040% with an average 2-week expansion rate of ≤0.010% between 8 and 12 weeks. If the 2-week expansion rate exceeds 0.010% within this range, the aggregate is slow reactive. Aggregates with 56-day expansions of 0.041% to 0.120% are moderately reactive, 0.121% to 0.240% are highly reactive, and those above 0.240% are very highly reactive. For SCM effectiveness, expansions below 0.020% are effective, those between 0.020% and 0.025% are uncertain, and expansions above 0.025% indicate the measure is not effective.

4.2.1.8 Norwegian Concrete Prism Test (NCPT)

The Norwegian Concrete Prism Test (NCPT), also recommended by RILEM as AAR-10, standardized in Norway NB32 [40], is similar to CPT but addresses the leaching effects by employing larger prisms (i.e., 100x100x450mm) [41]. This means that in relation to their volume, they have less exposed surface area from which alkalis can leach into the surrounding environment [42]. Similar to the CPT, the NCPT involves storing concrete prisms at 38°C in a humid environment (i.e., RH=100%) to induce ASR development. A fixed w/b ratio of 0.48 is defined addressing the sensitivity of this parameter on internal moisture level and alkali leaching [41]. Regarding aggregate volume, the procedure indicates that 60% of the mix should contain the potential reactive aggregate either coarse or fine [43].

This test is particularly valuable for assessing preventative measures, determining the required binder composition to produce non-reactive concrete [41]. The acceptance criteria differ depending on the purpose of the test. When evaluating the reactivity of aggregates combined with a high alkali cement, the threshold is set at 0.04% expansion after 52 weeks, same as CPT. However, for performance testing, where the focus is on assessing the effectiveness of preventative measures, the threshold of <0.030% expansion is applied after 52 weeks and <0.060% after 104 weeks [44]. Concrete incorporating PC, fly ash, and/or silica fume is only tested for 52 weeks, while all other binder combinations require testing for 104 weeks.

4.2.1.9 Summary of test methods

Over time, test methods have been developed to attend to specific project requirements, aggregate characteristics and regional demand [5,8]. They focus on addressing the limitations of previous tests, aiming to enhance the correlation between laboratory results and field performance [45]. For instance, The German Concrete Method (GCM) and the Danish Mortar Bar Test (TI-B51), both standardized, aim to simulate field conditions more precisely. The GCM uses concrete prisms (100x100x450mm) stored in fog chambers at 40°C for nine months, measured immediately with no cooling down period, the acceptance criteria for non-reactive aggregates is 0.06% at 9 months [46]. Conversely, the TI-B51 immerses mortar bar samples (40x40x160mm) in a NaCl solution at 50°C for 52 weeks, enabling alkalis from the solution to penetrate the concrete specimens. Compared to field behavior, GMC effectively identifies non-reactive and normally reactive

aggregates, while TI-B51 underestimates slow-reactive aggregates at 8 or 20 weeks but improves when the test period is extended to 26–52 weeks [46].

Another example is the Concrete Microbar Test (CMBT) developed in the 2000s aiming to rapidly assess the ACR potential of aggregates. The CMBT, similar to the AMBT, utilizes smaller prisms (i.e., 40 x 40 x 160 mm) while maintaining high temperatures (i.e., 80°C) and immersion in 1M NaOH for 30 days [47]. Since the method focuses on ACR, and some carbonate aggregates produce expansion only when used in larger particles, the aggregate size is increased to a range of 5 to 10 mm. When incorporated as a method recommended by RILEM (AAR-5), additional adjustments were made to the test period and aggregate distribution [2]. For instance, final readings are taken at 14 days, and aggregate particles exceeding 8 mm must be crushed.

Given the variety of methods to assess AAR potential, Table 4. 1 summarizes the key features of current the tests, highlighting differences in sample dimensions, shape, storage conditions, and test duration. Some tests focus on the aggregates, some on the mix design performance, and others on both the mix and the exposure performance [41].

Table 4. 1 – Laboratory procedures for evaluating the AAR potential of aggregates.

Procedure	Sample		Test Conditions		Test duration
	Shape	Size	Temperature	Storage	
Accelerated Mortar Bar Test – AMBT – (RILEM AAR-2.1)	Prism	25x25x285 mm^3	80° C	Samples immersed in a 1 M NaOH solution	14 days
Accelerated Mortar Bar Test – AMBT – (RILEM AAR-2.2 / ASTM C1260-22 / AS 1141.60.1)	Prism	40x40x160 mm^3	80° C	Samples immersed in a 1 M NaOH solution	14 days
Concrete Prism Test – CPT – (RILEM AAR-3 / CSA A23.2-14A / AS 1141.60.2)	Prism	75x75x250 mm^3	38° C	RH > 95 %	52 weeks
Accelerated Concrete Prism Test – ACPT – (RILEM AAR-4.1)	Prism	75x75x250 mm^3	60° C	RH > 95 %	20 weeks
Concrete Microbar Test – CMBT – (RILEM AAR-5)	Prism	40x40x160 mm^3	80° C	Samples immersed in a 1 M NaOH solution	14 days
Miniature concrete prism test – MCPT (AASHTO T380)	Prism	50x50x285 mm^3	60° C	Samples immersed in a 1 M NaOH solution	56 days
Danish Mortar Bar Test – TI-B51	Prism	40x40x160 mm^3	50° C	Samples immersed in a 1 M NaCl solution	52 weeks
Norwegian concrete prism test – NCPT – (RILEM AAR-10)	Prism	100×100×450 mm^3	38° C	RH > 95 %	52 weeks
Concrete Cylinder Test – CCT	Cylinder	f = 100 mm h = 200 mm	38° C, 50° C	RH > 95 %	15 weeks
German Concrete Method – GCM	Prism Cube	100x100x450 mm^3 300x300x300 mm^3	40° C	Samples storage in fog chambers	9 months
Alkali-Wrapped Concrete Prism Test – AW-CPT – (RILEM AAR-13)	This procedure can be combined with any of the above methods			Samples wrapped with water-holding material with alkali hydroxide solution (same as concrete pore solution)	

Despite the broad range of laboratory methodologies, including AMBT and CPT, certain limitations persist. These limitations can generate disparities between laboratory and field performance of concrete. Therefore, a more rigorous comparison and correlation analysis between laboratory and field outcomes is demanded. Thus, compiling existing literature data will guide establishing a reliable framework for predicting AAR-induced development in the field. This approach would enable the translation of laboratory results into predictive risk parameters for field performance, effectively bridging the gap between laboratory and field scenarios [7,8].

As, the evolution of ASR testing methods reflects ongoing efforts to balance accuracy, reliability, and practicality, future research should focus on refining existing methodologies outcomes to better reflect real-world conditions. This includes addressing issues such as alkali leaching and environmental simulation to ensure that laboratory results can be reliably translated into field performance predictions. The development of a robust database to consolidate findings from various tests can further enhance the understanding and prevention of AAR. Therefore, ensuring durability under diverse environmental conditions, contributing to more sustainable infrastructures.

4.3 Field studies on AAR development

Field studies are widely acknowledged as the most reliable method for investigating the long-term behavior against AAR [32]. These structures, exposed to varying environmental conditions over extended periods, have been essential in bridging the gap between laboratory performance tests and real-world behavior [6]. Figure 4. 1 presents a timeline of the established sites over time.

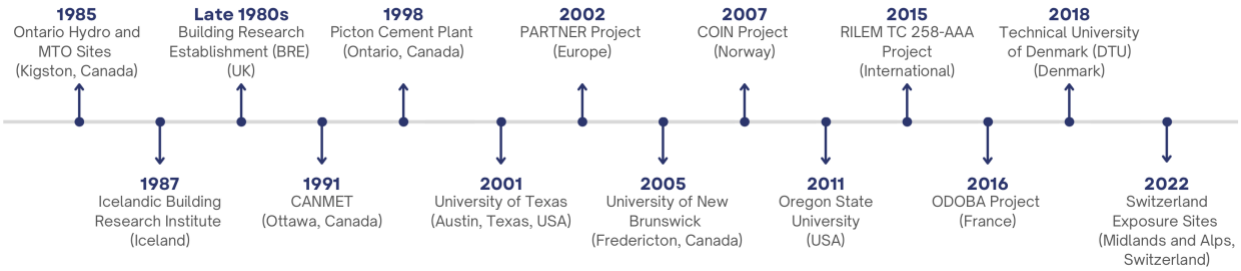


Figure 4. 1 – Timeline of established exposure AAR sites.

Since the 1960s, Ontario Hydro has conducted extensive studies to validate laboratory tests against field performance, particularly focusing on the effects of freezing, thawing, and deleterious aggregates on concrete durability [5,48]. In 1985, the Ontario Ministry of Transportation (MTO) built six concrete sidewalk sections using argillaceous dolomitic limestone aggregates, known for their tendency to induce alkali-carbonate reaction (ACR). The objective was to assess the effects of different alkali levels in cement (i.e., high and low-alkali concrete) and the use of slag in mitigating AAR. The study expanded in 1991, when an additional site was established using the highly reactive Spratt limestone aggregate, where a mix of unreinforced and reinforced concrete beams and slabs was tested [48,49]. These mixtures included SCMs like slag, fly ash, and silica fume, each designed to manage ASR reactivity.

In 1987, the Icelandic Building Research Institute constructed six air-entrained concrete walls using sea-dredged coarse aggregates from Hvalfjörður and Saltvík, along with a blend of fine aggregates from various locations [5]. These walls aimed to assess the reactivity potential of these aggregates under natural conditions. Subsequent field and laboratory investigations focused on exploring different factors such as cement alkali levels, silica fume content, and particle sizes of reactive aggregates [50]. By the late 1980s, the Building Research Establishment (BRE) in the United Kingdom established several exposure sites to evaluate how SCMs like fly ash, slag, metakaolin, and lithium-based admixtures could mitigate ASR in real-world settings [5] [51]. The concrete blocks used a range of reactive aggregates, including flint and greywacke, combined with high-alkali cement.

In 1991, Canada Centre for Mineral Energy Technology (CANMET) initiated a long-term field and laboratory research program in Ottawa, Canada to explore the reliability of SCMs and lithium-based admixtures in mitigating ASR [8,34,52,53]. More than 250 concrete mixtures incorporating reactive aggregates were cast, using different alkali contents and SCMs like fly ash, slag, and silica fume. Concrete blocks, prisms, and slabs were exposed outdoors and monitored over time for expansion. In 1998, at the Picton Cement Plant in Ontario, field studies aimed to evaluate ternary blends of high-alkali cement with silica fume and slag for their ability to mitigate ASR [5,54]. Seven pavement slabs were constructed using the highly reactive Spratt limestone and variations with high-alkali cement, silica fume, and slag.

The PARTNER project (2002-2006), funded by the European Community, aimed to establish a unified test procedure for evaluating the alkali-reactivity of aggregates across Europe [55]. A total

of 100 cubes were produced for 50 tests and it involved 24 partners from 14 countries, with both laboratory and field testing of 13 different aggregate types across eight field locations in Norway, Sweden, Denmark, Germany, France, Spain, Italy, and Portugal [46,56]. The goal was to assess the influence of environmental conditions on ASR development. Additionally, the project produced a petrographic atlas of potentially reactive rocks in Europe, providing a valuable resource for future assessments [12,57].

In 2001, The University of Texas at Austin, funded by TxDOT, established an exposure site in 2001 aimed at understanding the mechanisms of ASR and DEF, developing effective test methods and preventive measures, and transferring the knowledge gained to improve the durability of local transportation infrastructure [5,58]. A significant aspect of the project was the evaluation of variations in test methods, such as AMBT and CPT, for their ability to predict ASR-related expansion and cracking in concrete.

In 2005, the University of New Brunswick (UNB) initiated a research program near the Mactaquac Generating Station to evaluate concrete mixtures made with highly reactive Springhill greywacke aggregates and fly ash [5]. The research program was initiated to evaluate concrete mixtures for the potential reconstruction of the Mactaquac Generating Station, which is heavily affected by ASR [59].

The COIN project (2007-2014), conducted in Norway and linked with the RILEM TC 219-ACS initiative, aimed to develop reliable ASR performance testing methods by examining the impact of various parameters, such as curing and storage conditions, on concrete's internal moisture state, alkali leaching, and ASR expansion [60]. The project was divided into two parts: a study (Part I) that highlighted the critical role of alkali leaching in controlling ASR expansion, and a follow-up study (Part II) that tested 20 different concrete mixtures across two field exposure sites in Norway and Portugal. The project tested concrete mixtures with different cement types and alkali contents using control reactive aggregates from Norway (Ottersbo) and Canada (Spratt) [5].

In 2011, Oregon State University launched a field site dedicated to investigating the potential of fine lightweight aggregates in mitigating ASR in concrete [5]. This study aimed to evaluate how lightweight aggregates could enhance concrete durability, particularly in aggressive environments where ASR risk is high. Additionally, in June 2011, the exposure site at the University of Hawaii in Manoa part of the Federal Highway Administration's ASR Development and Deployment Program, exposed 40 different concrete mixtures [61]. These mixtures were produced using local

basaltic aggregates from various Hawaiian quarries, as well as imported sand from British Columbia (Canada), incorporating varying alkali contents, fly ash, and lithium-nitrate admixtures to assess their effectiveness in mitigating ASR. The blocks, placed at the Magoon Research and Instruction Facilities, are being closely monitored for length changes and the onset of cracking over time. The aim is to compare the data gathered with laboratory tests to evaluate the long-term predictive accuracy of ASR susceptibility under Hawaii's specific environmental conditions. Similarly, at the Lawrence, Massachusetts site, established in June 2012, concrete blocks from 73 different mixtures, including those with highly reactive aggregates, were produced [61]. By maintaining identical blocks across multiple locations, such as the Hawaii site and other sites in Texas, Ontario and New Brunswick, the aim it to assess the impact of environmental exposure on ASR progression, along with the effectiveness of mitigation strategies.

In 2015, the University of Toronto's Leaside site was established to evaluate the performance of SCMs in low-alkali concrete mixes containing reactive Sudbury and Spratt aggregates [5,62]. At the same year, the RILEM TC 258-AAA project expanded international efforts to evaluate ASR mitigation by exposing concrete cubes containing highly reactive aggregates and Class F fly ash to 10 different exposure sites across Europe and North America. These sites include Trondheim (Norway), Brevik (Norway), Düsseldorf (Germany), Paris (France), Lisbon (Portugal), Cascais (Portugal), Reykjavik (Iceland), Austin (USA), Treat Island (USA), and Ottawa (Canada).

In 2016, the Observatoire de la Durabilité des Ouvrages en Béton Armé (ODOBA) project in France was launched to study internal swelling reactions (ISR), including ASR, in large concrete structures, with a focus on applications in nuclear power plant containment [63]. Located at the Observatoire de la Durabilité des Enceintes (ODE) outdoor experimental platform in Cadarache, France, the project involves the construction of massive 8 m^3 concrete blocks that mimic the composition and structure of those used in French nuclear reactors. The blocks are heavily instrumented to monitor the effects of ISR under both natural and accelerated aging conditions.

By 2018, the Technical University of Denmark (DTU) established an outdoor exposure site to examine the long-term behavior of ASR-affected concrete [64]. This project aimed to compare natural exposure conditions with accelerated lab tests to evaluate how well laboratory methods predicted real-world expansion. Additionally, the site is testing the use of hydrophobic silane impregnation in reducing or delaying ASR, with continuous monitoring of moisture content and structural integrity.

Finally, in 2022, a Swiss project was initiated with two exposure sites—one in the Swiss Midlands and another in the Alps—to study ASR development under different environmental conditions [65]. The project aims to validate laboratory tests like CPT and Residual Expansion Test (RET) against long-term field data, focusing on concrete mixtures used in Swiss dam construction.

4.4 Scope of the work

This study aims to provide a detailed analysis of the discrepancies among the most widely adopted methods for assessing AAR in concrete. Through a comprehensive bibliometric review, the research identifies key contributors and maps research collaborations, providing insights into current research related to laboratory assessment for field performance (i.e., exposed field locations, sample sizes, and mix-design).

The primary objective is to deepen the understanding of existing literature on laboratory and field assessments of AAR, establishing a foundation for future research. By compiling and analyzing available data, this study seeks to develop a robust database that supports advanced analysis and improves the reliability of AAR predictions. This database will serve as a crucial resource for future studies aimed at refining predictive models of AAR occurrence in field structures.

4.5 Materials and Methods

To comprehensively evaluate the discrepancies in laboratory tests for assessing AAR occurrence in field structures, a detailed bibliometric analysis was conducted. This methodology systematically acquires, organizes, and evaluates existing literature, uncovering research trends and measuring their associations [66].

The overall procedure, from data to outcomes, is presented in Figure 4. 2. The initial step involved defining the purpose and scope of the bibliometric study, which aims to identify influential researchers and key studies focusing on evaluating AAR-induced deterioration in field structures (i.e., exposure sites). The research concentrated on literature correlating both laboratory and field performance, particularly focusing on newly constructed structures (i.e., blocks) rather than existing ones (i.e., dams, bridges).

A comprehensive search strategy was employed using a combination of keywords to ensure accuracy and relevance. The keywords included: “Alkali,” “silica,” “carbonate,” “aggregate,” “reaction,” “efficiency,” “combining,” “correlating,” “evaluation,” “divergence,” “prediction,” “monitoring,” “performance,” “comparing,” “comparison,” “field,” “in-situ,” “blocks,” “sites,” “exposed,” “laboratory,” “tests,” and “results.” Recognized databases such as Web of Science were utilized as primary sources. Optimization searching tools, such as Boolean operators (i.e., AND, OR, NOT), were applied to streamline search results.

After the initial search, a refinement step was implemented. Citation topics were limited to “concrete science,” “mechanics,” “testing and maintenance,” “computer vision and graphics,” “sustainability science,” “management,” and “geotechnical engineering,” ensuring that irrelevant topics were excluded from the analysis. The final data collection step involved reviewing all abstracts to ensure only publications covering both laboratory and field performance were included, thereby eliminating information not within the research scope.

After gathering the bibliometric data, VOSviewer software was used to perform a co-authorship clustering analysis, illustrating connections among researchers and trends over time. A quantitative analysis of the volume of co-authored documents per author and per country was conducted to understand the environments in which current field studies are explored. In this methodology, all co-authorship links were assigned equal weight as part of a full counting procedure. A summary of the main documents and network visualization of the results are provided in section 5.

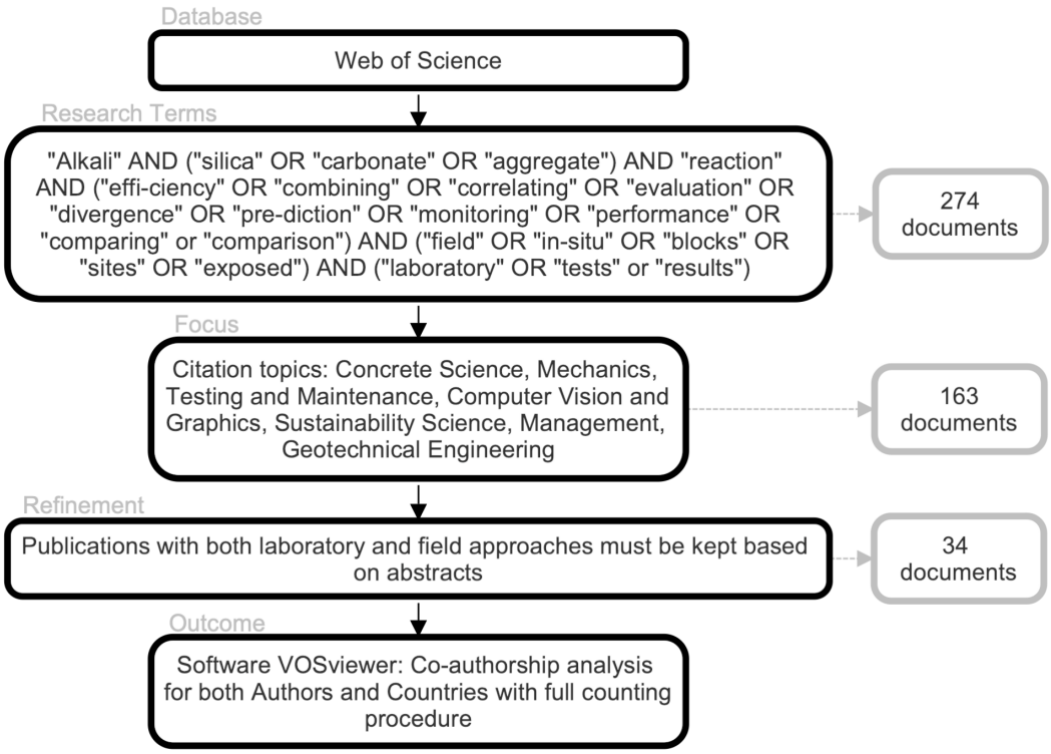


Figure 4. 2 – Bibliometric analysis flow: identifying the correlation between laboratory tests and field data for AAR assessment.

To further enhance the understanding of laboratory and field performance correlations, a robust database was developed [67]. This database, compiled from published data, includes experimental results from five projects involving laboratory accelerated tests conducted under different setups and field-exposed concrete members subjected to distinct environments [34,44,46,68,69]. The database categorizes information on material selection (i.e., aggregate, cement, SCMs, and admixtures) and mix proportions (Figure 4. 3). Additionally, it details the characteristics of laboratory and field samples (i.e., surface area, volume) along with their performance (i.e., expansion over time), test setup, and environmental conditions.

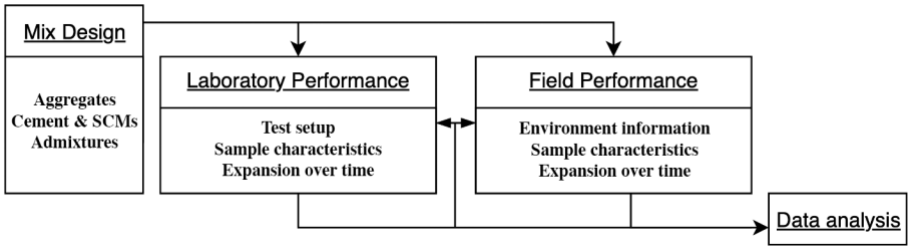


Figure 4. 3 – Database collection flowchart for assessing the variability from laboratory tests to predict field performance.

Therefore, this database integrates laboratory findings with field performance data, providing a foundation for advanced data analysis. This integration facilitates future research and enhances the predictive capabilities of accelerated laboratory tests regarding AAR in new concrete structures.

4.6 Results

The bibliometric data collection outcome of the 34 papers [7,8,30,31,37,44,48,54,55,59,69–92] is presented in Appendix A. It lists the authors' names, countries of affiliation, publication years, and citation counts, forming the basis for building the collaboration network for the bibliometric analysis.

4.6.1 Bibliometric Analysis

By implementing bibliometric analysis techniques, insights into the correlation between field and laboratory test performance for assessing AAR were highlighted. Figure 4. 4 displays the co-authorship network derived from the analysis conducted using VOS Viewer software. The network illustrates the authors' frequency of publication (i.e., size of nodes), research timeframe (i.e., color scheme), and collaboration strength (i.e., the thickness of lines).

Notably, prominent researchers such as Fournier, Thomas, Ideker, Drimalas, and Lindgard emerge as central nodes in the co-authorship network, indicating their significant influence and extensive collaboration within the AAR community. Their work spans several decades, with a noticeable increase in collaborative efforts from the early 2000s to the present, reflecting ongoing and evolving research interests in correlating accelerated laboratory test results with field performance.

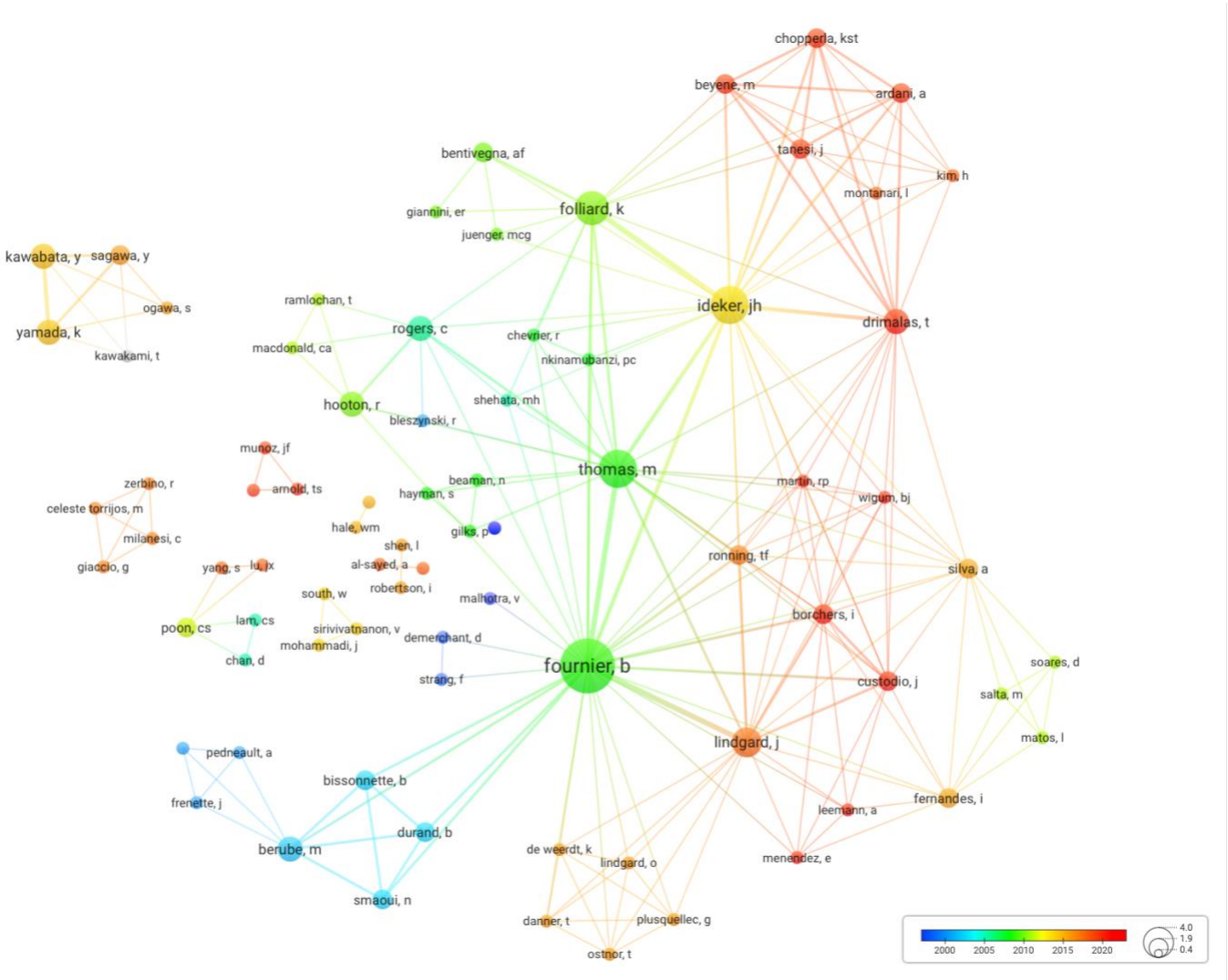
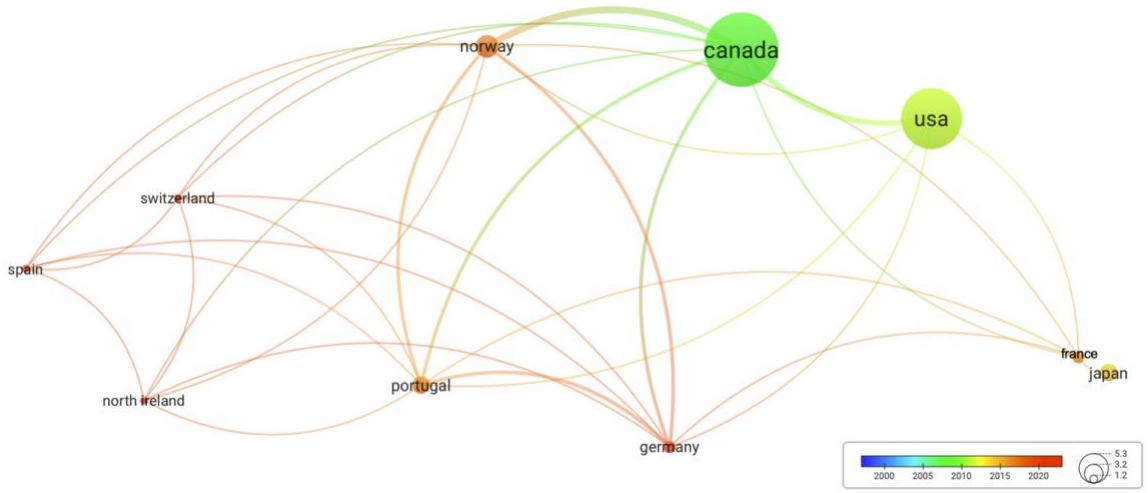


Figure 4. 4 – Network linkage from the co-authorship analysis for the 34 selected papers, available at weblink.

Accordingly, Figure 4. 5 illustrates the collaboration networks between countries, highlighting the significant contributions of North American researchers (i.e., Canada and the USA) in correlating laboratory and field performance of AAR over the years. The graph also reveals recent movements towards inter-laboratory partnerships in evaluating the impact of different environments

[44,46,55]. Furthermore, the network demonstrates strong international collaboration, especially among researchers from Canada, the USA, and Norway, underscoring the global nature of AAR challenges and the collective effort to address them.

Moreover, the increased frequency of publications and collaborations in recent years suggests a growing recognition of the importance of understanding and predicting AAR through both laboratory and field studies. This trend is likely to continue, driven by the need for more reliable and universally applicable predictive models to ensure the durability and serviceability of concrete infrastructure worldwide.



Canada [7,8,30,31,44,48,54,55,59,71,84–88] – Norway [7,30,44,55]
USA [31,37,44,69,72,73,77,78,84,89,90,92] – Japan [70,76,81] – Portugal [44,55,82]
Germany [44,55] – France [44,81] – Switzerland, Spain, North Ireland [55]

Figure 4. 5 – Network linkage from the co-authorship analysis for the 34 selected papers focusing on researchers' countries, available at weblink.

A meticulous evaluation of the documents uncovers information regarding field specimen dimensions, laboratory-performed tests, and evaluation protocols. As summarized in Table 4. 2, field studies predominantly utilize non-standardized block sizes, ranging from 100 mm to 3000 mm, nor structural elements, such as beams, slabs, and pavement blocks. The study objectives frequently converge on AAR mitigation solutions, emphasizing the use of SCMs in real-world environments. Aside from understanding the microstructure and devising new laboratory test procedures, evaluating the long-term performance of exposed blocks under different environmental conditions is also considered among the goals of the selected studies.

Table 4. 2 – Occurrence of field exposed structures by their characteristics.

Shape	Size	Surface area (m^2)	Volume (m^3)	Reference
Prism	400x400x700 mm^3	1.440	0.112	[31,59,80,84–86]
Prism	380x380x710 mm^3	1.368	0.103	[37,69,72]
Prism	400x400x600 mm^3	1.280	0.096	[70,76]
Prism	600x600x2000 mm^3	1.200	0.072	[48,84]
Prism	150x700x700 mm^3	1.400	0.074	[84,86]
Cube	300x300x300 mm^3	0.540	0.027	[44,55]
Prism	200x1200x4000 mm^3	11.680	0.960	[48]
Prism	286x910x910 mm^3	2.697	0.237	[92]
Cube	3000x3000x3000 mm^3	54.000	27.000	[59]
Prism	60x100x200 mm^3	0.076	0.001	[74]
Prism	1200x4000x2000 mm^3	30.400	9.600	[84]
Cube	100x100x100 mm^3	0.060	0.001	[44]
Cube	150x150x150 mm^3	0.135	0.003	[44]

The extensive bibliometric analysis reveals the absence of a universally accepted AAR assessment protocol, with over 20 evaluation methodologies identified (Table 4. 3). These methods range from simple expansion/length measurement (E), crack width (CW), porosity (P), and mechanical properties (i.e., compressive strength (CS), modulus of elasticity (ME), flexural strength (FS)) to chemical and microscopic evaluations (i.e., chemical analysis (CA), pore solution analysis (PS), chloride diffusion (CD), rapid chloride permeability test (RCPT), alkali-leaching analysis (AL), salt scaling (SS), thermogravimetric analysis (TGA), concrete air-void spacing factor (ASF), Raman microscopy (RM), scanning electron microscope (SEM)).

Additionally, more elaborate techniques, including petrographic analysis (PA), damage rate index (DRI), stiffness damage test (SDT), and qualitative damage assessment (QDA) have been used. Notably, only two non-destructive approaches, ultrasonic pulse velocity (UPV) and bulk electrical resistivity (BER), have been observed.

Therefore, since more than 20 different techniques have been used to assess the damage related to field-exposed blocks due to ASR in the studies, it indicates that a consensus has not yet been reached on a standardized assessment protocol. This variability implies that evaluations to identify the risk of ASR occurrence in the field, when compared to accelerated laboratory methods, differ for each study.

Table 4. 3 – Occurrence of testing procedures on exposed structures.

Test	Freq.	Reference
Expansion/length measurement (E)	7	[31,55,59,69,77,84,92]
Chemical analysis (CA)	4	[31,54,55,74]
Petrographic analysis (PA)	4	[31,37,48,72]
Damage rate index (DRI)	4	[44,55,85,86]
Scanning electron microscope (SEM)	4	[37,44,55,80]
Crack width (CW)	3	[55,85,86]
Pore solution analysis (PS)	3	[37,72,84]
Stiffness damage test (SDT)	3	[44,85,86]
Compressive strength (CS)	2	[44,80]
Modulus of elasticity (ME)	2	[44,80]
Flexural strength (FS)	2	[74,80]
Chloride diffusion (CD)	2	[48,54]
Porosity (P)	1	[72]
Rapid chloride permeability test (RCPT)	1	[48]
Alkali-leaching analysis (AL)	1	[76]
Salt scaling (SS)	1	[54]
Thermogravimetric analysis (TGA)	1	[72]
Concrete air-void spacing factor (ASF)	1	[44]
Raman microscopy (RM)	1	[55]
Qualitative damage assessment (QDA)	1	[55]
Ultrasonic pulse velocity (UPV)	1	[44]
Bulk electrical resistivity (BER)	1	[72]

Substantial information was obtained regarding the accelerated laboratory tests conducted to assess aggregates' reactivity. The data displays that ACPT, AMBT, AW-CPT, CCT, CPT, MCPT, and NCTP were among the tests performed. Notably, the CPT test was the most common, accounting for 36.4 % of the occurrences, followed by the AMBT test, accounting for 30.3 % of the occurrences. These observations reaffirm that the ABMT and CPT are among the most commonly used and trusted tests to determine the aggregate reactivity potential [1,8].

Figure 4. 6 demarcates regions with registered AAR investigations. Geographically, most ongoing research sites investigating long-term AAR performance are situated in the northern hemisphere. Notable sites in Canada include the Ontario Ministry of Transportation in Kingston (1985), CANMET studies in Ottawa (1991), Picton (1998), the University of Toronto's Leaside (2003), and the University of New Brunswick in Fredericton (2005) [5]. In the USA, significant research is conducted at the University of Texas in Austin (2001) and the Texas Department of Transportation's Cedar Park site. Additionally, two sites were implemented in the early 2010s as

part of the FHWA ASR Development and Deployment Program, located at the University of Hawaii in Manoa, Oahu Island, and at a DOT facility in Lawrence, Massachusetts, addressing AAR in warmer climates [5].

International collaboration efforts are also highlighted, such as the PARTNER project, initiated in the early 2000s, which involved 24 partners from 14 countries working together to evaluate the reliability of RILEM and local test accelerated laboratory test methods for AAR [5]. Another key effort is the COIN project, launched in 2007 in Norway, which aimed to assess the impact of different test setups and environmental temperatures on AAR development. The COIN project established exposure sites in both Portugal and Norway to evaluate warm versus cold weather comprehensively. Additionally, the RILEM TC 258-AAA project, which began in 2015 as a combination with the Norwegian KPN initiative, further exemplifies the collaborative approach to AAR research.

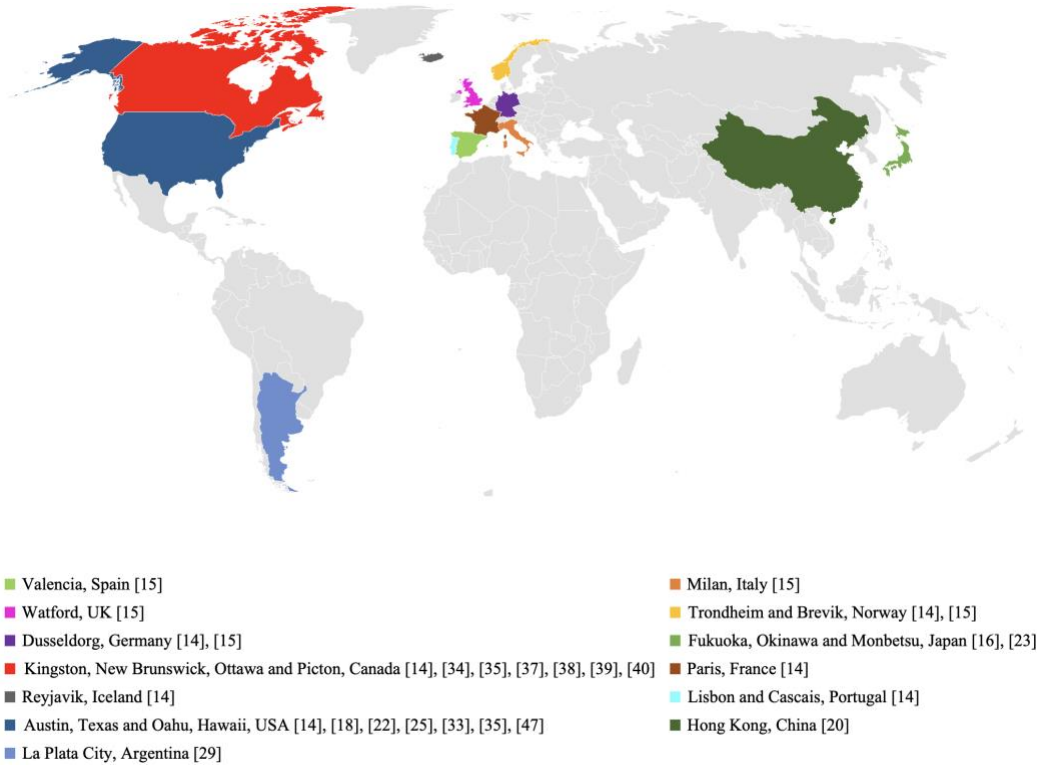


Figure 4. 6 – Location of exposed sites evaluating AAR development on field structures.

These projects brought together experts from multiple countries to assess standardized testing protocols and evaluate mitigating strategies for AAR in concrete structures, highlighting the global nature of this research and the collective effort to improve the understanding and management of AAR. The significance of these international collaboration efforts underscores the value of the current study, which aims to synthesize and build upon the extensive knowledge generated over decades. By consolidating data from these studies, this research contributes to developing more reliable and universally applicable predictive models for AAR, ultimately enhancing the durability and serviceability of concrete infrastructure worldwide.

4.6.2 Employed laboratory methods

AAR-affected field concrete has been a topic of study since 1960s [5], aiming to bridge laboratory and real performance discrepancies. A very relevant comparative and laboratory research program, implemented by Canadian researchers for over 30 years [34], has revealed significant insights. While both CPT and AMBT reliably identified alkali-reactive aggregates in this study, the correlation between CPT and exposed blocks gradually decreased in the presence of SCMs. Conversely, correlations improved when alkalis and/or higher temperatures were set during some laboratory tests.

A study by [31] indicated that aggregate performance in laboratory and field environments varied considerably, also AMBT results often disagreed with CPT. In fact, the AMBT exhibited reliability for specific aggregates, that fail with CPT. In terms of aggregate size, both tests successfully identified alkali-silica reactivity for fine aggregates, with reactivity levels varying greatly between testing setups. For coarse aggregates, the correlation between AMBT and CPT was less pronounced. Specifically, the CPT indicated significant expansion within various coarse aggregates, which exhibited innocuous behavior in AMBT. Therefore, it is not uncommon to accept the use of aggregates that pass the aggressive AMBT test yet show detrimental expansion in outdoor exposure block testing and CPT testing.

Similarly, AMBT was found to be an unreliable indicator of ASR [77]. Using aggregate from Hawaiian quarries, AMBT provided false positive occurrences of ASR, contradictory to ACPT, which was indicated to be reliable. Specimens that failed ACPT (i.e., classified as reactive) performed poorly in field exposure conditions, as expected by the laboratory outcome.

Given the inconsistencies highlighted, new testing protocols have attempted to incorporate comprehensive variables aiming for a resolution. The AW-CPT, proposed by [76], advocates that sample wrapping substantially reduces alkali leaching and minimizes drying compared to regular CPT. ASR-induced expansion of concrete blocks has demonstrated the accuracy of the AW-CPT in predicting concrete expansion under actual environmental conditions [27]. Furthermore, the authors reported that environmental conditions influenced the expansion behavior of ASR-induced concrete blocks. Based on a numerical simulation, rainfall was found to have a critical effect on the hygrothermal properties of concrete in the field. Similarly, [37] found that blocks exposed to elevated temperatures expanded at rates 4-5 times faster than their counterparts in milder settings.

Most studies show that SCMs reduce concrete expansion in exposed blocks when employed to mitigate ASR development. For example, concrete slabs and beams, subjected to the Canadian climate, demonstrated the efficacy of a ternary blend of 25% slag, 3.8% silica fume, and high-alkali Portland cement to prevent damaging ASR [48]. In terms of laboratory test procedures, [84] concluded that a reactive aggregate-SCM combination that passes the AMBT 14-day threshold of 0.10 % is unlikely to cause a damaging expansion in the field, and the 28-day expansion limit is unlikely to fail many reactive aggregate-SCM combinations. Thus, they recommend that a reactive aggregate-SCM combination should have AMBT results of less than 0.1% at 14 days to be used as a threshold. Investigations into CPT and MCPT for aggregate-SCM combinations revealed CPT's failure in predicting the behavior of exposed blocks containing SCMs, while MCPT was shown to be reliable [73]. The study also suggests that the hardened paste properties could predict the efficacy of aggregate-SCM combinations to prevent ASR, thus reducing performance-based testing.

Table 4. 4 summarizes the frequency of occurrence and comparisons between different testing protocols observed in the selected studies. The most frequently compared methods are the CPT (Concrete Prism Test) and AMBT (Accelerated Mortar Bar Test), which appear in multiple instances, highlighting their prevalence and significance in AAR research. In fact, AMBT appears frequently, with 6 correlations with CPT, 2 with ACPT, and 1 with NCPT. This reflects AMBT's significant role and sense of reliability over the testing setup in evaluating alkali-silica reactivity in aggregates.

Additionally, the CPT is noted to have correlations with other methods in 2 instances, indicating its widespread use and sense of reliability in the field. Comparisons with the ACPT (Accelerated Concrete Prism Test) appear in 3 instances, showing that ACPT is often used as a complementary method to CPT for scenarios requiring accelerated results. The AW-CPT (Alkali-Wrapped Concrete Prism Test) has 2 correlations with CPT and 1 with ACPT, aiming to address issues associated with alkali leaching. The NCPT (Norwegian Concrete Prism Test) has 1 correlation with both CPT and ACPT, suggesting a direct analysis.

The MCPT (Miniature Concrete Prism Test) and CCT (Concrete Cylinder Test) show fewer correlations, indicating their niche applications and limited use compared to the more widely used methods like CPT and AMBT.

Table 4. 4 – Occurrence correlation between testing protocols observed in the selected studies.

	CPT	ACPT	AW-CPT	NCPT	AMBT	MCPT	CCT
CPT	2 [85,86]	-	-	-	-	-	-
ACPT	3 [44,55,76]	0	-	-	-	-	-
AW-CPT	2 [70,76]	1 [76]	0	-	-	-	-
NCPT	1 [44]	1 [44]	0	0	-	-	-
AMBT	6 [31,44,59,77,80,84]	2 [44,69]	0	1 [44]	3 [48,54,74]	-	-
MCPT	1 [72]	1 [37]	0	0	0	0	-
CCT	0	1 [37]	0	0	0	1 [37]	0

4.7 Evaluation and future directions for AAR assessment methods

It has been confirmed that the ASR reactivity potential of aggregates is commonly evaluated using the AMBT and the CPT. Despite their widespread adoption, these tests often yield contradictory results, as summarized by the data gathered from the bibliometric analysis (Figure 4. 7). This figure illustrates the uncertainty zones (i.e., hashed red areas) when defining aggregates as reactive or non-reactive according to ASTM C1778 [26], based on the limits of 0.04% for CPT and 0.1% for AMBT. Such inconsistencies highlight the potential for false positives, questioning the reliability of these tests and the parameters influencing their outcomes.

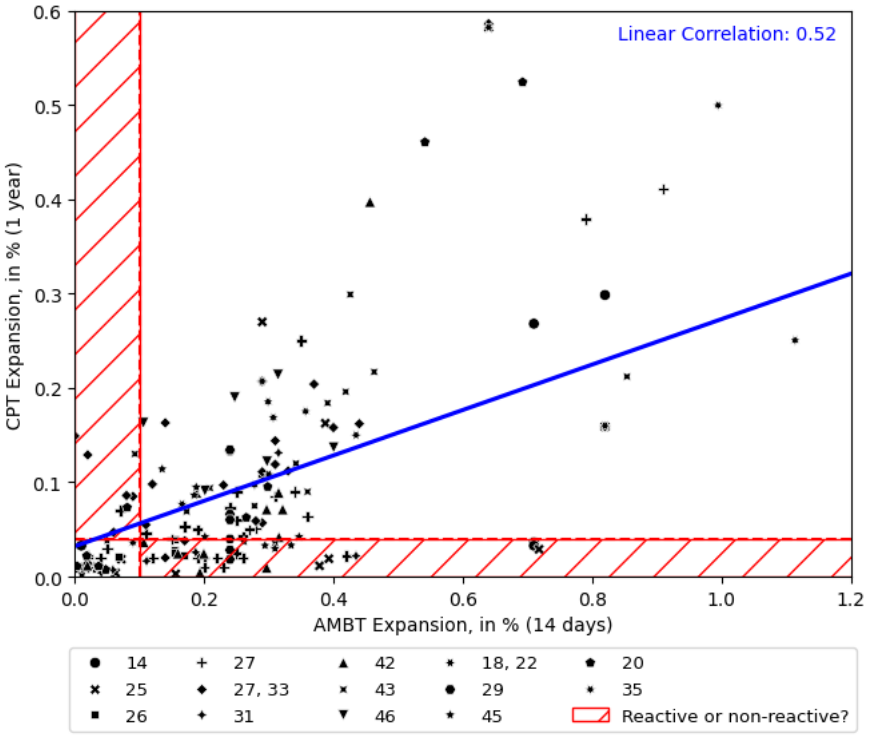


Figure 4. 7 – Comparison between CPT and AMBT results from the bibliometric analysis.

Despite extensive testing of AAR-induced concrete blocks, a proper correlation between short-term laboratory results and long-term field performance remains indefinable. Figure 4. 8a summarizes the approaches used in 34 papers from the bibliometric analysis, categorized by visual representation, expansion over time, and numerical simulation. Notably, numerical simulations were employed in only one study, while visual representations and data on expansion over time predominated.

Figure 4. 8b provides a detailed breakdown of the types of visual representations used, including composition (i.e., stocked column charts), distribution (i.e., histogram or scatter chart), relationship (scatter chart), and comparison (i.e., column chart or line chart). Relationships and comparisons were most frequently employed to represent field and laboratory outcomes, indicating that current interpretative approaches are predominantly descriptive. These methods do not adequately predict future behavior or examine the individual contributions of various variables to overall expansion.

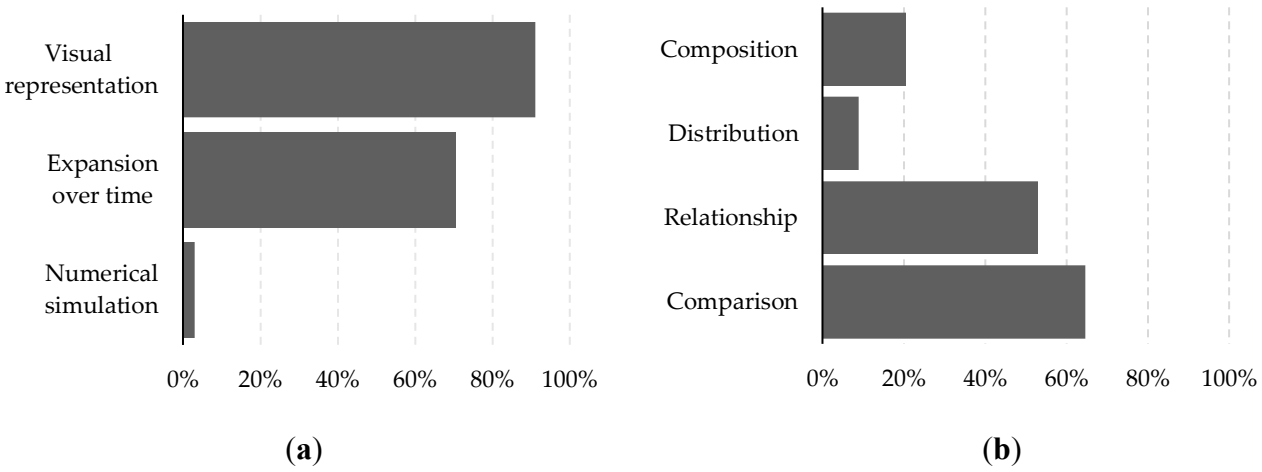


Figure 4. 8 – Approaches used in comparing laboratory and field performance against AAR: (**a**) Evaluation of visual representation, expansion over time, and numerical simulation; (**b**) Analysis of specific visual representation types: composition, distribution, relationship, and comparison.

Given the descriptive nature of current methodologies, research is needed to develop accurate predictive models using established methods. Laboratory tests, while influential, must be refined to more accurately represent field conditions, considering the numerous variables that can impact outcomes. To mitigate the risk of false non-reactive results, incorporating stochastic and artificial intelligence (AI) approaches into predictive models is advisable. These methods have proven successful in previous applications and can enhance the reliability of AAR assessments [93].

However, the implementation of these advanced tools has been slow, partly due to the scarcity of civil engineering professionals trained in information technology [94]. To address this gap, several strategies can be explored, including:

- Probabilistic Modeling: This method can explore the likelihood of AAR susceptibility in field concrete containing reactive aggregates, effectively capturing the uncertainties surrounding AAR risks.
- Monte Carlo Simulations: These simulations can generate random samples from probabilistic distributions, allowing for the analysis of multiple variable combinations and improving field AAR predictions.
- Machine Learning (ML) Algorithms: Training ML algorithms with existing laboratory and field data can enhance the prediction of AAR risks, facilitating more accurate decision-making.

In conclusion, leveraging stochastic and AI techniques can significantly improve AAR risk assessment and prediction. However, the effectiveness of these approaches depends heavily on the availability of a robust and comprehensive database. Such a database should encompass detailed laboratory and field data, including environmental conditions, aggregate properties, and historical performance. By integrating extensive datasets into predictive models, the accuracy and reliability of AAR assessments can be substantially enhanced, leading to more informed decision-making and better preventive strategies in concrete construction.

4.8 Database structure and content

Based on the needs previously presented, a solid database has been established [67] based on the main exposed fields according to published data [31,43,44,46,49,52,53,55,58,60,95,96]. The compiled database facilitates a nuanced comparison of laboratory tests and real-world performance, addressing the challenges in accurately predicting AAR. It includes detailed records of laboratory and field tests, environmental conditions, aggregate properties, and concrete performance data. Additionally, each entry, among the 1385 recordings in the database, includes both laboratory and field performance data, allowing for a direct comparison and analysis of discrepancies.

Figure 4. 9 represents the structure and relationships of the database entities and Appendix B contains the extended version including each entity's attribute. In summary, the database is structured into five main entities: Material Selection, Mix Information, Field Data, Laboratory Data, and Performance Metrics. Each one contains specific attributes that capture essential details related to AAR development.

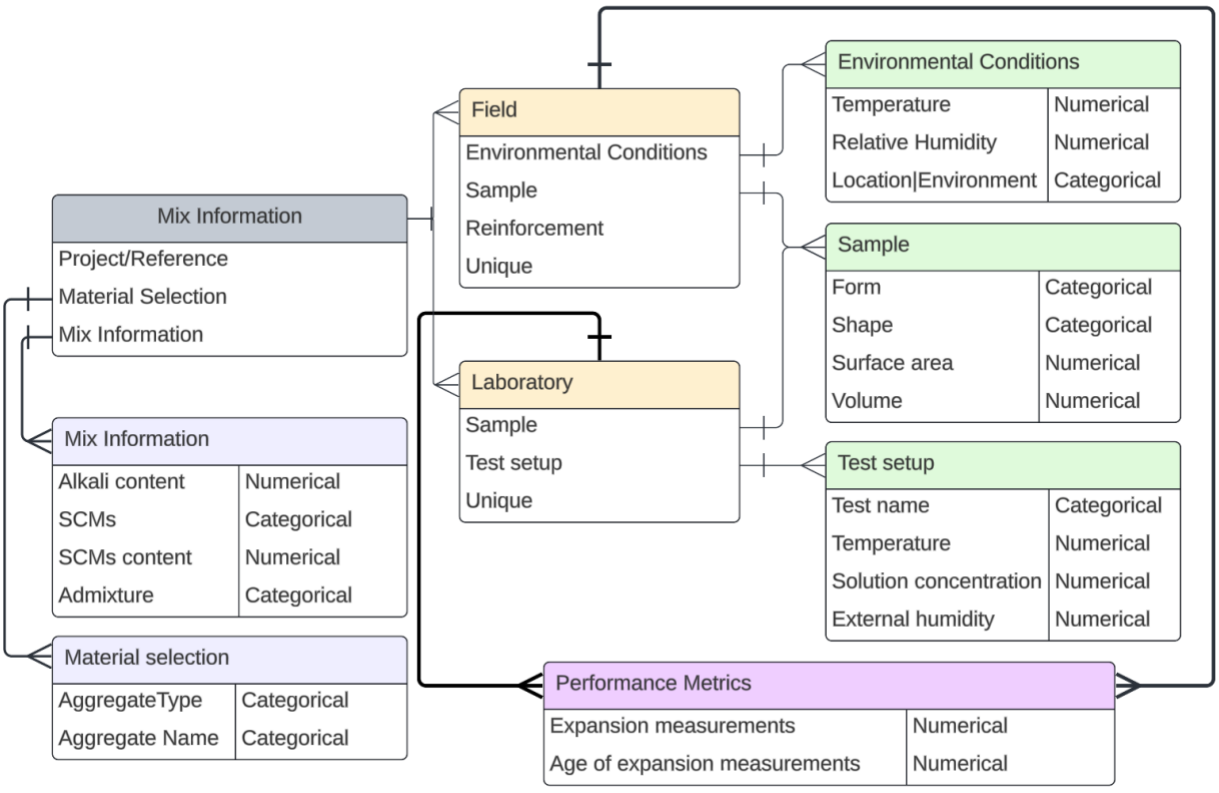


Figure 4. 9 – Database relationship diagram.

Material Selection includes attributes for aggregate type and aggregate name. Aggregate type (Figure 4. 10a) is categorical data indicating whether the aggregate is coarse, fine, or a combination of both. Aggregate name is categorical data referring to the origin of the aggregate (Figure 4. 10e), such as Alberta, Capitol - Austin, Texas; Cheyenne, WY; Daleville - Moscow, PA; DDS - Cleveland, Texas; Eagle Lake, Texas; El Indio, Texas; Fordyce - Mission, Texas; Hanson Little-River - Ashtown, AR; Helotes, Texas; Hylas - Rockville, VA; Jobe - El Paso, Texas; LG Everist - Dell Rapids, SD; Maryland - North East, MD; New Brunswick; New Ulm, MN; North Garden, VA; Nova Scotia; Omaha, NE; Ottersbo, Norway; Pioneer - Amarillo, Texas; Placitas - Albuquerque, New Mexico; Quebec; Richard Spur - Elgin, OK; San Antonio, Texas (NR); Spratt, ON; Stone Cold - Victoria, Texas; Stone Cold - Victoria, Texas + Wright - Robstown, Texas; Sudburry, ON; Wright - Robstown, Texas. The diversity of sources on aggregate information provides a baseline for assessing reactivity potential according to their respective mineralogy.

Alkali content (Figure 4. 10b) and SCMs content (Figure 4. 10c) are critical numerical parameters influencing AAR [27,34]. The histograms show a wide range of values, representing alkali content

in the concrete mix ranging from 0.32% to 1.29% (Figure 4. 10b) and employment of SCMs ranging from 0.0% to 76.0%, indicating the variability in mix designs in the dataset. This variability is essential for developing targeted mitigation strategies, as different alkali and SCMs contents have been shown to influence AAR progression. The categorical breakdown of SCMs types (Figure 4. 10d) and admixtures (Figure 4. 10f) highlights the complexity and diversity of materials used, which is crucial for addressing further options for mitigation strategies [97,98].

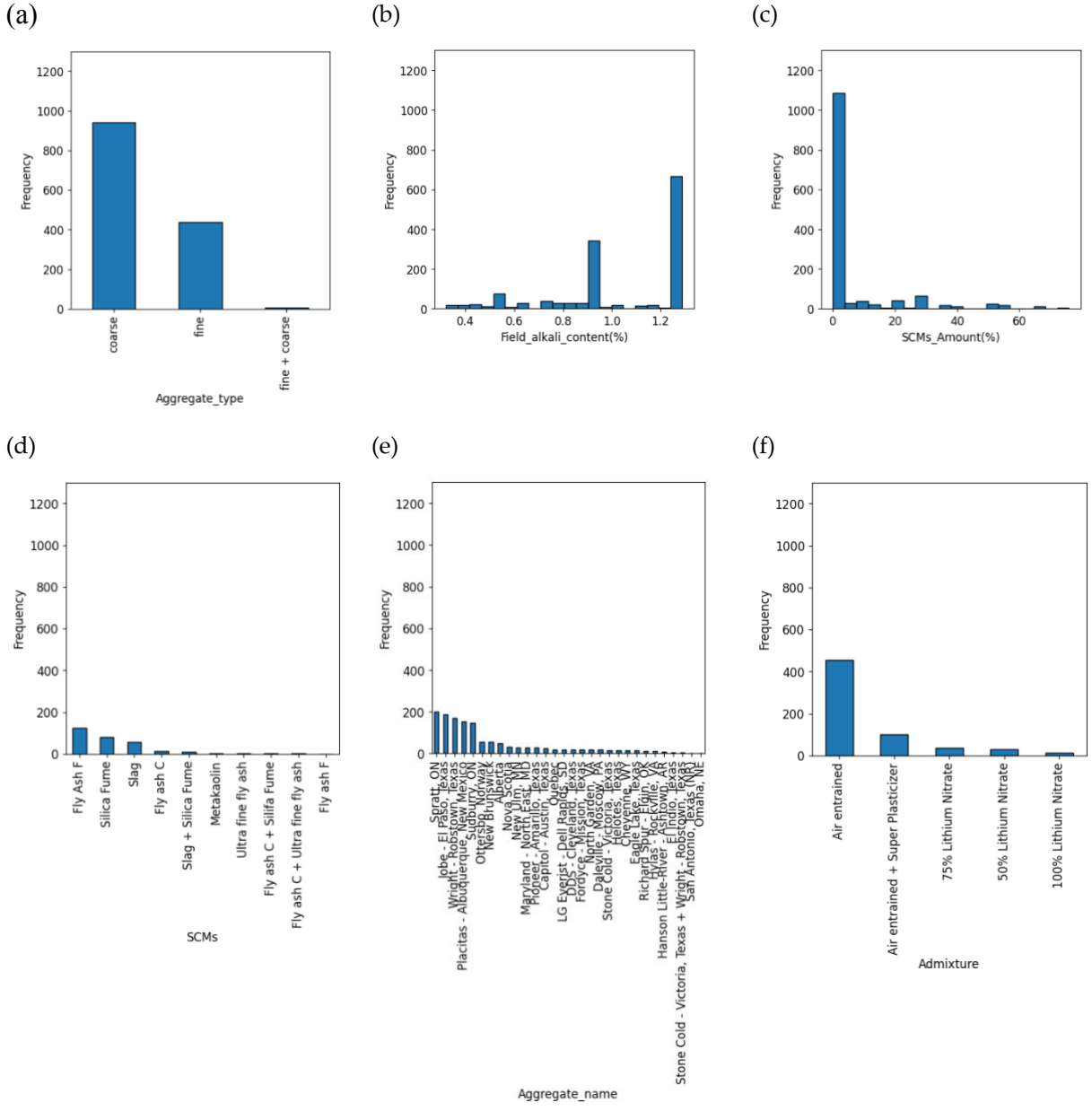


Figure 4. 10 – Histogram for the data distribution regarding (a) aggregate type, (b) field alkali content, (c) SCMs content, (d) SCMs type, (e) aggregate name, and (f) admixture.

Field data, including sample form (Figure 4. 11a), surface area (Figure 4. 11b), and volume (Figure 4. 11c), provide insights into the physical attributes of samples used in real-world applications. The environmental conditions, such as average temperature (Figure 4. 11d) and relative humidity (Figure 4. 11e), alongside the geographical location of samples (Figure 4. 11f), illustrate the diverse conditions under which AAR can occur (i.e., warm versus cold weather). This information allows for considering dynamic interactions between aggregates and the concrete matrix under varying environmental conditions, a limitation observed in traditional petrographic and chemical methods [1,2].

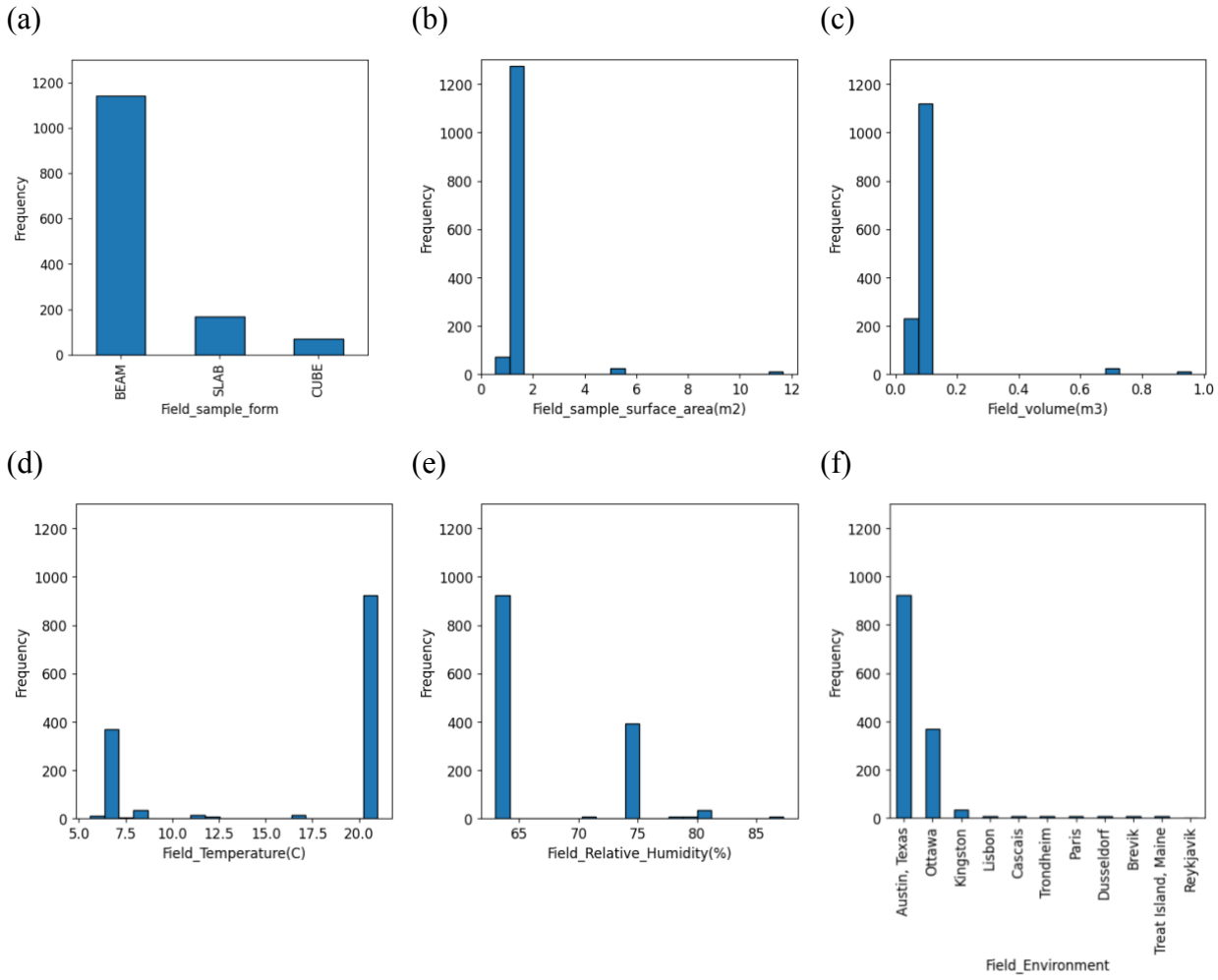


Figure 4. 11 – Histogram for the data distribution regarding field data for (a) sample form, (b) sample surface area, (c) sample volume, (d) temperature, (e) relative humidity, and (f) location.

The laboratory data segment, represented by test name (Figure 4. 12a), temperature (Figure 4. 12b), solution concentration (Figure 4. 12c), and sample attributes (Figure 4. 12d, Figure 4. 12e), provides a comprehensive overview of the conditions under which laboratory tests are conducted. The diversity in test setups and conditions, as shown in the histograms, helps address the criticism of existing test methods which may not accurately represent field conditions [31,32,69]. Additionally, by evaluating these variables along with test outcomes, the correlation between laboratory tests and actual field performance can be improved, reducing the risk of false positives and negatives [5,37].

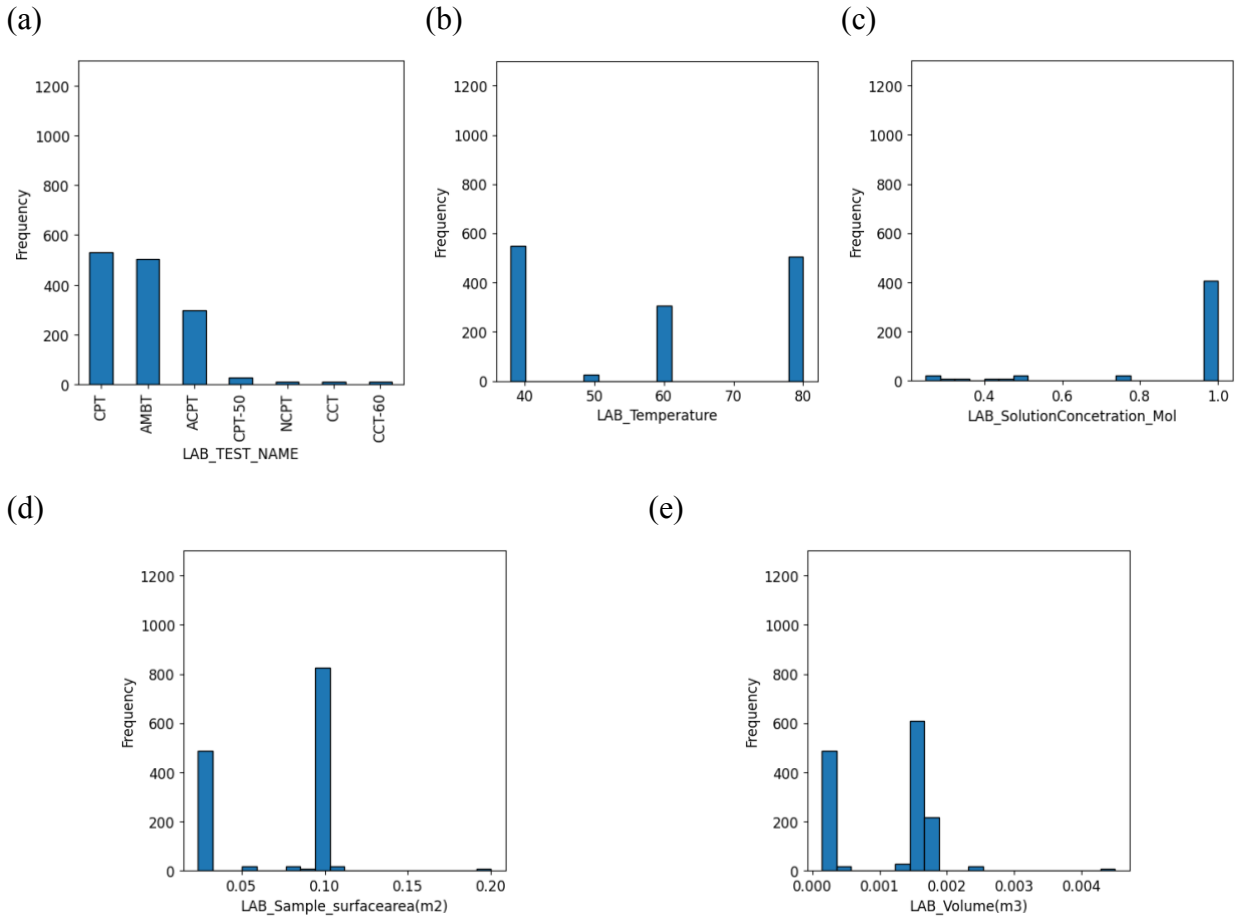


Figure 4. 12 – Histogram for the data distribution regarding accelerated laboratory data for (a) test name, (b) temperature, (c) solution concentration, (d) sample surface area, and (e) sample volume.

Performance metrics, particularly expansion measurements (Figure 4. 13a, Figure 4. 13c), offer a direct quantifiable measure of AAR impact. The data shows a broad range of expansion values, emphasizing the variability in AAR behaviour across different mixes and conditions. Field last expansion ranges from -0.049% to 1.358%, while laboratory outcome expansion ranges from -0.007% to 1.15%. Since the comparative analysis of laboratory and field expansion measurements reveals discrepancies, the broad range of the data can help enhancing accelerated tests predictive accuracy.

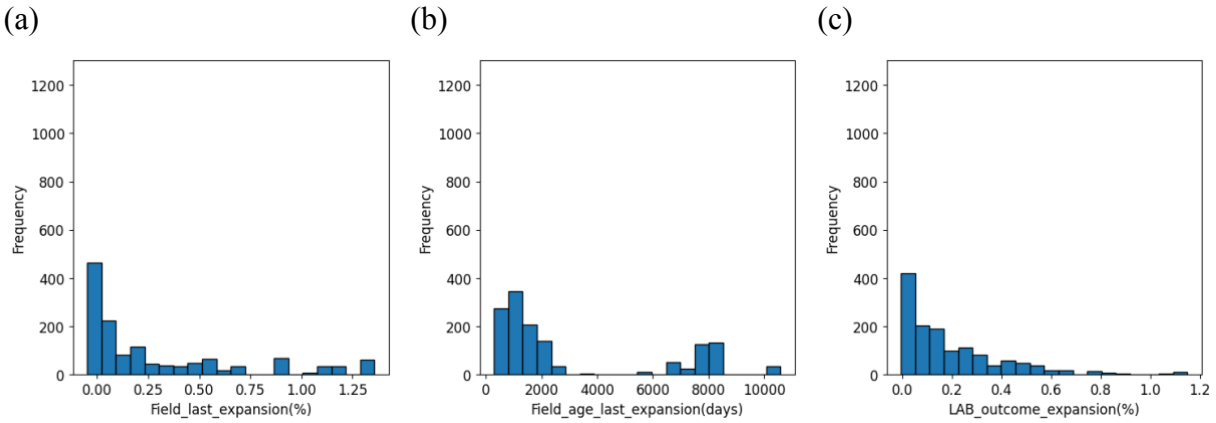


Figure 4. 13 – Histogram for the data distribution regarding performance metrics for (a) field last expansion, (b) age of last field expansion, and (c) laboratory outcome expansion.

Leveraging this comprehensive dataset allows for more robust statistical analyses and the development of predictive models that assertively account for the factors influencing AAR. Future research should focus on refining existing methodologies outcomes to reliably translat laboratory results into field performance predictions. This robust database consolidating data from various tests will further enhance the understanding and prevention of AAR, contributing to more durable and sustainable concrete infrastructure.

4.9 Conclusions

This study aimed to perform a comprehensive bibliometric analysis to critically review current laboratory and field methods for assessing AAR development. Additionally, it sought to establish a robust database integrating records of laboratory and field outcomes, supporting future research

in mitigating the risk associated with AAR development in new structures. The following conclusions were derived from the investigations:

- In addition to correlating laboratory and field performance of AAR-affected concrete, the study identified key research topics within the evaluated documents. This indicates trends for developing new accelerated laboratory procedures that vary storage conditions, mix designs, and sample attributes, as well as evaluate mitigation strategies.
- Regarding the evaluation of aggregate reactivity, the Accelerated Mortar Bar Test (AMBT) and Concrete Prism Test (CPT) are the most widely used, accounting for 30.3% and 36.4% of research occurrences, respectively. These tests were also the most commonly compared when evaluating block fields, highlighting their prominence in current testing methodologies.
- The extensive bibliometric review underscores the discrepancies between the AMBT and CPT. While both tests are critical for evaluating aggregate reactivity, there is a need for a reliable framework to accurately predict AAR occurrence in new structures.
- The current approaches to interpreting extensive data generated from outdoor fields remain largely descriptive. Future research should leverage advanced data analysis methods, including stochastic and artificial intelligence approaches, to enhance predictive models and ensure more reliable outcomes.

4.10 References

[1] B. Fournier, M.-A. Bérubé, Alkali–aggregate reaction in concrete: a review of basic concepts and engineering implications, 27 (2000).

[2] P.J. Nixon, I. Sims, eds., RILEM Recommendations for the Prevention of Damage by Alkali-Aggregate Reactions in New Concrete Structures, Springer Netherlands, Dordrecht, 2016. https://doi.org/10.1007/978-94-017-7252-5.

[3] M.G. Alexander, Alkali–aggregate reaction, in: Developments in the Formulation and Reinforcement of Concrete, Elsevier, 2019: pp. 87–113. https://doi.org/10.1016/B978-0-08-102616-8.00004-6.

[4] I. Sims, P. Nixon, RILEM Recommended Test Method AAR-0: Detection of Alkali-Reactivity Potential in Concrete—Outline guide to the use of RILEM methods in assessments

of aggregates for potential alkali-reactivity, Mat. Struct. 36 (2003) 472–479. https://doi.org/10.1007/BF02481527.

[5] B. Fournier, J. Lindgård, B.J. Wigum, I. Borchers, Outdoor exposure site testing for preventing Alkali-Aggregate Reactivity in concrete – a review., MATEC Web Conf. 199 (2018) 03002. https://doi.org/10.1051/matecconf/201819903002.

[6] J. Lindgård, Ö. Andiç-Çakır, I. Fernandes, T.F. Rønning, M.D.A. Thomas, Alkali–silica reactions (ASR): Literature review on parameters influencing laboratory performance testing, Cement and Concrete Research 42 (2012) 223–243. https://doi.org/10.1016/j.cemconres.2011.10.004.

[7] J. Lindgård, T. Østnor, B. Fournier, Ø. Lindgård, T. Danner, G. Plusquellec, K. De Weerdt, Determining alkali leaching during accelerated ASR performance testing and in field exposed cubes using cold water extraction (CWE) and μXRF, MATEC Web Conf. 199 (2018) 03004. https://doi.org/10.1051/matecconf/201819903004.

[8] B. Fournier, V. Malhotra, Evaluation of Laboratory Test Methods for Alkali-Silica Reactivity, Cement, Concrete, Aggr. 21 (1999) 173. https://doi.org/10.1520/CCA10431J.

[9] ASTM C295/C295M, Guide for Petrographic Examination of Aggregates for Concrete, ASTM International, 2019. https://doi.org/10.1520/C0295_C0295M-19.

[10] A.B. Poole, I. Sims, D.A. St John, Concrete petrography: a handbook of investigative techniques, Second edition, CRC Press, Boca Raton, 2019.

[11] C09 Committee, ASTM C856-20 Practice for Petrographic Examination of Hardened Concrete, (2020). https://doi.org/10.1520/C0856_C0856M-20.

[12] I. Fernandes, M.D.A. Ribeiro, M.A.T.M. Broekmans, I. Sims, eds., Petrographic Atlas: Characterisation of Aggregates Regarding Potential Reactivity to Alkalis, Springer Netherlands, Dordrecht, 2016. https://doi.org/10.1007/978-94-017-7383-6.

[13] CSA A23.2-27A, Standard Practices to identify degree of alkali-reactivity of aggregates and to identify measures to avoid deleterious expansion in concrete., in: CSA Standards A23.1–09/A23.2–09 Concrete Materials and Methods of Concrete Construction/Test Methods and Standard Practices for Concrete, 13th edition, CSA Group, Ontario, Canada, 2019: pp. 594–610.

[14] ASTM C289-07 Standard Test Method for Potential Alkali-Silica Reactivity of Aggregates (Chemical Method) (Withdrawn 2016), (2007).

[15] C09 Committee, ASTM C227-10 Test Method for Potential Alkali Reactivity of Cement-Aggregate Combinations (Mortar-Bar Method), (2010). https://doi.org/10.1520/C0227-10.

[16] M.D.A. Thomas, B. Fournier, K.J. Folliard, Alkali-Aggregate Reactivity (AAR) Facts Book, Book (2013) 224.

[17] ASTM C33-23 Specification for Concrete Aggregates, (2023). http://www.astm.org/cgi-bin/resolver.cgi?C33-03 (accessed August 6, 2024).

[18] M.-A. Bérubé, J. Frenette, Testing Concrete for AAR in NaOH and NaCl solutions at 38°C and 80°C, Cement and Concrete Composites 16 (1994) 189–198. https://doi.org/10.1016/0958-9465(94)90016-7.

[19] M.-A. Bérubé, B. Fournier, Canadian experience with testing for alkali-aggregate reactivity in concrete, Cement and Concrete Composites 15 (1993) 27–47. https://doi.org/10.1016/0958-9465(93)90037-A.

[20] R.E. Oberholster, G. Davies, An accelerated method for testing the potential alkali reactivity of siliceous aggregates, Cement and Concrete Research 16 (1986) 181–189. https://doi.org/10.1016/0008-8846(86)90134-1.

[21] ASTM C1260, Test Method for Potential Alkali Reactivity of Aggregates (Mortar-Bar Method), ASTM International, 2022. https://doi.org/10.1520/C1260-22.

[22] R.D. Hooton, C.A. Rogers, Development of the NBRI rapid mortar bar test leading to its use in North America, Construction and Building Materials 7 (1993) 145–148. https://doi.org/10.1016/0950-0618(93)90051-D.

[23] C. Lee, C.C. Liu, C.W. Wang, An Accelerated Concrete Prism Soaking Test for Evaluating the Alkali-Reactivity of Aggregates, in: Beijing, China, 2004.

[24] A. Shayan, H. Morris, A comparison of RTA T363 and ASTM C1260 accelerated mortar bar test methods for detecting reactive aggregates, Cement and Concrete Research 31 (2001) 655–663. https://doi.org/10.1016/S0008-8846(00)00491-9.

[25] B. Fournier, P.C. Nkinamubanzi, D. Lu, M.D.A. Thomas, K.J. Folliard, J.H. Ideker, Evaluating Potential Alkali-Reactivity of Concrete Aggregates – How Reliable are the Current and New Test Methods?, in: IBRACON, Rio de Janeiro, Brasil, 2006.

[26] ASTM C1778, Guide for Reducing the Risk of Deleterious Alkali-Aggregate Reaction in Concrete, ASTM International, 2022. https://doi.org/10.1520/C1778-22.

[27] S. Multon, M. Cyr, A. Sellier, P. Diederich, L. Petit, Effects of aggregate size and alkali content on ASR expansion, Cement and Concrete Research 40 (2010) 508–516. https://doi.org/10.1016/j.cemconres.2009.08.002.

[28] B. Lothenbach, K. De Weerdt, D. Hooton, J. Duchesne, A. Leemann, Can We Relate ASR Expansion to the Pore Solution Composition?, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 104–112. https://doi.org/10.1007/978-3-031-59419-9_13.

[29] B. Fournier, M.A. Bérubé, Application of the NBRI accelerated mortar bar test to siliceous carbonate aggregates produced in the St. Lawrence Lowlands (Quebec, Canada) part 2: Proposed limits, rates of expansion, and microstructure of reaction products, Cement and Concrete Research 21 (1991) 1069–1082. https://doi.org/10.1016/0008-8846(91)90067-R.

[30] J. Lindgård, B. Fournier, T.F. Rønning, M.D.A. Thomas, Alkali–aggregate reaction: performance testing, exposure sites and regulations, Proceedings of the Institution of Civil Engineers - Construction Materials 169 (2016) 189–196. https://doi.org/10.1680/jcoma.15.00077.

[31] B. Fournier, J.H. Ideker, K.J. Folliard, M.D.A. Thomas, P.-C. Nkinamubanzi, R. Chevrier, Effect of environmental conditions on expansion in concrete due to alkali–silica reaction (ASR), Materials Characterization 60 (2009) 669–679. https://doi.org/10.1016/j.matchar.2008.12.018.

[32] M. Thomas, B. Fournier, K. Folliard, J. Ideker, M. Shehata, Test methods for evaluating preventive measures for controlling expansion due to alkali–silica reaction in concrete, Cement and Concrete Research 36 (2006) 1842–1856. https://doi.org/10.1016/j.cemconres.2006.01.014.

[33] ASTM C1293, Test Method for Determination of Length Change of Concrete Due to Alkali-Silica Reaction, (2024). https://doi.org/10.1520/C1293_C1293M-23A.

[34] B. Fournier, R. Chevrier, A. Bilodeau, P.-C. Nkinamubanzi, N. Bouzoubaa, COMPARATIVE FIELD AND LABORATORY INVESTIGATIONS ON THE USE OF SUPPLEMENTARY CEMENTING MATERIALS (SCMs) TO CONTROL ALKALI-SILICA REACTION (ASR) IN CONCRETE, in: Brazil, 2016.

[35] J.H. Ideker, B.L. East, K.J. Folliard, M.D.A. Thomas, B. Fournier, The current state of the accelerated concrete prism test, Cement and Concrete Research 40 (2010) 550–555. https://doi.org/10.1016/j.cemconres.2009.08.030.

[36] AASHTO T380. (2019) Standard Method of Test for Potential Alkali Reactivity of Aggregates and Effectiveness of ASR Mitigation Measures (Miniature Concrete Prism Test, MCPT)., (2019).

[37] J. Tanesi, T. Drimalas, K.S.T. Chopperla, M. Beyene, J.H. Ideker, H. Kim, L. Montanari, A. Ardani, Divergence between Performance in the Field and Laboratory Test Results for Alkali-Silica Reaction, Transportation Research Record 2674 (2020) 120–134. https://doi.org/10.1177/0361198120913288.

[38] A. Ghanizadeh, M. Thomas, T. Drimalas, K.S. Teja, A. Parashar, J.H. Ideker, R. Lute, K. Folliard, Using the Miniature Concrete Prism Test (MCPT) to evaluate ASR preventive measures, (2022).

[39] E.R. Latifee, P.R. Rangaraju, Miniature Concrete Prism Test: Rapid Test Method for Evaluating Alkali-Silica Reactivity of Aggregates, J. Mater. Civ. Eng. 27 (2015) 04014215. https://doi.org/10.1061/(ASCE)MT.1943-5533.0001183.

[40] NB 32, Alkali–aggregate reactions in concrete, Test methods and Requirements to Test Laboratories, (2005).

[41] T.F. Rønning, J. Lindgård, I. Borchers, ASR performance testing concepts – RILEM AAR-10, AAR-11 and AAR-12, (2022).

[42] T.F. Ronning, B.J. Wigum, J. Lindgard, Recommendation of RILEM TC 258-AAA: RILEM AAR-10: determination of binder combinations for non-reactive mix design using concrete prisms-38 degrees C test method, MATERIALS AND STRUCTURES 54 (2021). https://doi.org/10.1617/s11527-021-01679-w.

[43] B.J. Wigum, L.T. Pedersen, B. Grelk, J. Lindgård, State-of-the art report: Key parameters influencing the alkali aggregate reaction, SINTEF Building and Infrastructure, 2006. https://www.sintef.no/globalassets/upload/byggforsk/partner/report-2.1-final-a06018.pdf.

[44] J. Custódio, J. Lindgård, B. Fournier, A. Santos Silva, M.D.A. Thomas, T. Drimalas, J.H. Ideker, R.-P. Martin, I. Borchers, B. Johannes Wigum, T.F. Rønning, Correlating field and laboratory investigations for preventing ASR in concrete – The LNEC cube study (Part I –

Project plan and laboratory results), Construction and Building Materials 343 (2022) 128131. https://doi.org/10.1016/j.conbuildmat.2022.128131.

[45] I. Borchers, J. Lindgård, C. Müller, Evaluation of laboratory test methods for assessing the alkali-reactivity potential of aggregates by field site tests, Materconstrucc 72 (2022) e286. https://doi.org/10.3989/mc.2022.17221.

[46] J. Lindgård, P.J. Nixon, I. Borchers, B. Schouenborg, B.J. Wigum, M. Haugen, U. Åkesson, The EU "PARTNER" Project — European standard tests to prevent alkali reactions in aggregates: Final results and recommendations, Cement and Concrete Research 40 (2010) 611–635. https://doi.org/10.1016/j.cemconres.2009.09.004.

[47] Z. Xu, X. Lan, M. Deng, M. Tang, A new accelerated method for determining the potential alkali-carbonate reactivity, Cement and Concrete Research 32 (2002) 851–857. https://doi.org/10.1016/S0008-8846(01)00758-X.

[48] R.D. Hooton, C. Rogers, C.A. MacDonald, T. Ramlochan, Twenty-Year Field Evaluation of Alkali-Silica Reaction Mitigation, MJ 110 (2013). https://doi.org/10.14359/51685905.

[49] C.A. MacDonald, Kingston Outdoor Exposure Site for ASR –29 Year Update, (2020).

[50] B.J. Wigum, G.J. Einarsson, ALKALI AGGREGATE REACTION IN ICELAND RESULTS FROM LABORATORY TESTING COMPARED TO FIELD EXPOSURE SITE, (n.d.).

[51] M.D.A. Thomas, SUMMARY OF BRE RESEARCH ON THE EFFECT OF FLY ASH ON AIKALI-SILICA REACTION IN CONCRETE, in: Melbourne, Australia, 1996.

[52] B. Fournier, A. Bilodeau, N. Bouzoubaa, P.-C. Nkinamubanzi, Field and Laboratory Investigations on the Use of Fly Ash and Li-Based Admixtures to Prevent ASR in Concrete, in: United Kingdom, 2018.

[53] B. Fournier, A. Ferro, V. Sivasundaram, CANMET/Industry Research Consortium on Alkali-Aggregate Reactivity in Concrete, EPRI and CANMET, Ontario, Canada, 2001. https://www.epri.com/research/products/1004031.

[54] R. Bleszynski, R.D. Hooton, M.D.A. Thomas, C. Rogers, Durability of Ternary Blend Concrete with Silica Fume and Blast-Furnace Slag: Laboratory and Outdoor Exposure Site Studies, MJ 99 (2002). https://doi.org/10.14359/12329.

[55] I. Fernandes, A. Leemann, B. Fournier, E. Menendez, J. Lindgård, I. Borchers, J. Custódio, PARTNER project post-documentation study. Condition assessment of field exposure site

cubes. Results of microstructural analyses, Cement and Concrete Research 162 (2022) 107006. https://doi.org/10.1016/j.cemconres.2022.107006.

[56] I. Borchers, C. Müller, Field Site Tests Established in the Partner Project for Evaluating the Correlation Between Laboratory Tests and Field Performance.PDF, in: Trondheim, Norway, 2008.

[57] I. Fernandes, M.A.T.M. Broekmans, P. Nixon, I. Sims, M. dos A. Ribeiro, F. Noronha, B. Wigum, Alkali-silica reactivity of some common rock types. A global petrographic atlas, QUARTERLY JOURNAL OF ENGINEERING GEOLOGY AND HYDROGEOLOGY 46 (2013) 215–220. https://doi.org/10.1144/qjegh2012-065.

[58] T. Drimalas, K.J. Folliard, J.H. Ideker, Findings from the University of Texas at Austin ASR Exposure Site after 20 Years, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 516–523. https://doi.org/10.1007/978-3-031-59349-9_59.

[59] S. Hayman, M. Thomas, N. Beaman, P. Gilks, Selection of an effective ASR-prevention strategy for use with a highly reactive aggregate for the reconstruction of concrete structures at Mactaquac generating station, Cement and Concrete Research 40 (2010) 605–610. https://doi.org/10.1016/j.cemconres.2009.08.015.

[60] J. Lindgård, T.F. Rønning, M.D.A. Thomas, B. Fournier, A.S. Silva, ASR - PERFORMANCE TESTING: MAIN FINDINGS IN THE NORWEGIAN COIN PROJECT, in: Sao Paulo – Brazil, 2016.

[61] M.D.A. Thomas, K.J. Folliard, F. Benoit, T. Drimalas, S. Garber, Methods for Preventing ASR in New Construction: Results of Field Exposure Sites, FHWA Office of Pavement Technology, 2013.

[62] S.U. Einarsdóttir, Modifications to Laboratory Test Methods to Evaluate the Beneficial Effects of Low-Alkali Binders on the Alkali-Silica Reaction, University of Toronto, 2017.

[63] C. Pelissou, B. Durville, S. Morin, Main Outcomes for AAR Studies on Large Scale Experiments in the ODOBA International Project, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 566–576. https://doi.org/10.1007/978-3-031-59349-9_65.

[64] K.K. Hansen, B. Grelk, R. Kofoed, J.B. Mørk, Field Investigation of Concrete Blocks with Different ASR Reactive Aggregate Types with and Without Surface Hydrophobic Impregnation, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 550–557. https://doi.org/10.1007/978-3-031-59419-9_64.

[65] A. Leemann, D. Sirtoli, J. Justs, M. Wyrzykowski, Establishing Two ASR Exposure Sites in Switzerland – Approach and Goals, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 532–539. https://doi.org/10.1007/978-3-031-59349-9_61.

[66] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, W.M. Lim, How to conduct a bibliometric analysis: An overview and guidelines, Journal of Business Research 133 (2021) 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070.

[67] A. Bergmann, L.F.M. Sanchez, Alkali-Aggregate Reaction (AAR) Comparative Data: Accelerated Laboratory Outcomes vs. Field Performance, (2024). https://doi.org/10.17632/5cscpggnhg.1.

[68] C.A. MacDonald, C. Rogers, R.D. Hooton, THE RELATIONSHIP BETWEEN LABORATORY AND FIELD EXPANSION – OBSERVATIONS AT THE KINGSTON OUTDOOR EXPOSURE SITE FOR ASR AFTER TWENTY YEARS, in: Austin, Texas, USA, 2012.

[69] J.H. Ideker, A.F. Bentivegna, K.J. Folliard, M.C.G. Juenger, Do Current Laboratory Test Methods Accurately Predict Alkali-Silica Reactivity?, MJ 109 (2012). https://doi.org/10.14359/51683914.

[70] Y. Kawabata, K. Yamada, T. Kawakami, Y. Sagawa, Environmental impacts on expansion of concrete due to alkali–silica reaction, Magazine of Concrete Research (2022) 1–11. https://doi.org/10.1680/jmacr.22.00158.

[71] D. Hooton, B. Fournier, Long-term alkali–silica mitigation of high-alkali concrete with cement replacements, Proceedings of the Institution of Civil Engineers - Construction Materials 175 (2022) 125–136. https://doi.org/10.1680/jcoma.21.00049.

[72] K.S.T. Chopperla, T. Drimalas, M. Beyene, J. Tanesi, K. Folliard, A. Ardani, J.H. Ideker, Combining reliable performance testing and binder properties to determine preventive

measures for alkali-silica reaction, Cement and Concrete Research 151 (2022) 106641. https://doi.org/10.1016/j.cemconres.2021.106641.

[73] J.F. Muñoz, C. Balachandran, T.S. Arnold, New Turner-Fairbank Alkali-Silica Reaction Susceptibility Test for Aggregate Evaluation, Transportation Research Record 2675 (2021) 798–808. https://doi.org/10.1177/03611981211004584.

[74] S. Yang, J.-X. Lu, C.S. Poon, Recycling of waste glass in dry-mixed concrete blocks: Evaluation of alkali-silica reaction (ASR) by accelerated laboratory tests and long-term field monitoring, Construction and Building Materials 262 (2020) 120865. https://doi.org/10.1016/j.conbuildmat.2020.120865.

[75] R.H. Haddad, A. Al-Sayed, Bond behavior between ASR-damaged concrete and CFRP sheets: Empirical modeling, Journal of Building Engineering 29 (2020) 101166. https://doi.org/10.1016/j.jobe.2019.101166.

[76] Y. Kawabata, K. Yamada, Y. Sagawa, S. Ogawa, Alkali-Wrapped Concrete Prism Test (AW-CPT) – New Testing Protocol Toward a Performance Test against Alkali-Silica Reaction–, ACT 16 (2018) 441–460. https://doi.org/10.3151/jact.16.441.

[77] I. Robertson, L. Shen, Field Evaluation of Concrete using Hawaiian Aggregates for Alkali Silica Reaction, MATEC Web Conf. 199 (2018) 03005. https://doi.org/10.1051/matecconf/201819903005.

[78] R.A. Deschenes, W. Micah Hale, Alkali-Silica Reaction in Concrete with Previously Inert Aggregates, J. Perform. Constr. Facil. 31 (2017) 04016084. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000946.

[79] V. Sirivivatnanon, J. Mohammadi, W. South, Reliability of new Australian test methods in predicting alkali silica reaction of field concrete, Construction and Building Materials 126 (2016) 868–874. https://doi.org/10.1016/j.conbuildmat.2016.09.055.

[80] G. Giaccio, M.C. Torrijos, C. Milanesi, R. Zerbino, Alkali–silica reaction in plain and fibre concretes in field conditions, Mater Struct 52 (2019) 31. https://doi.org/10.1617/s11527-019-1332-2.

[81] Y. Kawabata, K. Yamada, Evaluation of Alkalinity of Pore Solution Based on the Phase Composition of Cement Hydrates with Supplementary Cementitious Materials and its Relation to Suppressing ASR Expansion, ACT 13 (2015) 538–553. https://doi.org/10.3151/jact.13.538.

[82] A. Santos Silva, D. Soares, L. Matos, I. Fernandes, M.M. Salta, Alkali-Aggregate Reactions in Concrete: Methodologies Applied in the Evaluation of Alkali Reactivity of Aggregates for Concrete, MSF 730–732 (2012) 409–414. https://doi.org/10.4028/www.scientific.net/MSF.730-732.409.

[83] C.S. Lam, C.S. Poon, D. Chan, Enhancing the performance of pre-cast concrete blocks by incorporating waste glass – ASR consideration, Cement and Concrete Composites 29 (2007) 616–625. https://doi.org/10.1016/j.cemconcomp.2007.03.008.

[84] M.D.A. Thomas, B. Fournier, K.J. Folliard, M.H. Shehata, J.H. Ideker, C. Rogers, Performance Limits for Evaluating Supplementary Cementing Materials Using Accelerated Mortar Bar Test, MJ 104 (2007). https://doi.org/10.14359/18573.

[85] N. Smaoui, M.-A. Bérubé, B. Fournier, B. Bissonnette, B. Durand, Evaluation of the expansion attained to date by concrete affected by alkali–silica reaction. Part I: Experimental study, Can. J. Civ. Eng. 31 (2004) 826–845. https://doi.org/10.1139/l04-051.

[86] N. Smaoui, B. Fournier, M.-A. Bérubé, B. Bissonnette, B. Durand, Evaluation of the expansion attained to date by concrete affected by alkalisilica reaction. Part II: Application to nonreinforced concrete specimens exposed outside, Can. J. Civ. Eng. 31 (2004) 997–1011. https://doi.org/10.1139/l04-074.

[87] M.-A. Berube, J. Frenette, A. Pedneault, M. Rivest, Laboratory Assessment of the Potential Rate of ASR Expansion of Field Concrete, Cement, Concrete, Aggr. 24 (2002) 13. https://doi.org/10.1520/CCA10486J.

[88] D.P. DeMerchant, B. Fournier, F. Strang, Alkali–aggregate research in New Brunswick, 27 (2000).

[89] D. Lane, Comparison of Results from C 441 and C 1293 with Implications for Establishing Criteria for ASR-Resistant Concrete, Cement, Concrete, Aggr. 21 (1999) 149. https://doi.org/10.1520/CCA10428J.

[90] N. Whiting, Comparison of Field Observations with Laboratory Test Results on Concretes Undergoing Alkali-Silica Reaction, Cement, Concrete, Aggr. 21 (1999) 142. https://doi.org/10.1520/CCA10427J.

[91] A. Shayan, Prediction of Alkali Reactivity Potential of Some Australian Aggregates and Correlation With Service Performance, MJ 89 (1993). https://doi.org/10.14359/1240.

[92] E.R. Giannini, A.F. Bentivegna, K.J. Folliard, Coatings and overlays for concrete affected by alkali-silica reaction, in: Dresden, Germany, 2011.

[93] S. Bianchi, F. Biondini, Life-Cycle Assessment of Deteriorating RC Bridges Using Artificial Neural Networks, J. Infrastruct. Syst. 28 (2022) 04022005. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000672.

[94] M. Pregnolato, S. Gunner, E. Voyagaki, R. De Risi, N. Carhart, G. Gavriel, P. Tully, T. Tryfonas, J. Macdonald, C. Taylor, Towards Civil Engineering 4.0: Concept, workflow and application of Digital Twins for existing infrastructure, Automation in Construction 141 (2022) 104421. https://doi.org/10.1016/j.autcon.2022.104421.

[95] T. Drimalas, K.J. Folliard, A. Parashar, M.D.A. Thomas, A. Ghanizadeh, A.M. Hossack, B. Fournier, National Cooperative Highway Research Program, Transportation Research Board, National Academies of Sciences, Engineering, and Medicine, Alkali-Silica Reactivity Potential and Mitigation: Test Methods and State of Practice, Transportation Research Board, Washington, D.C., 2023. https://doi.org/10.17226/27435.

[96] K.J. Folliard, R. Barborak, T. Drimalas, L. Du, S. Garber, J. Ideker, T. Ley, S. Williams, M. Juenger, B. Fournier, M.D.A. Thomas, Preventing ASR/DEF in New Concrete: Final Report (0-4085-5), Texas Department of Transportation and the Federal Highway Administration, 2006.

[97] A. Leemann, L. Bernard, S. Alahrache, F. Winnefeld, ASR prevention — Effect of aluminum and lithium ions on the reaction products, Cement and Concrete Research 76 (2015) 192–201. https://doi.org/10.1016/j.cemconres.2015.06.002.

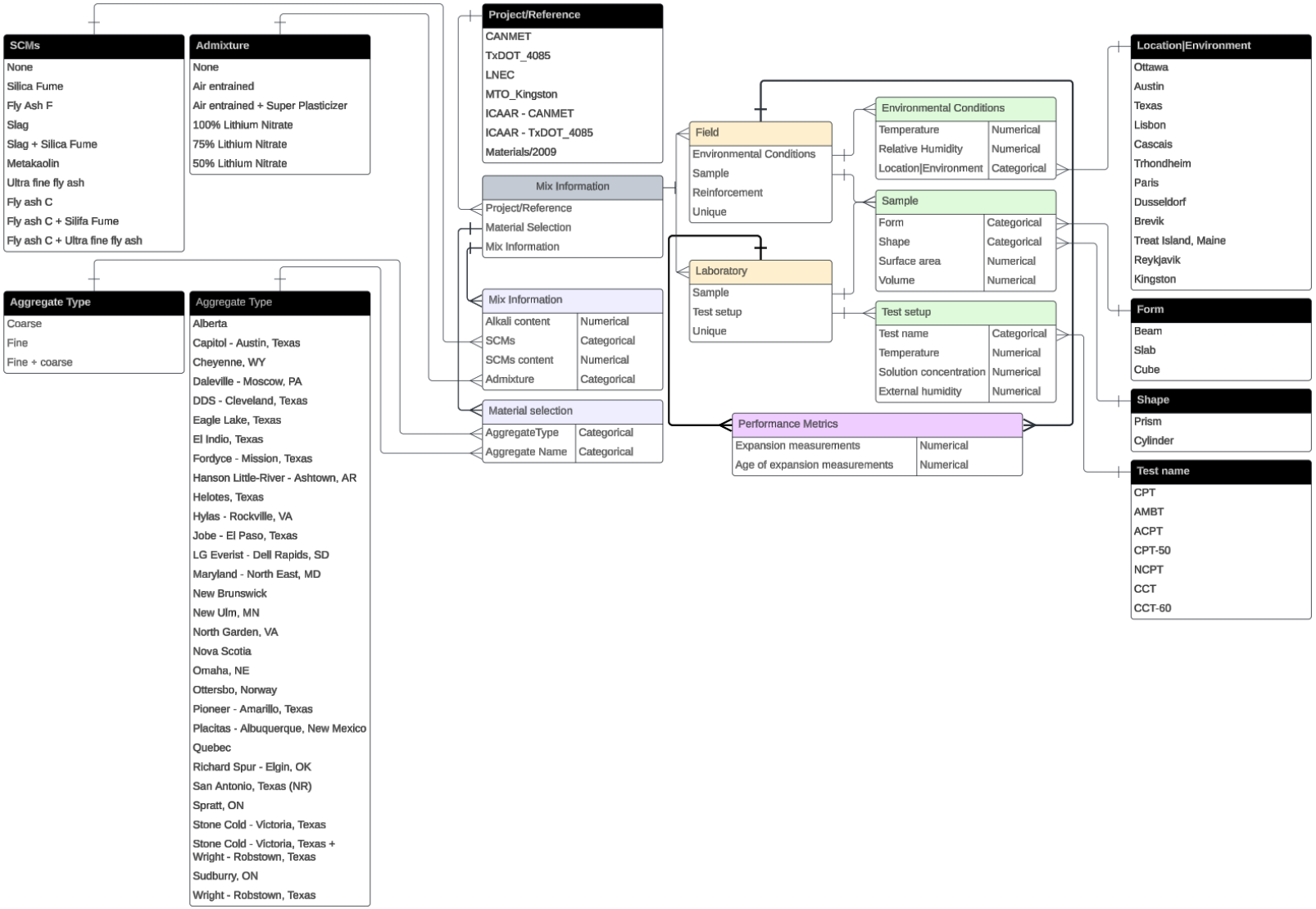
[98] K. de Weerdt, M. Ranger, M.E. Krüger, A. Bergmann, P. Hemstad, A. Leemann, B. Lothenbach, Impact of SCMs on alkali level and w/b ratio on alkali concentration in pore solution, Materials and Structures ((submited)).

4.11 Appendix A – List of papers resulting from the bibliometric analysis

REF	Authors	Countries/Affiliation	Publication Year	Citation Count
[15]	Fernandes, I; Leemann, A; Fournier, B; Menendez, E; Lindgard, J; Borchers, I; Custodio, J	Portugal; Switzerland/North Ireland; Canada; Spain; Norway; Germany; Portugal;	2022	35
[16]	Kawabata, Y; Yamada, K; Kawakami, T; Sagawa, Y	Japan; Japan; Japan; Japan	2022	29
[14]	Custodio, J; Lindgard, J; Fournier, B; Silva, AS; Thomas, MDA; Drimalas, T; Ideker, J; Martin, RP; Borchers, I; Wigum, BJ; Ronning, TF	Portugal; Norway; Canada; Portugal; Canada; USA; USA; France; Germany; Iceland; Norway	2022	23
[17]	Hooton, RD; Fournier, B	Canada; Canada	2022	23
[18]	Chopperla, KST; Drimalas, T; Beyene, M; Tanesi, J; Folliard, K; Ardani, A; Ideker, J	USA; USA; USA; USA; USA; USA	2022	57
[19]	Munoz, JF; Balachandran, C; Arnold, TS	USA; USA; USA	2021	38
[20]	Yang, SQ; Lu, JX; Poon, CS	China; China; China	2020	35
[21]	Haddad, RH; Al-Sayed, A	Jordan; Jordan	2020	45
[22]	Tanesi, J; Drimalas, T; Chopperla, KST; Beyene, M; Ideker, J; Kim, H; Montanari, L; Ardani, A	USA; USA; USA; USA; USA; USA; USA; USA	2020	28
[29]	Giaccio, G; Torrijos, MC; Milanesi, C; Zerbino, R	Argentina; Argentina; Argentina; Argentina	2019	56
[23]	Kawabata, Y; Yamada, K; Sagawa, Y; Ogawa, S	Japan; Japan; Japan; Japan	2018	69
[24]	Lindgard, J; Ostnor, T; Fournier, B; Lindgard, O; Danner, T; Plusquellec, G; De Weerdt, K	Norway; Norway; Canada; Norway; Norway; Norway; Norway	2018	9
[25]	Robertson, I; Shen, L	USA; USA	2018	5
[26]	Deschenes, RA; Hale, WM	USA; USA	2017	19
[27]	Sirivivatnanon, V; Mohammadi, J; South, W	Australia; Australia; Australia	2016	19
[28]	Lindgard, J; Fournier, B; Ronning, TF; Thomas, MDA	Norway; Canada; Norway; Canada	2016	34
[30]	Kawabata, Y; Yamada, K	Japan/France; Japan	2015	89
[31]	Hooton, RD; Rogers, CA; MacDonald, CA; Ramlochan, T	Canada; Canada; Canada; Canada	2013	23

[32]	Silva, AS; Soares, D; Matos, L; Fernandes, I; Salta, M	Portugal; Portugal; Portugal; Portugal; Portugal	2013	13
[33]	Ideker, J; Bentivegna, AF; Folliard, K; Juenger, MCG	USA; USA; USA; USA	2012	19
[47]	Giannini, ER; Bentivegna, AF; Folliard, K	USA; USA; USA	2012	12
[34]	Hayman, S; Thomas, MDA; Beaman, N; Gilks, P	Canada; Canada; Canada; Canada	2010	16
[35]	Fournier, B; Ideker, J; Folliard, K; Thomas, MDA; Nkinamubanzi, PC; Chevrier, R	Canada; USA; USA; Canada; Canada; Canada	2009	14
[36]	Lam, CS; Poon, CS; Chan, D	China; China; China	2007	15
[37]	Thomas, MDA; Fournier, B; Folliard, K; Shehata, MH; Ideker, J; Rogers, C	Canada; Canada; USA; Canada; USA; Canada	2007	26
[38]	Smaoui, N; Fournier, B; Berube, MA; Bissonnette, B; Durand, B	Canada; Canada; Canada; Canada; Canada	2004	10
[39]	Smaoui, N; Berube, MA; Fournier, B; Bissonnette, B; Durand, B	Canada; Canada; Canada; Canada; Canada	2004	26
[40]	Bleszynski, R; Hooton, RD; Thomas, MDA; Rogers, CA	Canada; Canada; Canada; Canada	2002	40
[41]	Berube, MA; Frenette, J; Pedneault, A; Rivest, M	Canada; Canada; Canada; Canada	2002	12
[42]	DeMerchant, DP; Fournier, B; Strang, F	Canada; Canada; Canada	2000	15
[43]	Fournier, B; Malhotra, VM	Canada; Canada	1999	10
[44]	Lane, DS	USA	1999	15
[45]	Whiting, NM	USA	1999	6
[46]	Shayan, A	Australia	1992	16

4.12 Appendix B – Database structure



Chapter 5: Assessing the reliability of laboratory test procedures for predicting concrete field performance against alkali-aggregate reaction (AAR)

Ana Bergmann and Leandro Sanchez

Abstract:

Alkali aggregate reaction (AAR) affected structures show reduced serviceability and premature distress in over 50 countries worldwide. Several laboratory test protocols have been proposed to evaluate the potential reactivity of aggregates by varying the conditions known to trigger and sustain the reaction. Among them, the most popular methods are the accelerated mortar bar test (AMBT) and the concrete prism test (CPT). Nevertheless, exposure site data, displaying the behaviour of concrete blocks exposed to real environmental conditions, has increased considerably recently, showing significant discrepancies between laboratory and concrete field performance. This study explores the reliability of laboratory tests, indicating moderate accuracy in predicting field performance for the AMBT and the CPT. The findings highlight an opportunity for recalibration of these methods through advanced analytical models that account for environmental conditions, alkali content, and the presence of SCMs to improve predictive accuracy. These measures will enhance concrete infrastructure safety by identifying risks associated with incorporating AAR-prone aggregates into new structures.

Keywords: Alkali-aggregate reaction (AAR), machine learning, field prediction, laboratory test, concrete durability.

5.1 Introduction

Alkali-aggregate reaction (AAR), a harmful durability mechanism affecting concrete structures worldwide, impacts the integrity and serviceability of structures in over 50 countries. This reaction, predominantly alkali-silica reaction (ASR), involves the interaction between silica phases in the aggregates and the alkali hydroxides present in the concrete pore solution; and originates secondary products, including both amorphous, gel-like phases, and

crystalline structures. These products can generate significant internal stresses, either through swelling upon water uptake or trough solidification pressure caused by confined crystallization and electrostatic repulsion. This leads to internal stresses and cracking, and reducing the mechanical properties of affected concrete [1–3].

It is widely accepted that the best approach to cope with ASR is to prevent the chemical reaction, for instance, by employing SCMs [1,4]. Therefore, reliable procedures to predict its occurrence are crucial for ensuring longer and safer concrete infrastructure. In this sense, conventional laboratory methods, such as the accelerated mortar bar test (AMBT) and the concrete prism test (CPT), are used for evaluating the potential reactivity of aggregates [1,5]. Although these methods are widely employed, they reveal discrepancies when their outcomes are compared to one another and with actual field performance [6,7]. This emphasizes the need to address such inconsistencies and improve the reliability of laboratory predictions for concrete against AAR.

A considerable increase in data from exposure concrete sites has been observed in recent years, accentuating the divergences between laboratory and field results [5,7–9]. As laboratory tests tend to accelerate the reaction process, they are influenced by various sources of errors and finish by not accounting for some critical real-world factors, such as environmental conditions [4,10]; hence, field-exposed elements are the most reliable source for selecting a ground truth reference for comparison regarding ASR development [10].

The enhanced availability of data and advanced data analysis techniques have prompted the investigation of the reliability of laboratory testing setups for field performance prediction. For instance, advanced data evaluation techniques from statistical classifications, such as confusion matrices, are known to accurately explore the performance of models in predicting the actual target, which could then offer insights regarding ASR development. More specifically, metrics such as accuracy, sensitivity, specificity, and precision provide detailed insights into the reliability of these tests. Accuracy reflects how well the accelerated test matches field results. Moreover, sensitivity measures its ability to detect true positives, while precision evaluates the correctness of positive predictions, minimizing false positives for AMBT and CPT. Specificity, on the other hand, measures the test's ability to correctly identify non-reactive cases, ensuring that false negatives are avoided.

The primary objective of this study is to examine the reliability of traditional laboratory methods in predicting ASR. By achieving this, the study collaborates to enhance the safety and suitable performance of concrete infrastructure by providing insights for a more accurate assessment of the risk associated with relying on accelerated tests for decision-making when incorporating AAR-prone aggregates in concrete.

5.2 Background

Once initiated, addressing AAR is both complex and uncertain, making prevention the most effective strategy to avoid such degradation. In this context, a wide number of laboratory test procedures have been developed to appraise the potential reactivity of aggregates and the efficiency of mitigation techniques before concrete pouring. These tests often differ in terms of specimen characteristics (e.g., shape, size), storage conditions (e.g., temperature, humidity, alkali levels), and mix compositions (e.g., crushed aggregates, added alkalis) [5,11]. The main goal is to accelerate the reaction kinetics, allowing for the rapid identification of aggregate reactivity, potential expansion of concrete mixtures, and the performance of preventive measures such as supplementary cementitious materials (SCMs). Amongst them, the accelerated mortar bar test (AMBT) and the concrete prism test (CPT) have been widely employed [12,13].

The AMBT, as per ASTM C1260, is performed by immersing mortar bars (i.e., 25x25x285 mm^3) made in a high-alkali solution (1M NaOH) at 80°C, with expansions measured over 14 to 28 days for mixtures without and with SCMs, respectively [12,14]. This accelerated method yields rapid results and is often recommended for preliminary screenings of aggregate reactivity. Aggregates are considered reactive in the AMBT if expansion surpasses the 0.1% threshold at 14 days [15]. However, the AMBT has been criticized for not accurately representing field conditions, potentially yielding false positives due to its harsh testing conditions, using mortar instead of concrete, and crushing process of the aggregates which can expose unstable minerals and exaggerate expansion results [4,10,16–18].

In contrast, the CPT, as per ASTM C1293, employs concrete prisms (i.e., 75x75x250 mm^3) and monitors expansion over a longer duration, typically 1 to 2 years for mixtures without

and with SCMs, under conditions of moderate temperature (i.e., 38°C) and high relative humidity (i.e., >95%) [13]. The reactivity threshold in CPT is generally set at 0.04% expansion at 1 year [15]. While the extended testing period allows for capturing long-term effects, the primary concern with CPT is the alkali's leaching, which reduces the alkali's content in the specimens over time, leading to lower ultimate expansion and possibly underestimating aggregate's reactivity [11,16,19]. Although measures such as covering specimens, increasing specimen size or accounting for leached alkalis are suggested to reduce the leaching effect, challenges remain due to issues such as shrinkage in sealed prisms and extensive leaching during submersion [5,11,19,20]. Moreover, CPT may still present issues for capturing the reactivity of slow-reacting aggregates, along with accounting for the influence of aggregate shape (i.e., flat particles from the crushing procedure demanding higher w/c lowering expansions), and the nature of non-reactive aggregates [4,15,17,21,22].

To reduce testing time and address the limitations of the traditional CPT, alternative methods have been proposed. For example, the Accelerated Concrete Prism Test (ACPT), a recommended method from RILEM (AAR-4.1) only for aggregate testing, increases the storage temperature to 60°C, reducing the test duration to 13 weeks [2,21]. However, the ACPT high-temperature setup can induce artificial reactions (i.e., affecting ASR-gel viscosity), cause excessive expansion that would not typically occur in field conditions and exacerbate leaching from the test prisms [4,19,21]. To address the alkali leaching, the Norwegian Concrete Prism Test (NCPT), standardized in Norway NB32 [23], that formed the basis for developing a slightly modified version of RILEM as AAR-10 [24] employs larger prisms (i.e., 100x100x450mm) to reduce the surface area-to-volume ratio while maintaining the same test conditions as CPT [25]. Additionally, the concrete cylinder test (CCT) and its variant CCT 60 are adaptations of CPT and ACPT, respectively, using concrete cylinders (i.e., diameter=200mm and height=100mm) rather than prisms [26].

In this sense, discrepancies between laboratory results and actual field performance have been reported [1,2,27,28]. Laboratory tests accelerate the reaction process and thus do not fully replicate the environmental conditions of field structures [10,19]. Therefore, several factors can affect the accuracy of the tests mentioned above. The concentration of alkalis and their movement within the concrete specimens play crucial roles. Alkalis can be recycled

during the test, and moisture movement can lead to changes in alkali concentration due to either ingress or leaching [19,29,30]. Higher moisture levels typically result in more leaching, while some alkalis may chemically bind to some cement paste hydrates. Additionally, the effect of temperature on alkali movement is also uncertain, adding complexity to the testing process. The size and preparation of specimens also significantly influence the test outcomes along with the test variations on pretreatment and storage conditions [8,11].

Several studies have highlighted the need for a stronger correlation between laboratory test results and field performance [31]. Moreover, it has been emphasized that understanding the limitations of current test setups and incorporating advanced analytical methods is crucial for enhancing predictive accuracy [31]. In this sense, the significant impact of factors such as alkali leaching or specimen size during accelerated testing, demands a more detailed correlation between the reactivity potential scales and field performance [10]. Furthermore, the variability in environmental exposure conditions, such as humidity and temperature, further complicates the relationship between laboratory test outcomes and field behaviour, requiring these variables to be accounted for better predictions [10,32].

This study aims to assess the reliability of traditional laboratory methods, specifically AMBT and CPT, in predicting AAR occurrence in the field. By examining a comprehensive database of laboratory and field results, this research seeks to understand the correlation of these methods with field behavior. Comparisons of the reactivity scales between laboratory and field contexts are developed to establish a clear correlation of time representativity. Finally, metrics such as accuracy, sensitivity, precision, and specificity are used to evaluate the performance of the test outcomes, distinguishing between mixtures with and without SCMs.

5.3 Scope of the work

To assess the reliability of traditional laboratory test methods, specifically AMBT and CPT referred to ASTM C1260 and ASTM C1293 in this work, respectively, in predicting the potential reactivity of aggregates and the efficiency of preventive measures, this study includes a few stages. First, a comprehensive database of laboratory results and corresponding field performance data is collected and organized to ensure its

representativeness and reliability. Then, a comparative analysis investigates the relationship between laboratory outcomes—focusing on the standard expansion thresholds of 0.1% for AMBT and 0.04% for CPT—and actual field performance, considering the threshold of 0.04% for labelling cases as reactive (i.e., YES) or non-reactive (i.e., NO). Also, an evaluation of the two methods through accuracy and sensitivity analysis to assess their reliability in predicting field behaviour is presented. Next, a K coefficient is introduced to further examine the correlation over time between accelerated laboratory outcomes and field performance, emphasizing the critical role of extended exposure periods in the field. Finally, the potential use of advanced analytical techniques, such as machine learning models, as tools to enhance the predictability of ASR development in the field is discussed.

5.4 Data preparation

5.4.1 Data collection

The experimental database is compiled from five projects involving laboratory results, conducted under different test setups, and 208 field-exposed concrete members, exposed to distinct environments [6–8,33,34]. As illustrated in Figure 5. 1, the database categorizes information regarding material selection (i.e., aggregate, cement, SCMs, and admixture) and mix proportions. Additionally, it includes details on the characteristics of laboratory and field samples (i.e., surface area, volume) along with their performance (i.e., expansion over time), laboratory test setup and environmental conditions.

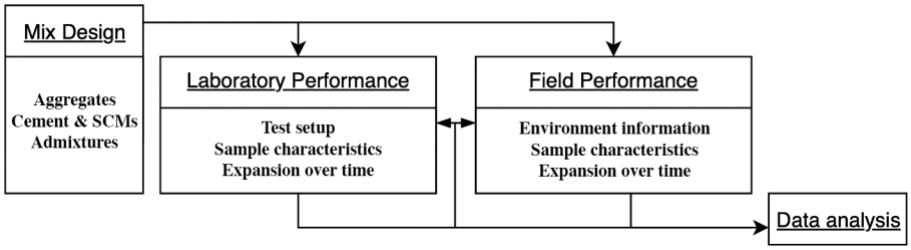


Figure 5. 1 – Database collection flowchart for assessing the reliability of laboratory tests to predict field performance.

5.4.2 Exploratory data analysis (EDA)

The exploratory data analysis (EDA) aims to differentiate patterns within the distribution across distinct testing setups and exposed environments. Figure 5.2 and Figure 5.3 present the variance observed for laboratory and field data, respectively. This step was conducted on the entire dataset, without filtering, to fully assess the available information.

The predominant testing setups (Figure 5.2a), comprise CPT(ASTM C1293) [13], and AMBT (ASTM C1260) [12], representing significant proportions of the data, accounting for 38.3%, and 36.5%, respectively. Additionally, the dataset includes results from the following testing setups: Accelerated concrete prism test (ACPT) [2], Norwegian concrete prism test (NCPT) [23], concrete cylinder test performed at 38^{o}C and 60^{o}C (CCT and CCT-60) [26], and a variation of the CPT at 50^{o}C (CPT-50) [26].

The laboratory samples display their surface areas and volumes that vary from 0.02375 to 0.20000 m^2 and 0.00014 to 0.00450 m^3 (Figure 5.2 b, Figure 5.2 c). Moreover, the analysis reveals variations from standard recommendations for some testing parameters (i.e., temperature, external humidity, and solution concentration), particularly in the concentration of sodium hydroxide solution (NaOH) in AMBT (Figure 5.2 d). These deviations range from 0.2 to 1 Mol/l, with a majority at 1 Mol/l, indicating potential influences on the test outcomes.

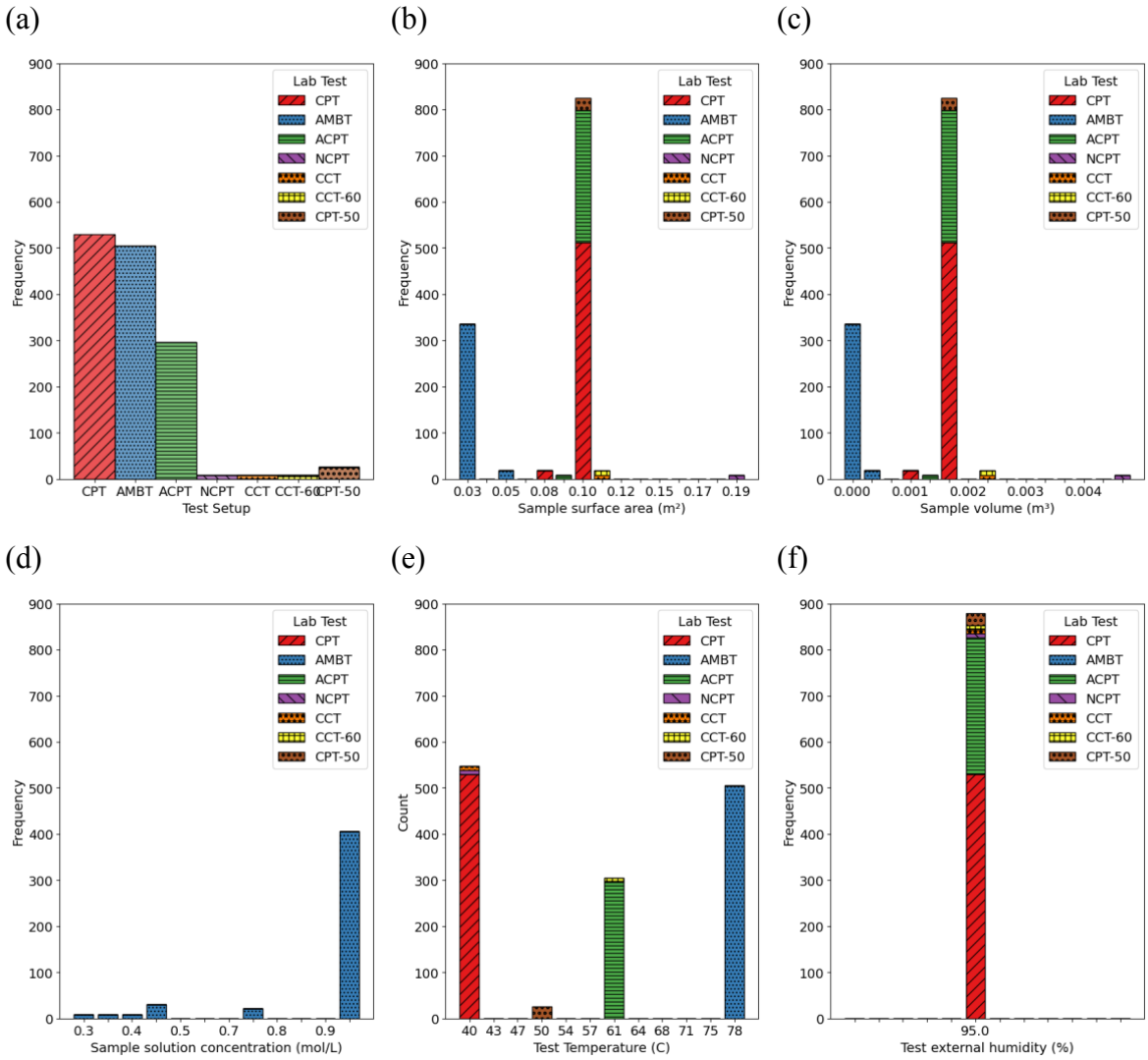


Figure 5. 2 – Database distribution of laboratory data by (a) testing, (b, c) sample characteristics and (d, e, f) their setup parameters.

The field data predominantly is observed in North American environments, with a significant contribution from Austin, US (66.6%), Ottawa, Canada (26.7%), and Kingston, Canada (2.6%). It includes the blocks geometry such as cubes and prisms (i.e., Long prisms and Flat prisms) (Figure 5.3c). Furthermore, the blocks vary in dimensions including cubes with 300x300x300mm^3, long prisms with 380x380x710mm^3, 400x400x700mm^3, 600x600x2000mm^3, and flat prims or slabs with 700x700x150mm^3 and 200x1200x4000mm^3. This is reflected in the variations of surface area (Figure 5.3a) and

volume (Figure 5.3b), indicating that most of the entries on the database represent surface area below 2 m^2 and volumes below 0.2 m^3.

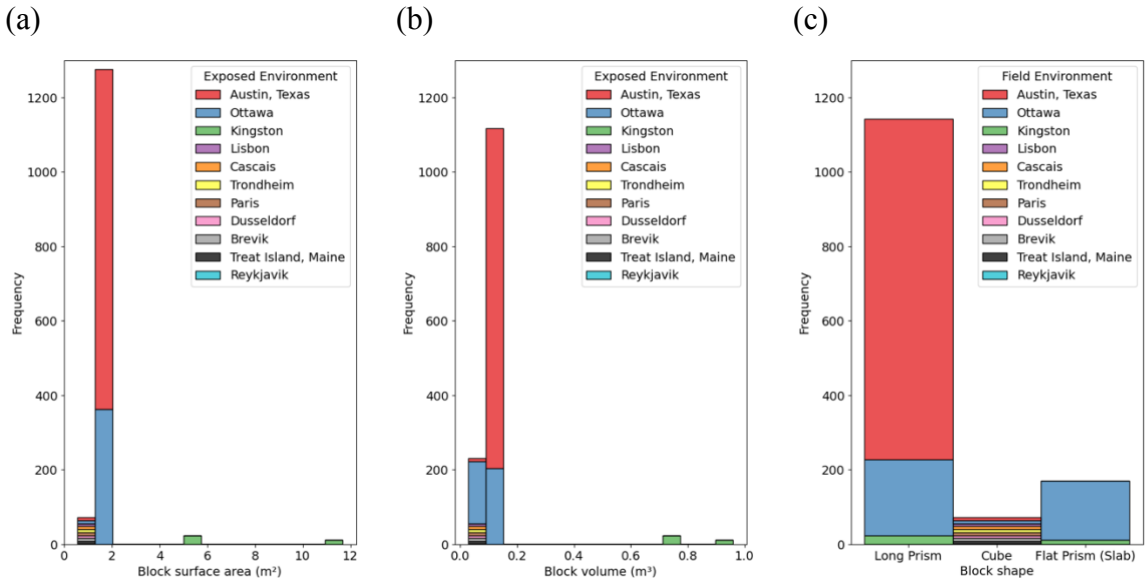


Figure 5. 3 – Database distribution of field data by exposed environment, in terms of sample (a) surface area, (b) volume, and (c) shape.

5.4.3 Data filtering

The filtering process of the database is crucial to refine the scope of the study. As the goal was to determine the reliability of current testing setups, the following criteria were used to ensure the relevance of the analysis:

1. **Laboratory and Field Correlation**: only field blocks with at least one corresponding laboratory performance result were included. This criterion recognizes that multiple laboratory results may correlate to a single field-exposed structure, ensuring a comprehensive understanding of the laboratory-field relationship.
2. **Focus on AMBT and CPT**: the analysis was exclusively focused on results from the Accelerated Mortar Bar Test (AMBT) and the Concrete Prism Test (CPT). These tests were selected due to their prevalent use. Consequently, data from other testing methods were excluded from this analysis to maintain consistency and relevance.

3. **Standard Compliance**: only test outcomes following current standards were selected. For the AMBT, this included specimens submerged in a NaOH 1 Mol/l solution at 80°C [12]. For the CPT, this included specimens stored at 38°C and over 95% RH [13]. Ensuring compliance with these standards ensures the reliability and comparability of the test results.
4. **Structural Criteria**: only concrete blocks without reinforcement and displaying geometry of long prisms (i.e., 380x380x710mm^3, 400x400x700mm^3, 600x600x2000mm^3) and cubes (i.e., 300x300x300mm^3) were selected to eliminate the potential influence of variations in these parameters on the test outcomes, thereby providing a clearer understanding of the material's inherent reactivity.
5. **Admixture Inclusion**: results without preventative admixtures, besides air entraining and superplasticizer, were included. This approach allows for the elimination of the effect of different admixtures on the results.
6. **Grouping by binders**: two main groups were studied: samples without SCMs and with SCMs. This distinction is essential for understanding the efficiency of the tests to recognize 2 distinct things: a) the reactivity of aggregate and b) the effect of preventative measures.
7. **Alkali Content Analysis**: to evaluate the probability of occurrence based on the cement alkali content, including any additional boosting in field mixtures (i.e., usually up to 1.25% of the cement), three main groups were selected for analysis based on data availability: low alkali content (less than 2.94kg/m^3), moderate alkali content (2.94kg/m^3 to 4.2 kg/m^3), and high alkali content (above 4.2 kg/m^3). This classification helps in understanding the influence of alkali content on the test reliability outcomes.

The filtering process formed the basis for the analysis presented in the subsequent sections. By employing these detailed filtering and grouping criteria, this study ensures a comprehensive and robust analysis of the data, providing valuable insights into the reliability of AMBT and CPT testing setups in predicting field performance for structures with and without SCMs.

5.5 Comparative analysis and performance evaluation

5.5.1 Comparative Analysis

Based on the filtered dataset, Figure 5. 4 illustrates the relationship between the outcomes from AMBT, CPT and field expansion at their respective intervals for mixes without SCMs. An initial visual analysis of the scatter plots (Figure 5. 4a, Figure 5. 4b) indicates a notable spread within the dates, suggesting some relationship between AMBT and CPT outcomes (Figure 5. 4c). It is worth noting that the thresholds used, 0.1% for AMBT at 14 days and 0.04% for CPT at one year, align with current standards [15]. For this comparison, the last recorded expansion for the field-exposed blocks was used, with an average field exposure of 8.3 years for the AMBT, ranging from 1.3 to 29 years, and an average of 7.6 years for the CPT, with a range of 1.1 to 27 years, as seen in Appendix A.

Next, a correlation analysis was conducted using the Pearson correlation coefficient (r), aiming to evaluate the strength and direction of the linear relationship between laboratory tests and field performance outcomes [35]. Given the values range from -1 to 1, indicating an inverse relationship for negative values and a direct relationship for positive values, the interpretation could indicate values equal to 0 with no linear correlation, from 0 to 0.4 as very weak and potentially negligible, 0.4 to 0.7 as weak and likely insignificant, 0.7 to 0.8 as moderate correlation and potentially meaningful, 0.8 to 0.9 strong correlation and 0.9 to 1 very strong correlation. The comparison shows that CPT has a weak positive relationship with field results, with a correlation coefficient of 0.66, while AMBT's correlation with field data is very weak and at 0.33. The correlation between AMBT and CPT is weak, at 0.49. However, relying solely on this correlation may not fully capture the complexity of reactivity outcomes, as it only measures the linear relationships and does not account for influencing parameters such as environmental factors. Therefore, analyzing additional metrics like accuracy and sensitivity is essential to gain a deeper understanding of each test's performance, as shown in Section 5.5.2.

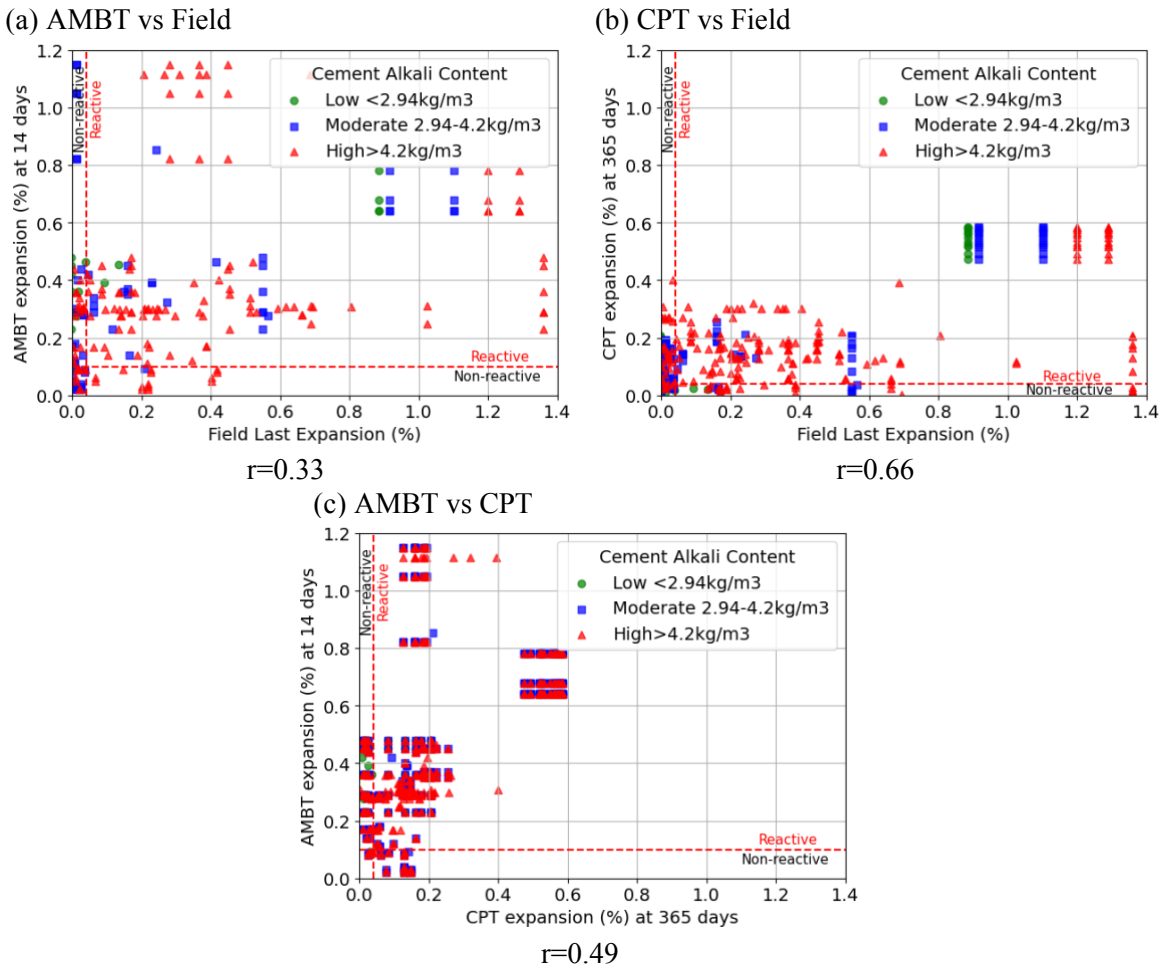


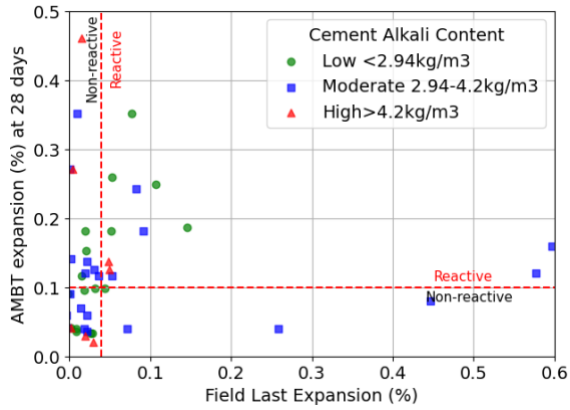
Figure 5. 4 – Correlation between (a) AMBT at 14 days and Field last expansion (%), (b) CPT at 365 days and Field last expansion (%), and (c) AMBT at 14 days and CPT at 365 days expansions (%) at their respective threshold for mixes without SCMs.

Following the analysis of mixes without SCMs, the filtered dataset for mixes containing SCMs is examined and presented in Figure 5. 5. This figure illustrates the relationship between the outcomes from AMBT, CPT, and field expansion at their respective intervals. The scatter plots (Figure 5. 5a, Figure 5. 5b) reveal a significant dispersion in the data, indicating some correlation between the AMBT and CPT results (Figure 5. 5c). The thresholds applied—0.1% for AMBT at 28 days and 0.04% for CPT at two years—are consistent with current standards. Additionally, blocks exposed to field conditions exceeding 0.04% expansion at the last recorded age are considered susceptible to ASR [11,12]. For

mixes containing SCMs, the evaluation for both AMBT and CPT was conducted at an average field exposure of 13.3 years, with a range from 1.7 to 27 years (Appendix A).

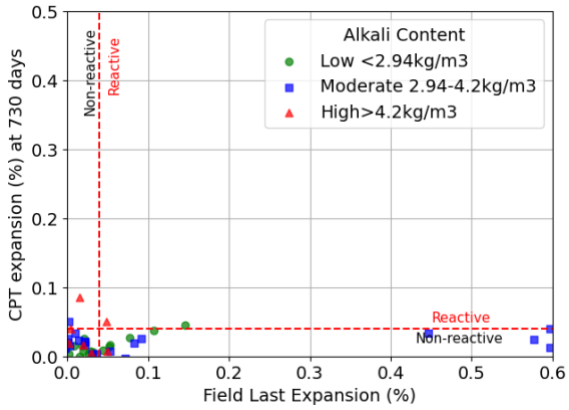
The comparative analysis between CPT results and field data shows a very week and potentially negligible correlation, with a coefficient of 0.18. The correlation between AMBT and field data is even weaker nearly to non-existent, with a coefficient of 0.04. Finally, the relationship between AMBT and CPT results is weak and positive, with a correlation coefficient of 0.62. Additionally, the p-values for AMBT vs. field (0.7958) and CPT vs. field (0.2284) are both greater than 0.05, suggesting that these correlations are not statistically significant at the 5% significance level. The lack of statistical significance implies that, despite differences in the correlation coefficients, other factors such as environmental conditions or the type of SCMs used, as well as potential nonlinear relationships, may influence the reactivity outcomes. Therefore, analyzing additional metrics like accuracy and sensitivity is essential to gain a deeper understanding of each test's performance, as shown in Section 5.5.2.

(a) AMBT vs Field



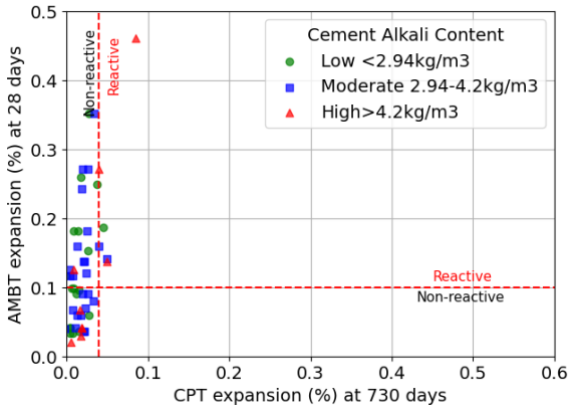
r=0.04

(b) CPT vs Field



r=0.18

(c) AMBT vs CPT



r=0.62

Figure 5. 5 – Correlation between (a) AMBT at 14 days and Field last expansion (%), (b) CPT at 365 days and Field last expansion (%), and (c) AMBT at 14 days and CPT at 365 days expansions (%) at their respective threshold for mixes with SCMs.

5.5.2 Performance Evaluation

Accuracy is based on the binary indicator, either reactive (i.e., YES) or non-reactive (i.e., NO). Concordance in test and field outcomes signifies accuracy, while divergence suggests failure (i.e., accuracy = correct predictions/total number of predictions). In this sense, for mixes without SCMs, this binary classification yields an accuracy rate of 68% for CPT at the 365-day benchmark in predicting field performance, whereas AMBT similarly stands at 72% at the 14-day mark, when compared to the last recorded expansion, meaning averages of 8.3 and 7.6 for AMBT and CPT, respectively (Appendix A).

A detailed exploration of the test's performance is conducted through a sensitivity analysis via a confusion matrix. This analysis quantifies the proportions of true positives (TP), true negatives (TN), and their false counterparts: false positives (FP) and false negatives (FN). Therefore, an integrative analysis generates new metrics that enhance the comprehension of the test outcomes. It evaluates the sensitivity (i.e., TP/(TP+FN)), specificity (i.e., TN/(TN+FP)), precision (i.e., positive predictive value or TP/(TP+FP)), and the F-1 score (i.e., the harmonic mean of precision and specificity).

Figure 5. 6 summarizes the counts for the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), offering a comprehensive view of both tests' predictive capabilities in assessing field performance outcomes.

The AMBT without SCMs achieved an accuracy of 72%, indicating that it correctly classified 72% of the cases overall. The precision shows that 74% of the positive predictions made by the AMBT were correct, meaning it is quite reliable to predict reactive cases. Additionally, AMBT exhibited a strong sensitivity, correctly identifying 90% of the actual reactive cases. However, its specificity is significantly lower, indicating that it only correctly identified 32% of non-reactive cases, leading to a higher number of false positives. The F1 score for AMBT is 0.82, demonstrating a strong balance between precision and recall, making it an effective test for identifying reactivity for mixes without SCMs.

For CPT without SCMs, an accuracy of 68% was observed, meaning it correctly classified 68% of the cases. Its precision, as for the AMBT, shows that 74% of the predicted reactive cases were correctly classified. The sensitivity of CPT was slightly lower than AMBT, indicating that it identified 85% of the true positive cases, still a strong performance. The specificity of CPT is low, meaning it correctly classified only 25% of non-reactive cases, leading to more false positives. The F1 score for CPT is 0.79, reflecting a solid balance between precision and recall, though slightly lower than AMBT.

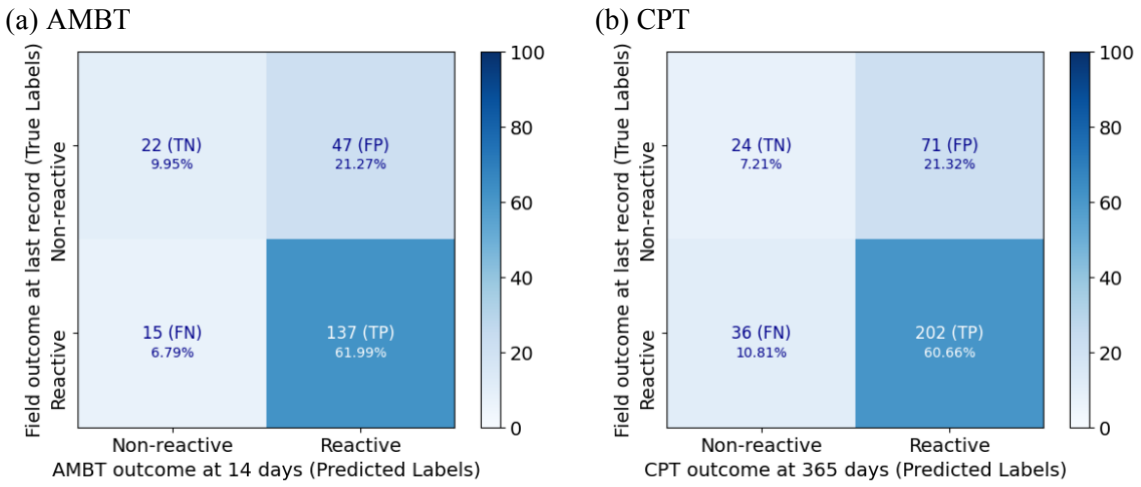


Figure 5. 6 – Confusion matrix representing true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for (a) AMBT outcomes and (b) CPT outcomes for mixes without SCMs, in percentage (%).

Recognizing that a linear relationship alone may not fully capture the complexity of reactivity outcomes, a detailed analysis of accuracy and sensitivity was also conducted for mixes with SCMs. For mixes containing SCMs, the evaluation for both AMBT and CPT was conducted at an average field exposure of 13.3 years, with a range from 1.7 to 27 years (Appendix A). Figure 5. 7 provides a detailed summary of the counts for true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), offering a comprehensive overview of both tests' predictive capabilities in evaluating field performance outcomes. In this context, AMBT and CPT for mixes with SCMs achieved the same accuracy of 76% in predicting field performance.

For CPT with SCMs, the precision indicated that 50% of the positive predictions were correct, meaning that half of the cases identified as reactive were truly reactive. However, the low sensitivity of CPT (i.e., 19%) suggests the test frequently misses reactive cases (i.e., high false negatives). On the other hand, its specificity demonstrates that it correctly identified 91% of the non-reactive cases, meaning it performs well at classifying negative cases but struggles with positive detection. The F1 score for CPT, which balances precision and recall, is 0.27, reflecting its overall challenge in identifying true positive cases effectively.

In contrast, AMBT with SCMs demonstrated stronger sensitivity, correctly identifying 75% of the true positives, significantly outperforming CPT in detecting reactive cases. Its precision, however, remains the same as CPT at 0.50, meaning that while AMBT is good at identifying reactive cases, only half of its positive predictions are accurate. AMBT's specificity of 0.64 is lower than CPT's, indicating that it correctly identified 64% of the non-reactive cases, making it more prone to false positives. The F1 score for AMBT is 0.60, suggesting a better balance between precision and recall compared to CPT.

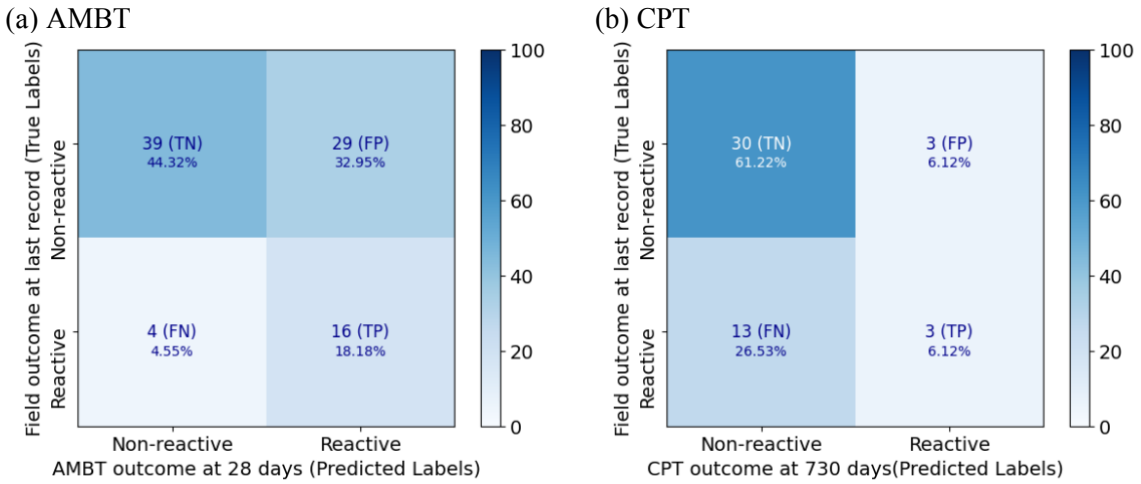


Figure 5. 7 – Confusion matrix representing true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for (a) AMBT outcomes and (b) CPT outcomes for mixes with SCMs, in percentage (%).

5.6 Accelerated laboratory vs field equivalency analysis over time

The coefficient K is introduced to better explore the correlation over time between accelerated laboratory outcomes and field performance. This coefficient is defined as the ratio of long-term field performance expansion (R_{lt}) to the accelerated laboratory test expansion (R_{at}) at a given time. The K coefficient allows for a comparative analysis to understand how well the accelerated tests predict long-term behavior in actual field conditions.

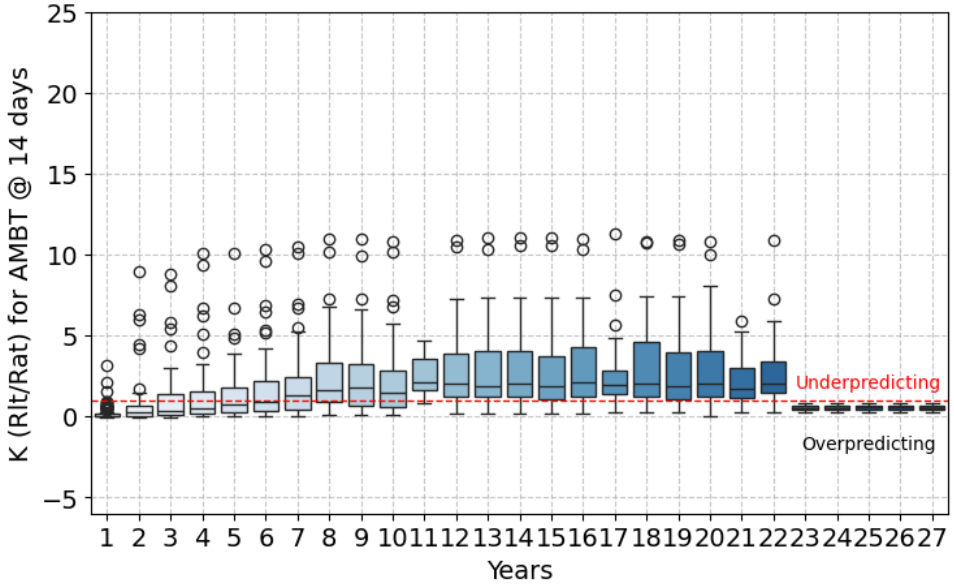
$$K = \frac{R_{lt}}{R_{at}}$$

The primary purpose of the coefficient K is to establish a relationship between the actual field performance and the accelerated test results. By calculating K, a first indication of the validity of accelerated tests as predictors of long-term behavior can be concluded.

- K>1: Indicates that the field expansion is greater than the laboratory expansion. This could suggest that the laboratory test underpredicts the actual field performance.
- K<1: Indicates that the field expansion is less than the laboratory expansion. This could suggest that the laboratory test overpredicts the actual field performance.
- K=1: Indicates a perfect correlation where the accelerated laboratory expansion matches the field performance expansion.

Figure 5. 8a and Figure 5. 8b show the K coefficient for AMBT at 14 days and CPT at 365 days over 27 years of exposed concrete blocks without SCMs, respectively. This analysis focuses on the overall averages of all parameters, without considering specific factors like environmental conditions and alkali loading, which are discussed in detail in Section 5.7. In Figure 5. 8a, the average K values are generally below one for the first six years, indicating that AMBT overpredicts the field performance during this period. After six years, the K values surpass one, suggesting that laboratory results underpredict field performance. Figure 5. 8b shows that the K values for CPT start below one but increase over time, indicating that up to two years of block exposure, CPT overpredicts the field performance but becomes underpredicting it as time progresses. Additionally, the variability in CPT results remains higher compared to AMBT throughout the evaluated data. This divergence arises from variations of parameters affecting the reaction kinetics such as exposure environment, alkali contents, and aggregate type. Moreover, a lower variability is evident for the periods between 23 years to 27 years, representing two distinct mixtures within a single project. In this case, the blocks are exposed to the same environment conditions (i.e., Kingston, Canada) and only one source of aggregate (i.e., Spratt) is being tested, highlighting the more controlled nature of this dataset. Yet, this demonstrates they are mostly in the overprediction zone for AMBT and underprediction zone for CPT.

(a) AMBT



(b) CPT

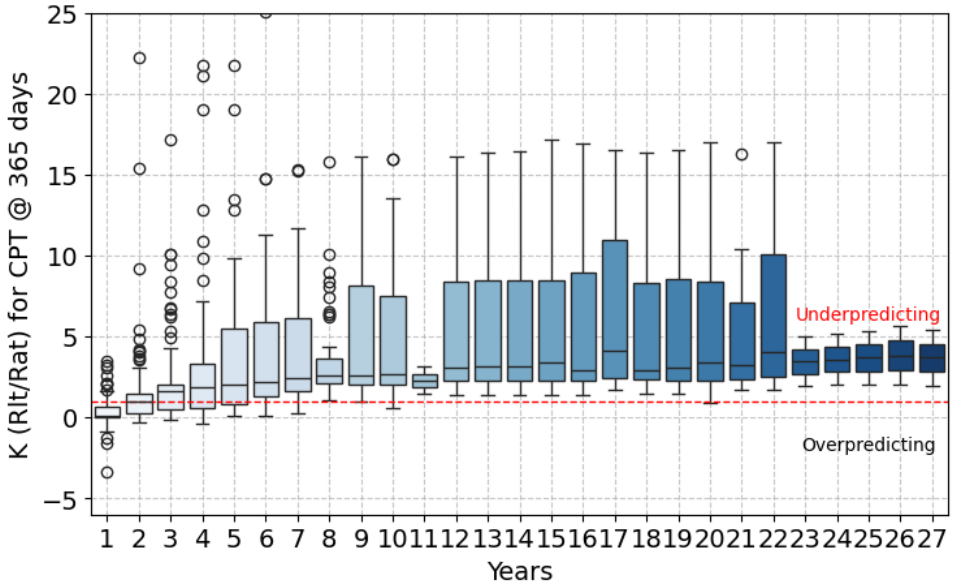
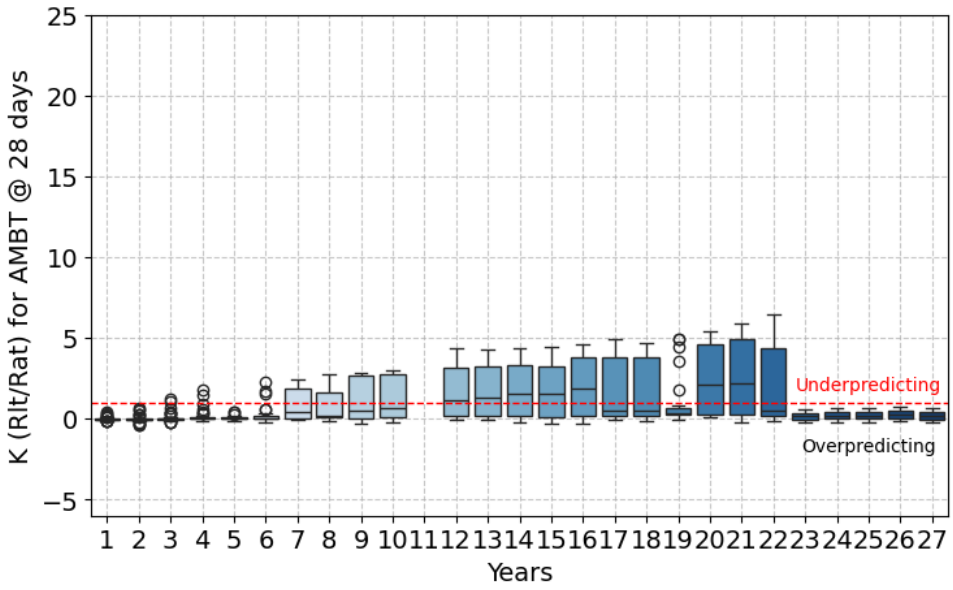


Figure 5. 8 – K coefficient for (a) AMBT and (b) CPT over 27 years of exposed concrete blocks without SCMs.

Figure 5.9a and **Figure 5.9b** illustrate the K coefficient for AMBT at 28 days and CPT at 730 days over 27 years of exposed concrete blocks with SCMs. In **Figure 5.9a,** the K values for AMBT vary along the k=1 mark throughout the 27 years. **Figure 5.9b**, however, shows significant variability in K values for CPT, which tend to increase over time. This indicates an initial overprediction of field performance by CPT, which transitions to underpredicting

field performance after approximately 6 years. Additionally, lower variability is noted between 23 and 27 years, representing a sole project exposed to the same environmental conditions (i.e., Kingston, Canada), using Spratt aggregate, and showing AMBT overpredicting and CPT fluctuating between overprediction and underprediction.

(a) AMBT



(b) CPT

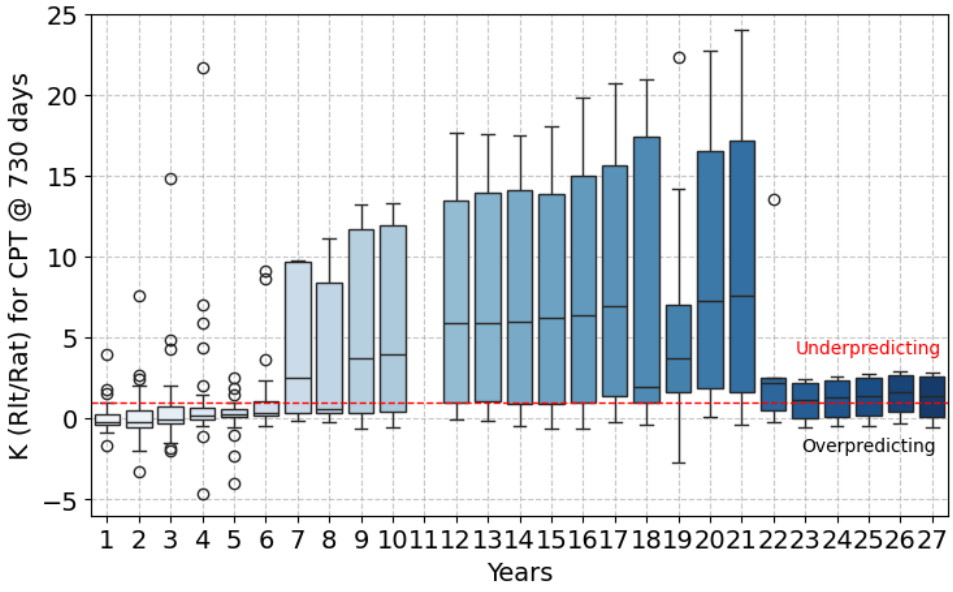


Figure 5. 9 – K coefficient for (a) AMBT and (b) CPT over 27 years of exposed concrete blocks with SCMs.

5.7 Discussion

This study's results reveal significant insights into the performance of accelerated tests (i.e., AMBT and CPT) in predicting AAR in field-exposed concrete blocks, both with and without SCMs. The findings highlight the relative strengths and limitations of these tests, suggesting the need for improved predictive accuracy through multifactorial analysis and integration of laboratory and field data.

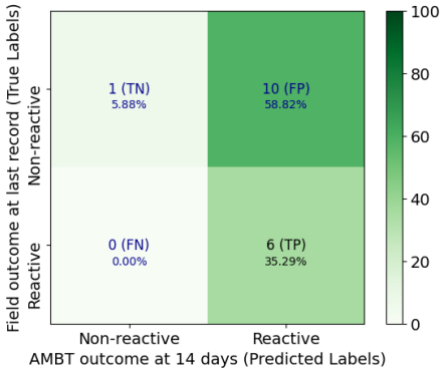
Figure 5. 4, Figure 5. 5 and subsequent linear regression analyses reveal moderate correlations between CPT results and field performance (i.e., 0.66 for mixes without SCMs and 0.18 for mixes with SCMs), but weaker correlations for AMBT (i.e., 0.33 for mixes without SCMs and 0.04 for mixes with SCMs). Specifically, the sole linear correlation coefficients between CPT and field data could suggest that CPT better predict long-term field behavior. This aligns with prior studies indicating CPT's extended duration and controlled conditions better mimic actual field conditions, despite not accounting for all environmental variables [2,10,36]. Conversely, AMBT's weaker correlation with field data highlights its tendency to overpredict reactivity due to its aggressive testing environment. This is consistent with literature suggesting AMBT's high-alkali immersion conditions accelerate reactions beyond typical field scenarios, thus skewing results towards higher reactivity predictions [1,2]. These discrepancies underscore the limitations of accelerated tests in capturing the nuanced interactions present in real-world conditions.

However, the evaluation of test accuracy and sensitivity (Figure 5. 6 and Figure 5. 7) using confusion matrices reveals that CPT and AMBT have similar accuracy for mixes without SCMs (i.e., 68% vs 72%, respectively) and the same accuracy for mixes with SCMs (i.e., 76%). In this sense, the sensitivity and specificity metrics indicate that AMBT is generally more reliable in identifying true positive cases of reactivity for mixes without SCMs. This supports the notion that although, in the overall context, CPT may better simulate field conditions [37], the more intense testing conditions of AMBT result in fewer false negatives, making it more reliable for identifying reactive cases.

When filtering and comparing the results of laboratory tests that have been boosted in the laboratory (i.e., 5.25 kg/m3 Na_2O_{eq} for CPT as per ASTM C1293) against varying field alkali levels (i.e., low, moderate, and high), the confusion matrix (Figure 5.10) indicates that CPT

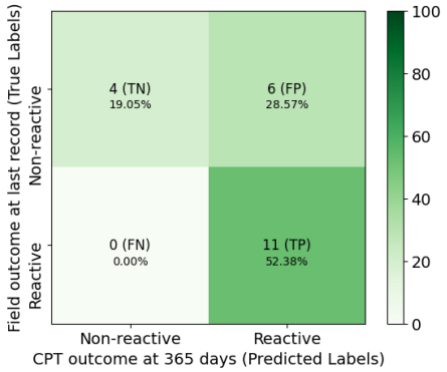
results better reflect the behavior of conventional concrete (i.e., 2.6 kg/m³ Na_2O_{eq}) or low alkali field conditions compared to AMBT. However, the lower CPT responses at moderate and high alkali field levels can be associated with alkali leaching issues related to the test. This phenomenon may compromise its ability to fully capture the influence of elevated alkali concentrations in field conditions, affecting its performance under such scenarios. Nevertheless, CPT's higher accuracy (71%), precision (65%), and specificity (40%) in low alkali conditions are attributed to its performance in better identifying true negatives (TN) and true positives (TP), alongside fewer false negatives (FN), when compared to AMBT.

(a) AMBT – Low alkali
(<2.94 kg/m^3)



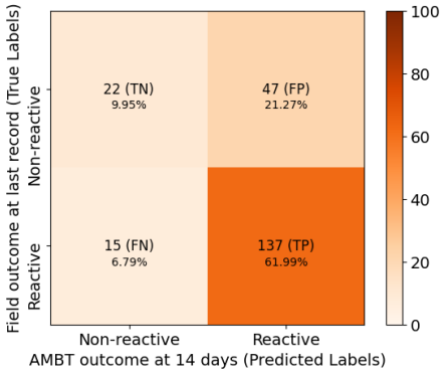
Accuracy – 41% | Precision – 38%
Sensitivity – 100% | Specificity – 9%

(b) CPT – Low alkali
(<2.94 kg/m^3)



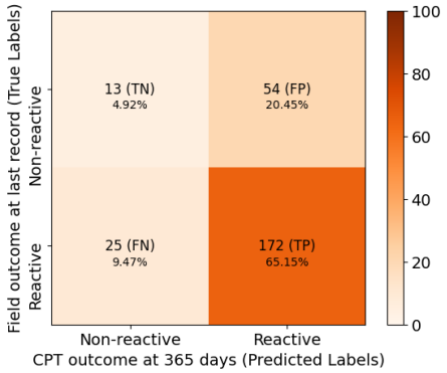
Accuracy – 71% | Precision – 65%
Sensitivity – 100% | Specificity – 40%

(c) AMBT – Moderate alkali
(>2.94 kg/m^3;<4.2 kg/m^3)



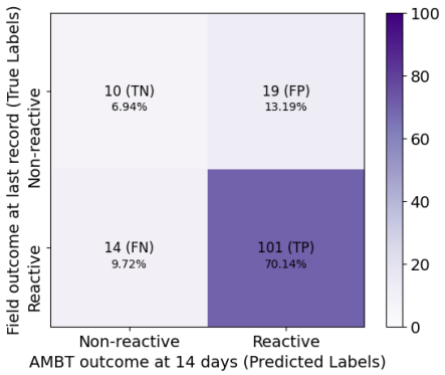
Accuracy – 41% | Precision – 74%
Sensitivity – 90% | Specificity – 32%

(d) CPT – Moderate alkali
(>2.94 kg/m^3;<4.2 kg/m^3)



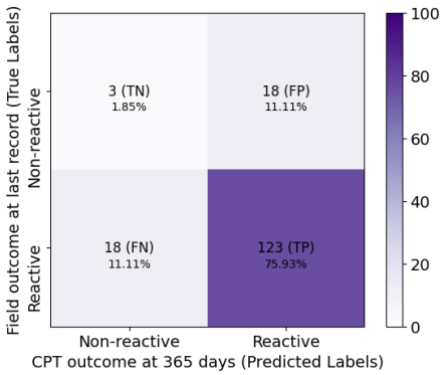
Accuracy – 41% | Precision – 38%
Sensitivity – 100% | Specificity – 9%

(e) AMBT – High alkali (>4.2 kg/m^3)



Accuracy – 77% | Precision – 84%
Sensitivity – 88% | Specificity – 34%

(f) CPT – High alkali (>4.2 kg/m^3)



Accuracy – 78% | Precision – 87%
Sensitivity – 87% | Specificity – 14%

Figure 5. 10 – K coefficient for (a) AMBT and (b) CPT over 27 years of exposed concrete blocks with SCMs.

For mixes with SCMs, CPT has low sensitivity (0.19), indicating it struggles to identify true reactive cases, but its high specificity (0.91) makes it effective at detecting non-reactive cases. In contrast, AMBT shows much higher sensitivity (0.75), making it more reliable for identifying reactive cases, though its precision (0.50) remains similar to CPT, meaning both tests have a comparable rate of false positives. This lower precision for both tests may be attributed to the inability to fully account for the mitigating effects of SCMs under accelerated conditions, which may not accurately represent the slower, more complex interactions in field environments. Additionally, SCMs, known for their capacity to mitigate ASR through mechanisms such as pozzolanic reactions and reduced permeability, may not fully exhibit these benefits under AMBT, once it is not a performance test [19,32].

Yet, when implementing the time variable through the coefficient K analysis over 27 years (Figure 5. 8 and Figure 5. 9), valuable insights into the long-term predictive capability of AMBT and CPT are clarified. This underscores the importance of validating accelerated laboratory results against field performance over extended exposure periods. Generally, for mixes without SCMs, the K values for both tests initially overpredict field performance ($K < 1$) but transition to underpredicting ($K > 1$) over time. This trend highlights the temporal limitations of accelerated tests, which may not fully capture the delayed nature of AAR expansion observed in the field.

For mixes with SCMs, AMBT's K values remain closer to one indicating a more consistent estimation of field performance. CPT, however, shows initial overprediction, transitioning to underprediction after about six years. This suggests that while SCMs enhance long-term durability, their impact on mitigating ASR is not fully captured, necessitating more nuanced data analysis approaches that account for the effects of SCMs and environmental exposure [4,10]. Also, the alkali content in mixes tends to interfere with test outcomes, reflecting a complex interaction of such variables in predicting field performance through accelerated tests [19].

5.7.1 Environmental influence

Figure 5. 11 provides an analysis of the K coefficient under different environmental conditions over 22 years of exposed concrete blocks compared with both AMBT and CPT tests, distinguishing between warm (i.e., Austin, Texas) and cold (i.e., Ottawa, ON) environments for mixes without SCMs. The K values for AMBT in warm environments fall mostly in the underprediction zone over time. This indicates that the accelerated test conditions may not fully account for the increased reactivity that occurs in warmer field conditions over the long term. Yet, in cold environments, the K values for AMBT remain relatively below one, indicating a consistent overprediction of field performance. This suggests that AMBT may be more reliable in colder conditions, although some variability still exists [10]. Consequently, the stability of K values in cold environments implies that the effects of temperature on reactivity are less pronounced, allowing AMBT to provide a more accurate prediction over the long term.

For CPT, the K values in warm environments show a clear transition from underprediction to overprediction in the early ages (i.e., 1-2 years). The variability increases significantly over time, reflecting the complex interaction between prolonged exposure and warm conditions, which the accelerated test struggle to replicate accurately [32]. This trend indicates that CPT, while initially overpredicting, may begin to underestimate reactivity due to the sustained high temperatures. In cold environments, the K values for CPT remain below one for a longer period, up to 9 years, before transitioning to a distribution that falls in the

underprediction zone. The lower variability compared to warm environments suggests that CPT may provide more consistent results in colder climates, although the transition indicates potential long-term predictive challenges [28].

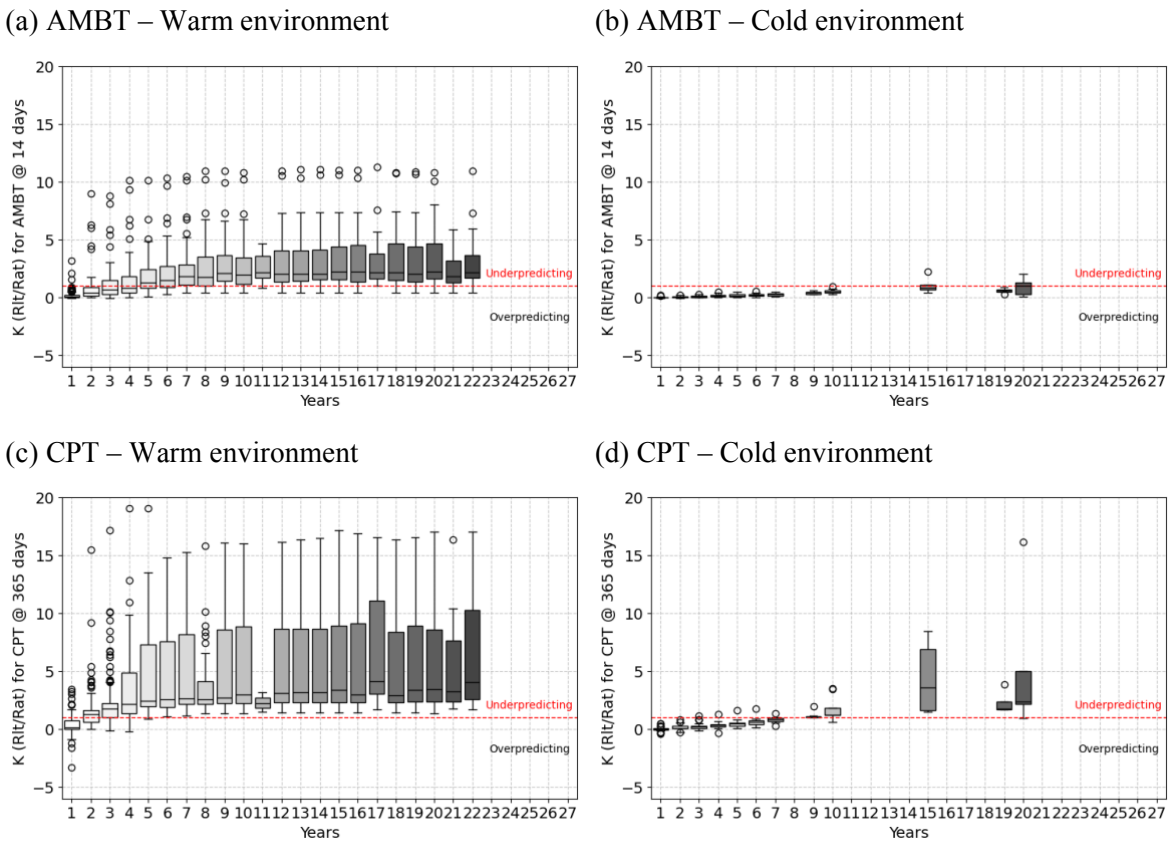


Figure 5. 11 – K coefficient for (a, b) AMBT and (c, d) CPT over 22 years of exposed concrete blocks for mixes without SCMs for (a, c) Warm and (b, d) Cold environments.

For mixes with SCMs, Figure 5. 12, provides an analysis of the K coefficient under different environmental conditions over 22 years of exposed concrete blocks comparing with both AMBT and CPT tests, distinguishing between warm (i.e., Austin, Texas) and cold (i.e., Ottawa, ON) environments. The K values for AMBT in warm environments show low variation but transition to the underprediction zone after around three years. This indicates that while SCMs help stabilize the reactivity predictions initially, the accelerated test conditions in warmer environments eventually lead to underprediction [32]. In cold environments, the K values for AMBT with SCMs show minimal variation and stay mostly

in the overpredicting zone, further underscoring the stabilizing effect of SCMs in different environmental conditions.

For CPT with SCMs, the K values in warm environments also show much higher variability compared to cold environments, and an earlier transition from underprediction to overprediction zones (i.e., 2 years). In cold environments, the K values for CPT with SCMs remain relatively stable and below one up to 19 years, indicating that SCMs contribute to more consistent and reliable long-term predictions in colder climates.

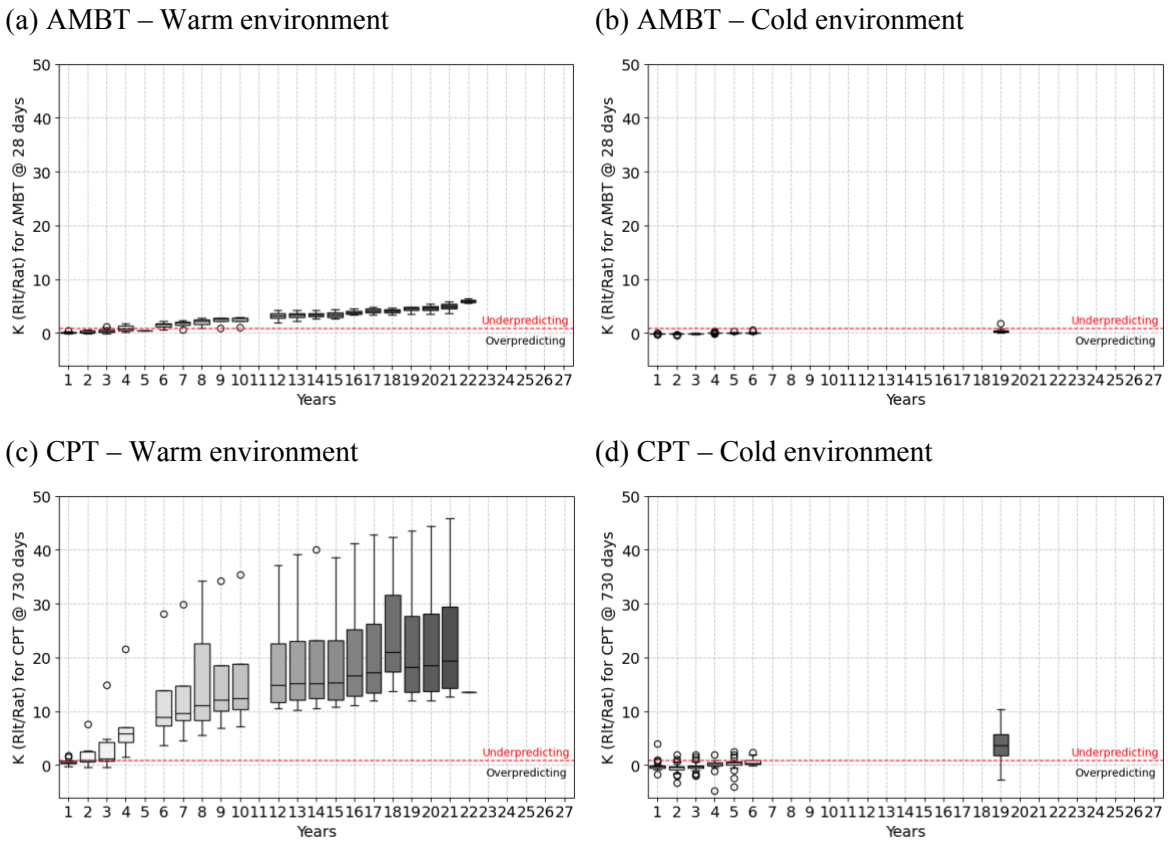


Figure 5. 12 – K coefficient for (a, b) AMBT and (c, d) CPT over 22 years of exposed concrete blocks for mixes without SCMs for (a, c) Warm and (b, d) Cold environments.

The analysis of K values under different environmental conditions underscores the necessity of considering environmental factors in the predictive modelling of ASR. Both AMBT and CPT exhibit results influenced by temperature, which accelerated tests may not fully replicate [2].

5.7.2 Alkali loading influence

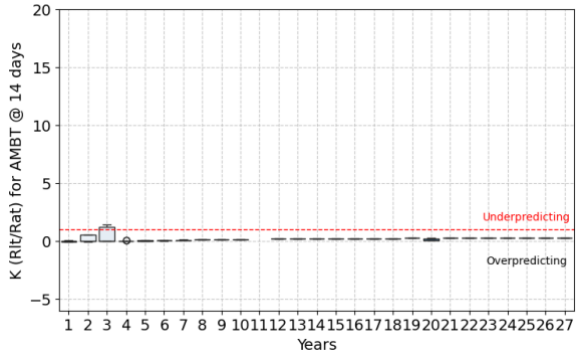
Figure 5. 13 illustrates the K coefficient for AMBT at 14 days and CPT at 365 days over 27 years of exposed concrete blocks for mixes without SCMs categorized by low (i.e., 2.94kg/m^3), moderate (i.e., between 2.94kg/m^3 and 4.20 kg/m^3), and high alkali content (i.e., above 4.20 kg/m^3). Notably, the influence of alkali content on the K coefficients for both AMBT and CPT is significant.

For low and moderate alkali content mixtures, AMBT (Figure 5. 13a and Figure 5. 13c) shows a consistent trend of overprediction over time, which aligns with the literature indicating that accelerated tests often exaggerate the reactivity potential due to higher alkali availability during the test [2]. In contrast, CPT (Figure 5. 13b and Figure 5. 13d) initially overpredicts field performance but begins to underpredict after about six and three years for low and moderate alkalis, respectively. This shift suggests that prolonged exposure periods reveal reactivity that was not captured by the CPT.

For high alkali content, the trend is similar for both AMBT and CPT. Although initially, the tests are overpredicting (K<1) field performance, they transition to underpredicting the field performance at 6 years for AMBT and 2 years for CPT. The earlier transition for CPT compared to AMBT indicates that AMBT might be better suited for capturing the initial reactivity responses but also struggles with long-term prediction [36].

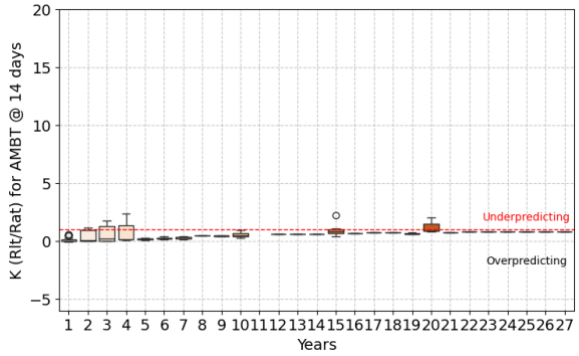
The variability in K values for high alkali content mixes highlights the complex interaction between alkali content and environmental conditions, which are challenging to replicate accurately in accelerated tests. This variability reflects that higher alkali content generally leads to more pronounced reactivity, which accelerated tests might not fully capture over time [4]. The observed trends suggest that in all alkali levels, CPT initially overpredict field performance while AMBT remains in the overpredicting zone for low and moderate alkalis. However, as the exposure period extends, both tests transition to underpredicting field performance, reflecting the limitations of accelerated tests in simulating long-term behaviour.

(a) AMBT – Low alkali
(<2.94 kg/m^3)

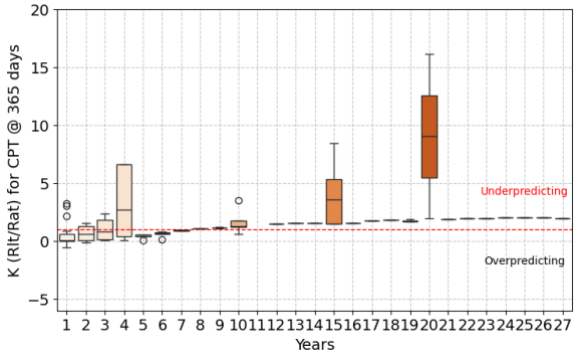


(b) CPT – Low alkali
(<2.94 kg/m^3)

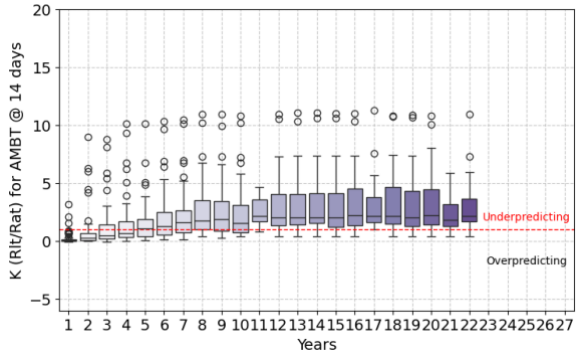
(c) AMBT – Moderate alkali
(>2.94 kg/m^3;<4.2 kg/m^3)



(d) CPT – Medium alkali
(>2.94 kg/m^3;<4.2 kg/m^3)



(e) AMBT – High alkali
(>4.2 kg/m^3)



(f) CPT – High alkali
(>4.2 kg/m^3)

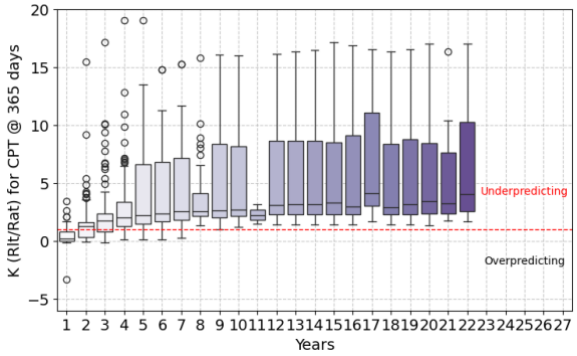


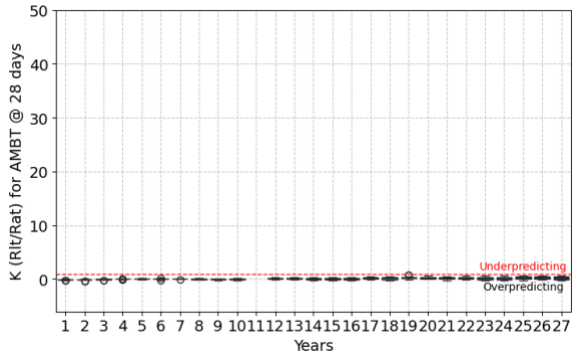
Figure 5. 13 – K coefficient for (a, c, e) AMBT and (b, d, f) CPT over 27 years of exposed concrete blocks for mixes without SCMs for (a, b) low, (c, d) medium, (e, f) high alkali content.

For mixes with SCMs, as exposed in Figure 5. 14a to Figure 5. 14f, the K coefficient for AMBT at 28 days and CPT at 730 days over 27 years of exposed concrete blocks is

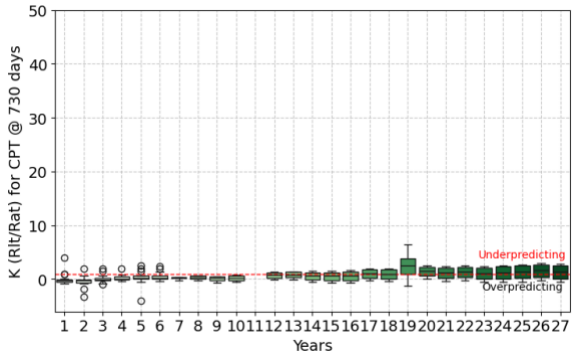
categorized by alkali content: low (i.e., 2.94kg/m^3), moderate (i.e., between 2.94kg/m^3 and 4.20 kg/m^3), and high alkali content (i.e., above 4.20 kg/m^3).

In low alkali content mixes with SCMs, both AMBT and CPT tend to remain in the overpredicting zone, yet the variations of CPT surpass the line especially around 16 years of predictions. In moderate alkali content mixes with SCMs, AMBT and CPT initially overpredict and then transition to underprediction at around 6 and 5 years, respectively. However, a much higher variability in the test outcomes is observed for CPT. Finally, a small amount of data is observed for high alkali content mixes with SCMs, yet, AMBT remains in the overprediction zone, while CPT transitions to the underprediction zone at 19 years. This suggests that while SCMs can be effective in controlling reactivity under certain conditions (i.e., type and amount employed, aggregate reactivity and environment), high alkali content mixtures are linked to reduced performance in long-term predictions from accelerated laboratory tests [10].

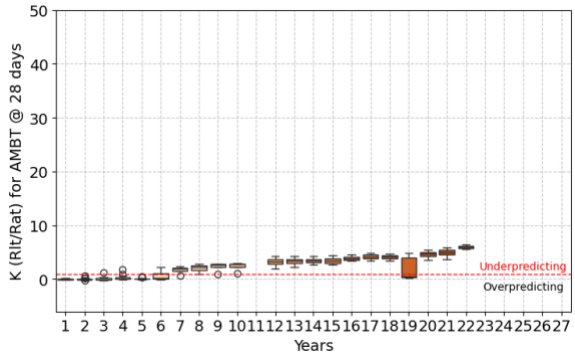
(a) AMBT – Low alkali
(<2.94 kg/m^3)



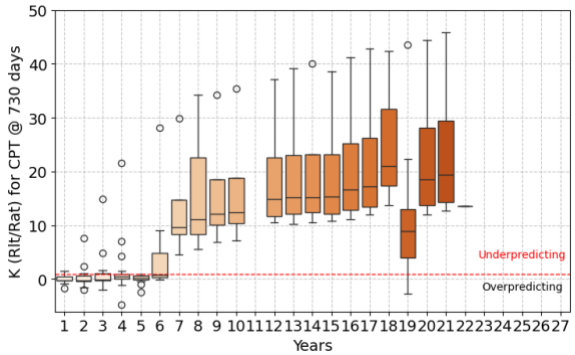
(b) CPT – Low alkali
(<2.94 kg/m^3)



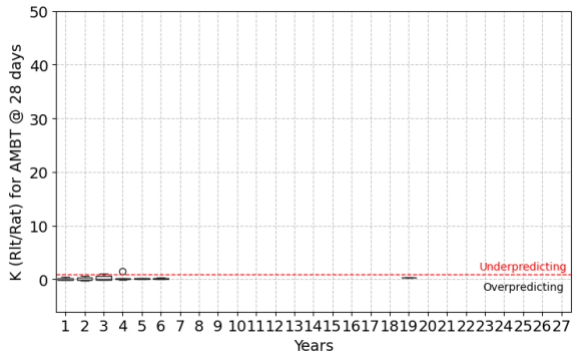
(c) AMBT – Moderate alkali
(>2.94 kg/m^3;<4.2 kg/m^3)



(d) CPT – Medium alkali
(>2.94 kg/m^3;<4.2 kg/m^3)



(e) AMBT – High alkali
(>4.2 kg/m^3)



(f) CPT – High alkali
(>4.2 kg/m^3)

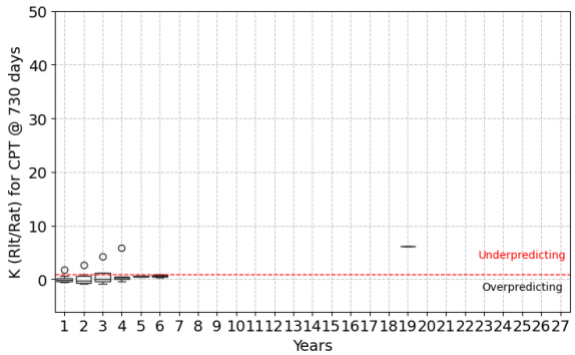


Figure 5. 14 – K coefficient for (a, c, e) AMBT and (b, d, f) CPT over 27 years of exposed concrete blocks for mixes with SCMs for (a, b) low, (c, d) moderate, (e, f) high alkali content.

Therefore, further enhancing test outcomes' prediction requires a comprehensive analysis that includes mix design details (i.e., alkali content, SCMs), aggregate employed and environmental conditions. In addition, incorporating these factors into multifactorial and

probabilistic models could significantly improve the reliability of predictions, supporting more effective decision-making in mitigating ASR risks.

5.8 Conclusion

Focusing on the accurate assessment of AAR in new concrete structures, this study evaluated the reliability of current testing methods. The key findings are as follows:

- For mixes without SCMs, AMBT (ASTM C1260) overpredicts reactivity due to accelerated testing conditions but ensures fewer reactive cases are missed, making it valuable when false negatives are to be avoided. CPT (ASTM C1293) provides more balanced field performance results but may miss some reactive cases. For mixes with SCMs, AMBT (ASTM C1260) remains reliable for detecting reactivity but may result in more false positives, while CPT (ASTM C1293) is more accurate for non-reactive cases but underestimates reactivity when SCMs are present.
- The impact of temperature in reactivity predictions shows that for mixes without SCMs, AMBT (ASTM C1260) tends to underpredict in warm conditions while remaining more stable in the overprediction zone in cold climates. CPT (ASTM C1293) transitions between underprediction and overprediction across both temperature regimes. For mixes with SCMs, AMBT (ASTM C1260) initially offers stable predictions but underpredicts after three years in warm conditions, while CPT (ASTM C1293) exhibits significant variability, transitioning between underprediction and overprediction after two years.
- The influence of alkali content is significant for both mixes without and with SCMs. For mixes without SCMs, AMBT (ASTM C1260) at low and moderate alkali levels tends to remain in the overprediction zone, while at high alkali content, AMBT (ASTM C1260) and CPT (ASTM C1293) at all levels (low, moderate, and high) transition from overprediction to underprediction. For mixes with SCMs, AMBT (ASTM C1260) stays in the overprediction zone for both low and high alkali contents, while CPT (ASTM C1293) at low alkali levels remains stable around K=1. However, AMBT (ASTM C1260) at moderate levels, and CPT (ASTM C1293) at moderate and high alkali levels, shift from overpredicting to underpredicting reactivity.

The observed variability and performance metrics suggest that traditional accelerated tests such as AMBT (ASTM C1260) and CPT (ASTM C1293) have limitations in predicting long-term ASR behaviour. The presence of SCMs and varying alkali content complicate predictions, highlighting the need for multifactorial and probabilistic models that incorporate environmental conditions, alkali content, and SCM effects to improve ASR prediction reliability and support better decision-making in mitigating ASR risks.

5.9 References

[1] B. Fournier, M.-A. Bérubé, Alkali-aggregate reaction in concrete: a review of basic concepts and engineering implications, Canadian Journal of Civil Engineering 27 (2000) 167–191. https://doi.org/10.1139/l99-072.

[2] P.J. Nixon, I. Sims, eds., RILEM Recommendations for the Prevention of Damage by Alkali-Aggregate Reactions in New Concrete Structures, Springer Netherlands, Dordrecht, 2016. https://doi.org/10.1007/978-94-017-7252-5.

[3] A. Leemann, M. Góra, B. Lothenbach, M. Heuberger, Alkali silica reaction in concrete - Revealing the expansion mechanism by surface force measurements, Cement and Concrete Research 176 (2024) 107392. https://doi.org/10.1016/j.cemconres.2023.107392.

[4] J. Lindgård, Ö. Andiç-Çakır, I. Fernandes, T.F. Rønning, M.D.A. Thomas, Alkali–silica reactions (ASR): Literature review on parameters influencing laboratory performance testing, Cement and Concrete Research 42 (2012) 223–243. https://doi.org/10.1016/j.cemconres.2011.10.004.

[5] I. Sims, P. Nixon, RILEM Recommended Test Method AAR-0: Detection of Alkali-Reactivity Potential in Concrete—Outline guide to the use of RILEM methods in assessments of aggregates for potential alkali-reactivity, Mat. Struct. 36 (2003) 472–479. https://doi.org/10.1007/BF02481527.

[6] J. Custódio, J. Lindgård, B. Fournier, A. Santos Silva, M.D.A. Thomas, T. Drimalas, J.H. Ideker, R.-P. Martin, I. Borchers, B. Johannes Wigum, T.F. Rønning, Correlating field and laboratory investigations for preventing ASR in concrete – The LNEC cube

study (Part I – Project plan and laboratory results), Construction and Building Materials 343 (2022) 128131. https://doi.org/10.1016/j.conbuildmat.2022.128131.

[7] C.A. MacDonald, C. Rogers, R.D. Hooton, The relationship between laboratory and field expansion - observations at the Kingston outdoor exposure site for ASR after twenty years, in: 14th International Conference on Alkali-Aggregate Reaction (ICAAR), Austin, Texas, USA, 2012.

[8] B. Fournier, R. Chevrier, A. Bilodeau, P.-C. Nkinamubanzi, N. Bouzoubaa, Comparative Field and Laboratory Investigations on the Use of Supplementary Cementing Materials (SCMs) to Control Alkali-Silica Reaction (ASR) in Concrete, in: H. de M. Bernardes, N.P. Hasparyk (Eds.), 15th International Conference on Alkali-Aggregate Reaction (ICAAR), São Paulo, Brazil, 2016.

[9] I. Borchers, J. Lindgård, C. Müller, Evaluation of laboratory test methods for assessing the alkali-reactivity potential of aggregates by field site tests, Materiales de Construcción 72 (2022) e286. https://doi.org/10.3989/mc.2022.17221.

[10] B. Fournier, J.H. Ideker, K.J. Folliard, M.D.A. Thomas, P.-C. Nkinamubanzi, R. Chevrier, Effect of environmental conditions on expansion in concrete due to alkali–silica reaction (ASR), Materials Characterization 60 (2009) 669–679. https://doi.org/10.1016/j.matchar.2008.12.018.

[11] J. Lindgard, M.D.A. Thomas, E.J. Sellevold, B. Pedersen, O. Andic-Cakir, H. Justnes, T.F. Ronning, Alkali-silica reaction (ASR)-performance testing: Influence of specimen pre-treatment, exposure conditions and prism size on alkali leaching and prism expansion, CEMENT AND CONCRETE RESEARCH 53 (2013) 68–90. https://doi.org/10.1016/j.cemconres.2013.05.017.

[12] ASTM C1260, Test Method for Potential Alkali Reactivity of Aggregates (Mortar-Bar Method), ASTM International, 2022. https://doi.org/10.1520/C1260-22.

[13] ASTM C1293, Test Method for Determination of Length Change of Concrete Due to Alkali-Silica Reaction, (2024). https://doi.org/10.1520/C1293_C1293M-23A.

[14] R.E. Oberholster, G. Davies, An accelerated method for testing the potential alkali reactivity of siliceous aggregates, Cement and Concrete Research 16 (1986) 181–189. https://doi.org/10.1016/0008-8846(86)90134-1.

[15] ASTM C1778, Guide for Reducing the Risk of Deleterious Alkali-Aggregate Reaction in Concrete, ASTM International, 2022. https://doi.org/10.1520/C1778-22.

[16] M. Thomas, B. Fournier, K. Folliard, J. Ideker, M. Shehata, Test methods for evaluating preventive measures for controlling expansion due to alkali–silica reaction in concrete, Cement and Concrete Research 36 (2006) 1842–1856. https://doi.org/10.1016/j.cemconres.2006.01.014.

[17] M.-A. Bérubé, B. Fournier, Canadian experience with testing for alkali-aggregate reactivity in concrete, Cement and Concrete Composites 15 (1993) 27–47. https://doi.org/10.1016/0958-9465(93)90037-A.

[18] S. Multon, M. Cyr, A. Sellier, P. Diederich, L. Petit, Effects of aggregate size and alkali content on ASR expansion, Cement and Concrete Research 40 (2010) 508–516. https://doi.org/10.1016/j.cemconres.2009.08.002.

[19] J. Lindgård, T. Østnor, B. Fournier, Ø. Lindgård, T. Danner, G. Plusquellec, K. De Weerdt, Determining alkali leaching during accelerated ASR performance testing and in field exposed cubes using cold water extraction (CWE) and μXRF, MATEC Web Conf. 199 (2018) 03004. https://doi.org/10.1051/matecconf/201819903004.

[20] S.U. Einarsdottir, R.D. Hooton, Modifications to ASTM C1293 That Allow Testing of Low-Alkali Binder Systems, ACI Materials Journal 115 (2018). https://doi.org/10.14359/51702350.

[21] J.H. Ideker, B.L. East, K.J. Folliard, M.D.A. Thomas, B. Fournier, The current state of the accelerated concrete prism test, Cement and Concrete Research 40 (2010) 550–555. https://doi.org/10.1016/j.cemconres.2009.08.030.

[22] B. Fournier, P.C. Nkinamubanzi, D. Lu, M.D.A. Thomas, K.J. Folliard, J.H. Ideker, Evaluating Potential Alkali-Reactivity of Concrete Aggregates – How Reliable are the Current and New Test Methods?, in: S. Kuperman, N.P. Hasparyk (Eds.), II Simpósio Sobre Reação Álcali-Agregado Em Estruturas de Concreto, Rio de Janeiro, Brasil, 2006.

[23] N. 32 Norwegian Concrete Association, Alkali–aggregate reactions in concrete, Test methods and Requirements to Test Laboratories, (2005).

[24] T.F. Ronning, B.J. Wigum, J. Lindgard, Recommendation of RILEM TC 258-AAA: RILEM AAR-10: determination of binder combinations for non-reactive mix design using concrete prisms-38 degrees C test method, MATERIALS AND STRUCTURES 54 (2021). https://doi.org/10.1617/s11527-021-01679-w.

[25] T.F. Rønning, J. Lindgård, I. Borchers, ASR performance testing concepts – RILEM AAR-10, AAR-11 and AAR-12, in: A.L. Batista, A.S. Silva, I. Fernandes, L.O. Santos, J. Custodio, C. Serra (Eds.), 16th International Conference on Alkali-Aggregate Reaction in Concrete, Lisbon, Portugal, 2022.

[26] S. Stacey, K. Folliard, T. Drimalas, M.D.A. Thomas, An accelerated and more accurate test method to ASTM C1293: The concrete cylinder test, in: H. de M. Bernardes, N.P. Hasparyk (Eds.), 15th International Conference on Alkali-Aggregate Reaction (ICAAR), São Paulo, Brazil, 2016.

[27] J. Tanesi, T. Drimalas, K.S.T. Chopperla, M. Beyene, J.H. Ideker, H. Kim, L. Montanari, A. Ardani, Divergence between Performance in the Field and Laboratory Test Results for Alkali-Silica Reaction, Transportation Research Record 2674 (2020) 120–134. https://doi.org/10.1177/0361198120913288.

[28] T.N. Nguyen, L.F.M. Sanchez, J. Li, B. Fournier, V. Sirivivatnanon, Correlating alkali-silica reaction (ASR) induced expansion from short-term laboratory testings to long-term field performance: A semi-empirical model, Cement and Concrete Composites 134 (2022). https://doi.org/10.1016/j.cemconcomp.2022.104817.

[29] A. Santos Silva, D. Soares, L. Matos, I. Fernandes, M.M. Salta, Alkali-Aggregate Reactions in Concrete: Methodologies Applied in the Evaluation of Alkali Reactivity of Aggregates for Concrete, Materials Science Forum 730–732 (2012) 409–414. https://doi.org/10.4028/www.scientific.net/MSF.730-732.409.

[30] M.-A. Berube, J. Frenette, A. Pedneault, M. Rivest, Laboratory Assessment of the Potential Rate of ASR Expansion of Field Concrete, Cement, Concrete and Aggregates 24 (2002) 13. https://doi.org/10.1520/CCA10486J.

[31] N. Whiting, Comparison of Field Observations with Laboratory Test Results on Concretes Undergoing Alkali-Silica Reaction, Cement, Concrete and Aggregates 21 (1999) 142. https://doi.org/10.1520/CCA10427J.

[32] D. Hooton, B. Fournier, Long-term alkali–silica mitigation of high-alkali concrete with cement replacements, Proceedings of the Institution of Civil Engineers - Construction Materials 175 (2022) 125–136. https://doi.org/10.1680/jcoma.21.00049.

[33] J. Lindgård, P.J. Nixon, I. Borchers, B. Schouenborg, B.J. Wigum, M. Haugen, U. Åkesson, The EU "PARTNER" Project — European standard tests to prevent alkali reactions in aggregates: Final results and recommendations, Cement and Concrete Research 40 (2010) 611–635. https://doi.org/10.1016/j.cemconres.2009.09.004.

[34] J.H. Ideker, A.F. Bentivegna, K.J. Folliard, M.C.G. Juenger, Do Current Laboratory Test Methods Accurately Predict Alkali-Silica Reactivity?, ACI Materials Journal 109 (2012). https://doi.org/10.14359/51683914.

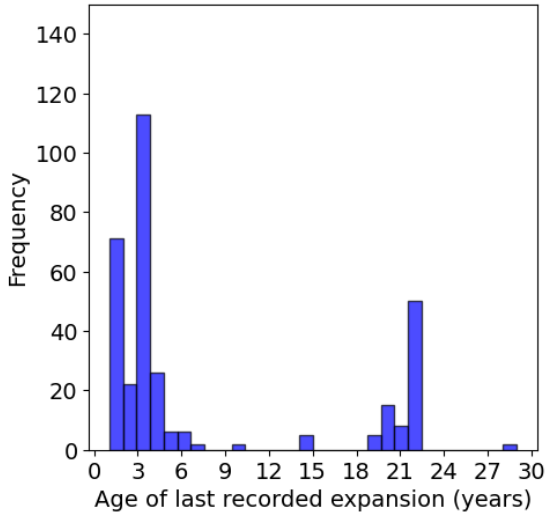
[35] J. Lazar, J.H. Feng, H. Hochheiser, Statistical analysis, in: Research Methods in Human Computer Interaction, Elsevier, 2017: pp. 71–104. https://doi.org/10.1016/B978-0-12-805390-4.00004-2.

[36] J. Lindgård, T.F. Rønning, M.D.A. Thomas, B. Fournier, A.S. Silva, ASR - Performance testing: main findings in the Norwegian COIN project, in: H. de M. Bernardes, N.P. Hasparyk (Eds.), 15th International Conference on Alkali-Aggregate Reaction, São Paulo, Brazil, 2016.

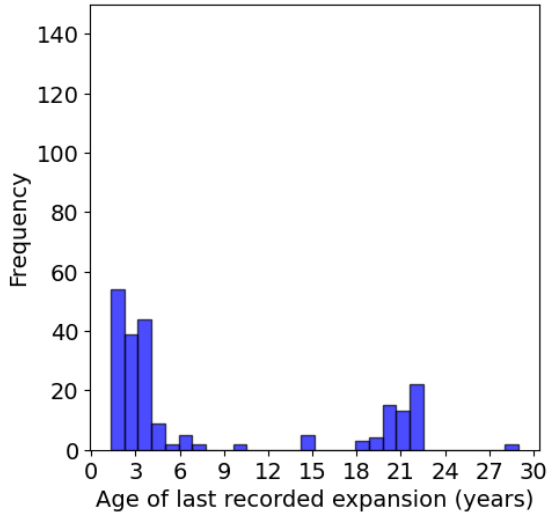
[37] B. Fournier, J. Lindgård, B.J. Wigum, I. Borchers, Outdoor exposure site testing for preventing Alkali-Aggregate Reactivity in concrete – a review, MATEC Web of Conferences 199 (2018) 03002. https://doi.org/10.1051/matecconf/201819903002.

5.10 Appendix A – Distribution of field age of the evaluated blocks

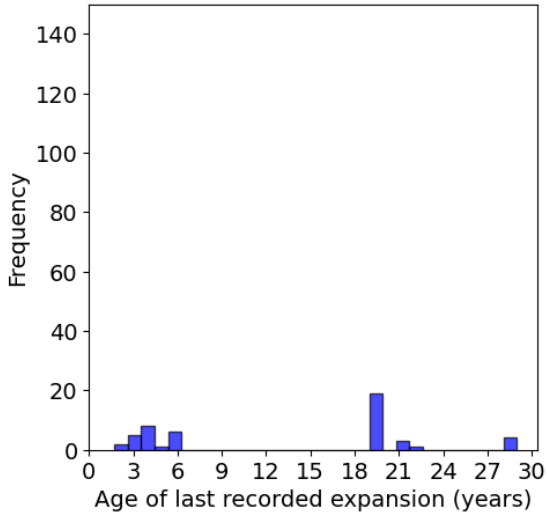
(a) CPT without SCMs



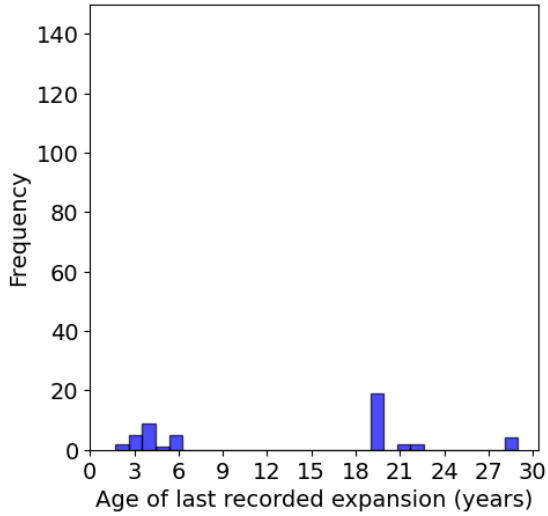
(b) AMBT without SCMs



(c) CPT with SCMs



(d) AMBT with SCMs



Chapter 6: Multifactorial analysis of AAR development: Integrating laboratory and field data with statistical and probabilistic modelling

Ana Bergmann and Leandro F. M. Sanchez

Abstract: Alkali aggregate reaction is amongst the most harmful durability issues affecting concrete infrastructure worldwide. Preventative measures are the most effective strategy against AAR, as remediation is challenging once the reaction is triggered; however, while accelerated laboratory tests, such as the AMBT and CPT, are commonly used to assess aggregates' reactivity, discrepancies between laboratory results and field performance remain unquantified. Therefore, this paper presents a multifactorial analysis integrating laboratory and field data with statistical and probabilistic modelling to assess the risk of AAR occurrence in the field. Through traditional statistical methods, Bayesian analysis, and Beta distribution model, the impact of parameters leading to AAR development was assessed resulting in the probability distribution of AAR occurrence given laboratory test outcomes, environment and alkali loading. The findings show that AMBT outperforms in identifying non-reactive cases for both mixes with and without SCMs, while both AMBT and CPT demonstrate similar reliability for detecting reactive cases, with CPT slightly better at higher alkali levels. However, misclassification risks from both tests increase in warm environments with high alkali levels and no SCMs, whereas cold conditions with low alkali and SCMs improve the reliability of reactive outcome classification.

Keywords: Alkali-Aggregate Reaction, Probabilistic modelling, Bayesian analysis, Logistic regression

6.1 Introduction

Alkali-aggregate reaction (AAR) is recognized as one of the most critical deterioration mechanisms affecting concrete infrastructures worldwide; it is currently reported in over 50 countries [1–7]. Both types of AAR—alkali-silica reaction (ASR) and alkali-carbonate reaction (ACR)— are triggered when dissolved alkali hydroxides present in the concrete pore

solution react with unstable mineral phases within the aggregates [8,9]. Over the years, it became quite evident that the best approach against AAR is its prevention; in this, context, a wide number of laboratory test procedures were developed to appraise the potential reactivity of aggregates and the efficiency of preventive measures prior to concrete pouring.

Amongst those, the accelerated mortar bar test (AMBT) and the concrete prism test (CPT) are widely employed nowadays [8,10]. These methods classify aggregates based on their reactivity levels, ranging from non-reactive to very highly reactive [11]. However, numerous studies have highlighted significant discrepancies between laboratory test outcomes and the actual performance of concrete in field conditions [12–18]. Although the extent of these discrepancies remains unquantified, the accumulation of data gathered from field exposure has increased substantially in the last decades [14,15,19,20]. Given these discrepancies, it is crucial to assess the reliability of laboratory tests in predicting field performance. This includes understanding the factors influencing the variations in reliability, such as alkali content, aggregate type, and environmental conditions beyond current descriptive and deterministic approaches (i.e., mean, median, standard deviation, etc.). Consequently, integrating stochastic methods and artificial intelligence to analyze the parameters may offer more precise predictions of AAR occurrence in the field, thereby enhancing the understanding of the risks associated with AAR in new concrete structures.

In this sense, this study aims to evaluate the probability of AAR occurrence in field conditions based on laboratory test results, i.e., AMBT and CPT. Bayesian analysis and Beta distribution models are employed to estimate the probability of AAR occurrence while incorporating data uncertainties. Additionally, logistic regression is utilized to predict field reactivity likelihood based on laboratory outcomes, alkali loading, and environmental conditions. This probabilistic approach provides a nuanced understanding of AAR risk, offering a robust framework for preventing and mitigating AAR in concrete structures.

6.2 Background

6.2.1 Laboratory and field tests to appraise AAR potential

Alkali-aggregate reaction (AAR) represents a significant durability issue for concrete structures, triggered by the reaction of the alkali hydroxides present in the concrete pore solution with unstable mineral phases encountered withing the aggregates, leading to expansion and cracking [21]. AAR is developed in two forms: alkali-silica reaction (ASR) and alkali-carbonate reaction (ACR). ASR involves the reaction of alkali hydroxides with reactive silica in aggregates, forming a secondary reaction product, i.e., ASR gel, that swells upon moisture uptake inducing internal stresses and resulting in cracking [9,22,23]. ACR, although less common and understood, involves the reaction of alkali hydroxides with carbonate rocks, such as dolomitic limestone, leading to the dedolomitization of the aggregate and subsequent expansion [24,25].

Aiming to predict the potential for AAR development in the field, accelerated laboratory testing methods have been proposed. These tests typically involve variations in sample characteristics (e.g., shape, dimensions), storage conditions (e.g., temperature, relative humidity, solution alkali concentration), and mix design (e.g., crushed materials, additional alkalis) [10,26]. Additionally, their primary objective is to accelerate the reaction kinetics, allowing for the rapid determination of aggregate reactivity, the potential expansion of concrete mixtures, and the assessment of mitigation effectiveness, such as the use of supplementary cementitious materials (SCMs). Among the accelerated tests, the AMBT and the CPT are the most widely used.

The AMBT is performed by immersing mortar bars made in a high-alkali solution (1M NaOH) at 80°C, measuring expansion over 14 to 28 days for mixtures without and with SCMs, respectively [27,28]. This accelerated method provides rapid results, commonly suggested for initial screenings of aggregate reactivity. The threshold for considering aggregates as reactive in the AMBT is set at 0.1% expansion [11]. However, the test is questioned for exposing unstable minerals due to crushing and increasing the aggregate surface area, while modifying their texture, which could lead to higher expansion levels than

in concrete mixtures [29,30]. Additionally, the severe test conditions and short duration may overestimate reactivity and not capture slower reactions.

In contrast, the CPT uses concrete prisms and measures expansion over a longer period, usually 1 to 2 years (without and with SCMs, respectively), under conditions of slightly elevated temperature (i.e., 38°C) and high relative humidity (i.e., 95%) [31]. The threshold for reactivity in the CPT is generally set at 0.04% expansion [11]. Although the long duration of the CPT allows for capturing long-term effects, its major issue is alkali leaching, which reduces the alkali content of the specimens over time, leading to lower ultimate expansion measurements and thus reactivity of the aggregates [32].

To assess and calibrate accelerated laboratory results, exposure sites were developed. These sites, where concrete members are subjected to natural environmental conditions, provide valuable data to assess the performance of laboratory tests alongside actual field performance. For instance, the importance of long-term field exposure for understanding ASR development under varying environmental conditions has been reported to significantly influence the reaction kinetics ultimate expansion and induced deterioration [33]. Moreover, the limitations of accelerated laboratory tests underscore the need to quantify their reliability, through advanced analytical methodologies.

6.2.2 Analytical methodologies

Uncertainty is an unavoidable aspect of engineering, arising from the inherent randomness in natural processes (i.e., aleatory uncertainty) and the limitations in knowledge or predictive models (i.e., epistemic uncertainty), which are managed through statistical and probabilistic methods [34]. These approaches allow engineers to quantify the likelihood of various outcomes and assess the risks associated with them, leading to more informed decisions. In summary, statistics can be understood as the science of data. This involves collecting, summarizing, organizing, analyzing, interpreting data, and building models [35,36]. Consequently, statistical methods are mathematical tools used to analyze, interpret, and make inferences from data. These methods are essential in research for identifying patterns, testing hypotheses, and making decisions based on data. Statistical methods can be broadly

categorized into two general branches, including descriptive statistics and inferential statistics [36,37].

Descriptive statistics focuses on data collection, organization and presentation. However, it doesn't allow making conclusions beyond the analyzed data. In descriptive statistics, three general methods are used, including measures of central tendency (i.e., mean, median, and mode), measures of spread (i.e., range, quartiles and percentiles, variance, and standard deviation) and graphical displays (i.e., dot plots, frequency distribution, histograms, frequency polygons, bar graphs, pie chart, box and whiskers plot) [37].

In contrast, inferential statistics makes use of the features of a sample to make assumptions about the unknown parameters in a certain population [38]. As a result, it is concerned with drawing conclusions based on a dataset by extending beyond the available information. In inferential statistics, both deterministic and probabilistic approaches can be employed. Conversely, a fixed threshold can be used to deterministically label an aggregate as reactive or non-reactive. In contrast, probabilistic analysis incorporates the variability to estimate the likelihood of various outcomes rather than a single fixed result (i.e., the probability of being reactive in the field or not). Moreover, inferential statistics methods include sampling theory, estimation theory, hypothesis testing, and regression analysis. In estimation theory, the prediction or estimation of a population parameter based on the available data is the target relying on point estimate, interval estimate and confidence interval techniques. Additionally, it accounts for techniques such as maximum likelihood estimation and minimum mean squared error estimation to make assumptions. Given the inherent variability and uncertainty of datasets, Bayesian interference is generally employed to create a probability model for a set of data, summarizing results as probability distributions of parameters [39,40]. Otherwise, regression analysis reveals the mathematical relationship between variables finding an expression to best represent the dataset under evaluation.

This study performs a multifactorial analysis by integrating laboratory and field data with statistical and probabilistic modelling techniques. By implementing Bayesian analysis and Beta distribution models, the probability of AAR occurrence in field conditions based on laboratory test results is evaluated. Risks associated with the use of laboratory tests to predict field performance are then computed.

6.3 Scope of the work

The primary objectives of this work are to evaluate the reliability of AMBT and CPT outcomes by assessing the probability of AAR occurrence in the field based on their outcomes. To achieve this goal, the methodology involves the following steps:

- Data collection and database development: a comprehensive database encompasses both laboratory and field conditions, including variables such as alkali content, aggregate type (i.e., reactivity levels), presence and amount of supplementary cementitious materials (SCMs), and environmental factors (i.e., temperature and relative humidity);
- Model development and validation: descriptive statistical analysis summarizes and understands the collected data, identifying the central tendency and distribution of key parameters. Correlation analysis identifies relationships between variables influencing AAR development. Bayesian analysis and Beta distribution develop a probabilistic model to estimate the likelihood of AAR occurrence, incorporating uncertainties in the collected data.
- Results interpretation and application: probability distributions for mixtures without and with SCMs are analyzed, and the implications of the probabilistic estimates for field reactivity occurrence are discussed. Finally, the limitations of the study and directions for future research are presented.

The proposed approach provides a robust framework for bridging the gap between laboratory testing and field performance, offering a measurable understanding of the associated risks of developing AAR in the field given the use of an accelerated test outcome, and contributing to the development of effective mitigation strategies for new concrete structures.

6.4 Data Collection and Database

This study integrates data from both laboratory experiments and field measurements obtained from published documents, including journal papers, conference papers, and technical reports, covering a period from 2001 to 2024, reflecting results of exposed fields from the early 1990s [13,16,17,20,41–48]. Only datasets containing comprehensive information from laboratory to field conditions were included to ensure consistency and reduce errors. It is important to note that concrete members exposed in the field are referred to as “blocks”.

The inclusion criteria required detailed descriptions of both the testing protocols and the resulting outcomes. Only results from standardized laboratory protocols were selected to minimize variations in testing conditions, focusing on AMBT and CPT as the most used worldwide. Therefore, the dataset includes detailed information on test conditions, variables such as concrete alkali content, aggregate type, presence and amount of SCMs, and expansion measurements over time. For exposed concrete members, additional data on environmental conditions, such as location, temperature, and humidity were also collected.

More specifically, Figure 6. 1 represents the information for laboratory and field included in the dataset.

Laboratory	Field
• Test condition: Temperature, solution concentration, and relative humidity. • Sample details: Geometry, dimensions, volume, and surface area. • Aggregate details: Type (i.e., fine or coarse) and source (i.e., country or region of origin). • SCMs: Type and amount. • Expansion measurements: Laboratory outcomes (i.e., expansion at a given time), last measured expansion, age of last measured expansion and expansion over time.	• Sample details: Geometry, dimension, volume, surface area, and reinforcement. • Aggregate: Type and origin. • Supplementary cementitious materials (SCMs): Type and amount. • Alkali content: Amount. • Admixtures: Type and amount. • Environment conditions: Annual average temperature and relative humidity. • Expansion measurements: Last measured expansion, age of last measured expansion and expansion over time.

Figure 6. 1 – Database information for laboratory and field entries.

An entry example demonstrating the detailed information collected is presented hereafter:

- Laboratory: Aggregate tested: Spratt, ON, Aggregate type: Coarse, SCMs: None, SCMs amount: 0%, Sample shape: prism, Sample surface area (m^2): 0.10125, Sample volume (m^3): 0.0016875, Test name: CPT, Test temperature (^oC): 38, Solution concentration (Mol): N/A, External humidity (%): 95, Laboratory outcome expansion (%): 0.184, Laboratory last

measured expansion (%): 0.225, Laboratory age of last recorded expansion (days): 2190, Expansion over time.

- Field: Mix-design name: M25-SP+, Aggregate tested: Spratt, ON, Aggregate type: Coarse, Concrete alkali content (kg/m^3): 5.225, SCMs: None, SCMs amount: 0%, Admixture: Air entrained: Sample shape: beam, Sample surface area (m^2): 1.44, Sample volume (m^3): 0.112, Reinforcement: none, Environment: Ottawa, Field annual average temperature (^{o}C): 6.9, Field annual average relative humidity (%): 75, Field last measured expansion (%): 0.349, Field age of last recorded expansion (days): 6935, Expansion over time.

As previously mentioned, the dataset primarily reflects results from AMBT [27] and CPT [31] tests, which are the focus of this study. The classification of aggregate reactivity for each test follows ASTM C1778 standard protocol [11], with thresholds of 0.10% and 0.04% for AMBT and CPT at 14 and 365 days, respectively.
To enhance the evaluation of parameters, the data are subdivided as follows:

- No SMCs: Reactivity levels (R0, R1, R2, R3) as per ASTM C1778; alkali-content (low, moderate, high); environment.
- With SMCs: Alkali-content (low, moderate, high); SCMs type; environment.

Finally, the main variables used for the model development are summarized in Appendix A. It is worth noting that for the binary variable of field reactivity (i.e., YES | NO), a threshold of 0.04% was selected, given that this is the expansion level at which concrete typically starts showing visible external cracks. Additionally, the last recorded expansion data was adopted as the reference for the field's ultimate expansion in this analysis, reflecting the entire history of the reaction for each member under evaluation.

6.5 Model development and validation

To develop a probability model for the occurrence of field reactivity, the following steps were performed (Figure 6. 2).

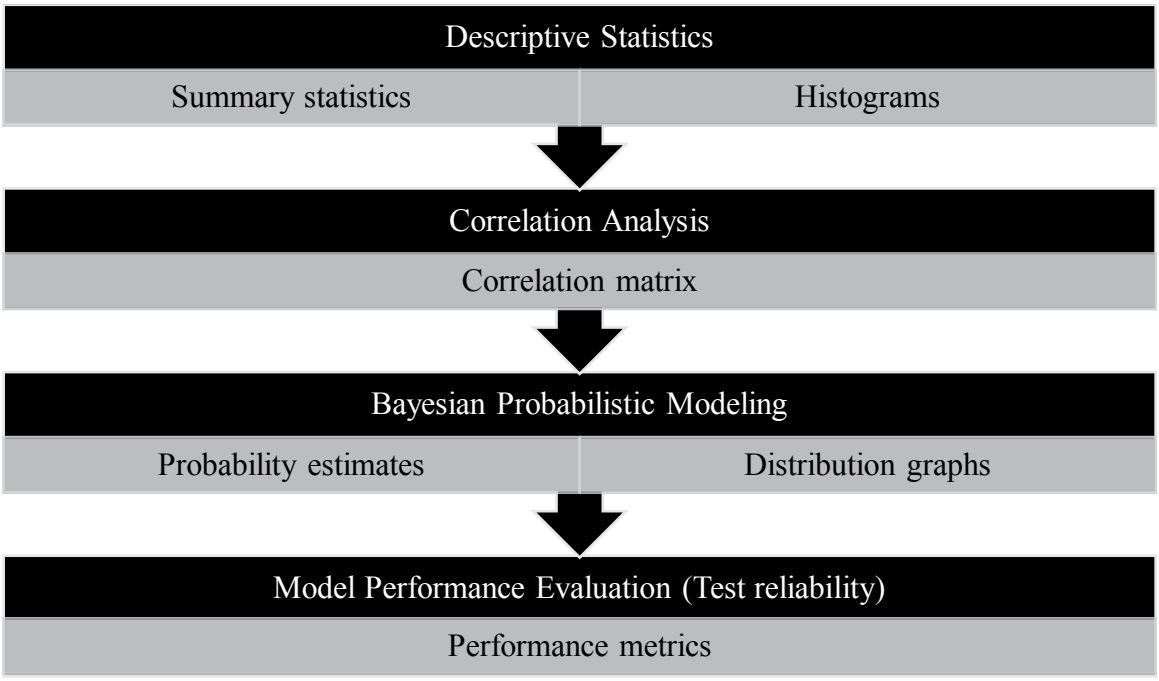


Figure 6. 2 – Model development and validation steps and outcomes.

Initially, descriptive statistics were calculated to summarize the central tendency, dispersion, and distribution of the dataset, highlighting key characteristics such as environmental conditions, mix design, and laboratory and field performance. Pearson correlation coefficients identified linear relationships between variables, determining key correlations for reactivity with and without SCMs. Then, a Bayesian probabilistic modelling approach was employed to estimate reactivity probabilities, incorporating uncertainties and calculating confidence intervals via Beta distribution. Finally, the model performance was evaluated using Brier score to measure the probabilistic prediction accuracy. This comprehensive approach supports understanding the reliability of laboratory outcomes in predicting AAR occurrence, aiding in risk management and decision-making.

6.6 Descriptive statistics

Descriptive statistics, including mean, median, standard deviation, variance, and range, were calculated to summarize the central tendency, dispersion, and distribution shape of the data (i.e., skewness calculation). Histograms were used to visualize the distribution of each variable within the subsets of the dataset. Only blocks without reinforcement, not confined and without mitigation admixtures were selected for this study. The analysis was performed for two main groups: without SCMs and with SCMs.

Dataset without SCMs

Figure 6. 3 and Appendix B highlight the distributions for the groups without SCMs, consisting of 867 records. The numerical parameters evaluated for the descriptive statistics provide insights into the environment (i.e., field temperature and RH), mix-design (i.e., alkali content, aggregate type), and laboratory and field performance (i.e., expansion).

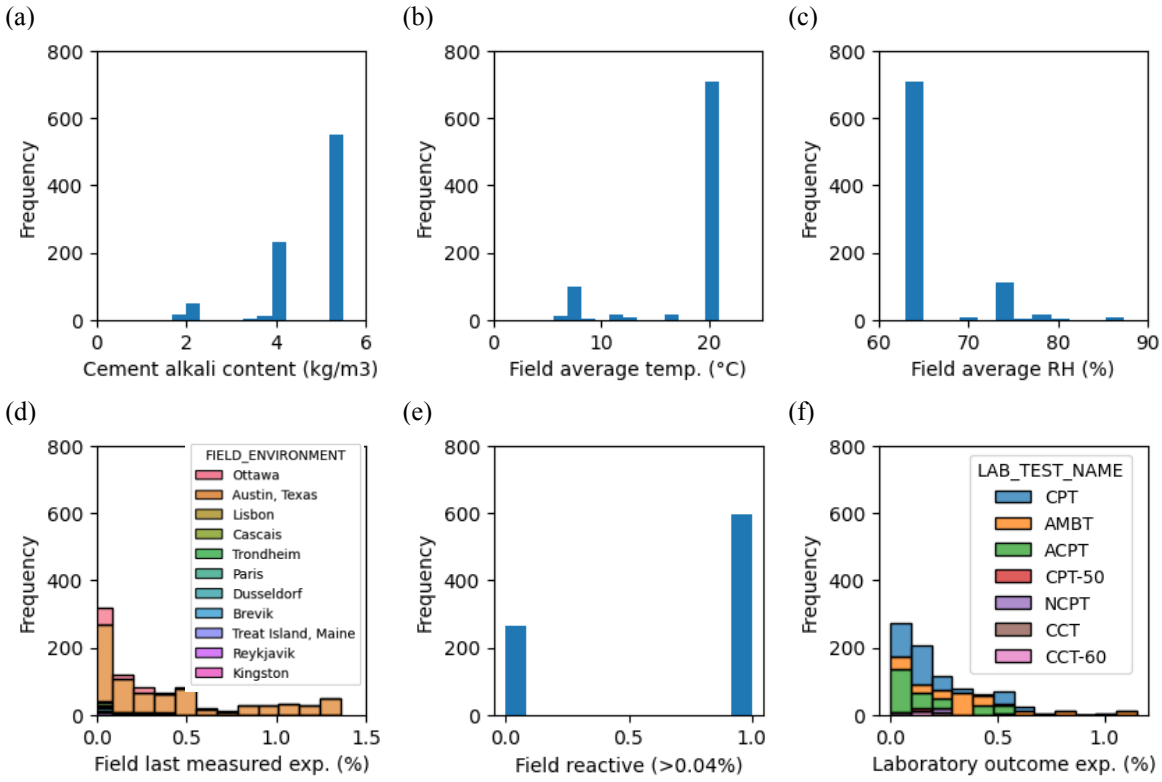


Figure 6. 3 – Histograms for the subset data without SCMs: (a) alkali content; (b) field annual average temperature; (c) field annual average relative humidity; (d) field last measured expansion; (e) field reactive over the threshold of 0.04%, 0 for No and 1 for Yes; (f) laboratory outcome expansion.

Considering the environment, the annual average field temperature is 18.3°C, with values ranging from 5.6°C to 21°C. Two main groups of field exposure are identified: colder and warmer environments (6°C and 21°C, respectively). The relative humidity values are high, varying from 63% to 87.3%. For the mix design, the mean alkali content of the cement used (i.e, normally Portland Cement GU Type in Canada or ASTM Type I in the USA) considering extra alkalis in boosted systems is 4.68 kg/m3 (SD = 0.93 kg/m3), with minimum and maximum values of 1.67 kg/m3 and 5.5 kg/m3, respectively. The skewness of -1.51 suggests a left-skewed distribution where values are mostly high with a few low records. Coarse aggregates constitute 65.2% of the tested samples (Appendix C). Yet, Jobe - El Paso, Texas, a fine aggregate, is the most frequently tested aggregate (15.7% of samples), followed by Placitas - Albuquerque, New Mexico; Wright – Robstown, Texas; and Spratt, ON, with 111 (12.9%), 96 (11.1%) and 87 (10.1%) occurrences, respectively.

Laboratory expansion comprises results from 7 different test setups (i.e., CPT, ACPT, AMBT, CPT-50, NCPT, CCT, CCT-60). For the CPT, the mean expansion is 0.19% (SD = 0.17%), with values ranging from 0.00% to 0.59%. The interquartile range (IQR) is from 0.08% to 0.26% and has a moderately right-skewed distribution (1.13), where most values are low with few high expansions. For the AMBT, the mean expansion is much higher, equal to 0.37% (SD = 0.28%), with values ranging from 0.02% to 1.15%. The IQR is from 0.17% to 0.45% and has a right-skewed distribution (1.27).

Finally, the mean expansion last recorded in the field is 0.36% (SD = 0.41%), with values ranging from -0.02% to 1.36%. The IQR spans from 0.02% to 0.53% and presents a right-skewed distribution (1.13) where most values are low with a few higher values stretching the distribution. Approximately 69% of the exposed blocks are reactive, based on the 0.04% threshold.

Dataset with SCMs

Figure 6. 4 and Appendix D highlight the distributions for the groups with SCMs, consisting of 157 records. The descriptive statistics for numerical variables indicate environmental parameters (i.e., field temperature and RH), mix design (i.e., alkali content, aggregate type), and performance metrics (i.e., expansion).

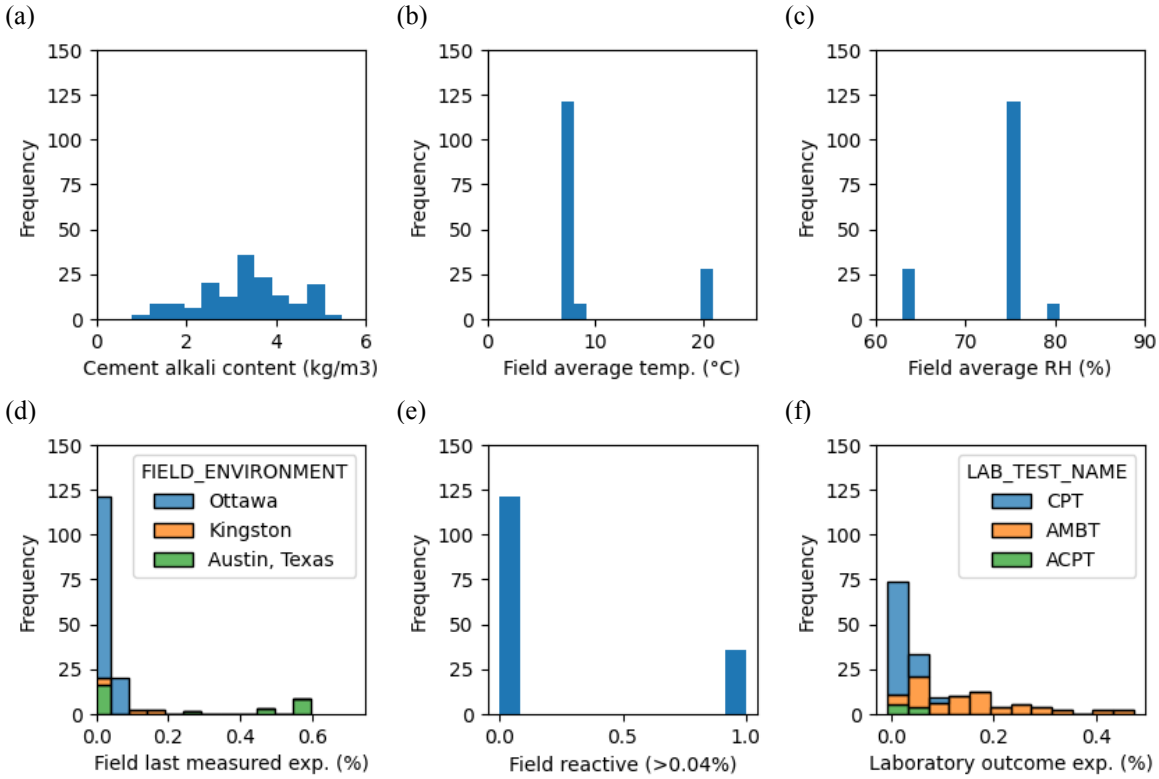


Figure 6. 4 – Histograms for the subset data with SCMs: (a) alkali content; (b) field annual average temperature; (c) field annual average relative humidity; (d) field last measured expansion; (e) field reactive over the threshold of 0.04%, 0 for No and 1 for Yes; (f) laboratory outcome expansion.

The above plots show that the annual average field temperature is 9.49°C (SD = 5.39°C), with values ranging from 6.9°C to 21°C. As observed, most samples are from colder environments, i.e., lower temperatures. Furthermore, the annual RH average value is 73.14% (SD = 4.89%), with values ranging from 63% to 80.5%. Regarding the mix design used, the mean alkali content present in the cement considering extra alkalis from the boosting process is 3.36 kg/m^3 (SD = 1.01 kg/m^3), with minimum and maximum values of 0.79 kg/m^3 and 5.47 kg/m^3. The IQR for alkali content is between 2.73 kg/m^3 and 4.16 kg/m^3. The skewness of -0.26 suggests a slightly left-skewed distribution close to being symmetric. The most frequently tested aggregate (Appendix E) is Spratt, ON**,** a coarse aggregate, with 36 occurrences, making up 22.9% of the total samples. The mean SCMs amount is 28.73% (SD = 17.01%), with values ranging from 7.5% to 76%, indicating a wide variability in the SCM

content (Appendix F). The SCMs recorded are Fly Ash (F, C, Ultrafine), Silica Fume, Slag, Slag + Silica Fume, Metakaolin, Fly as C + Silica Fume, and Fly ash C + Ultrafine fly ash.
Laboratory expansion comprises results from 3 different test setups (i.e., CPT, ACPT, AMBT). For the CPT, the mean expansion is 0.02% (SD = 0.02%), with values ranging from 0.00% to 0.09%. The interquartile range (IQR) is from 0.01% to 0.03% and has a right-skewed distribution (1.41), where most values are low with few high expansions. For the AMBT, the mean expansion is much higher, equal to 0.15% (SD = 0.11%), with values ranging from 0.01% to 0.47%. The IQR is from 0.06% to 0.22% and has a right-skewed distribution (0.93).
Finally, the mean expansion last recorded in the field is 0.06% (SD = 0.14%), with values ranging from -0.01% to 0.60%. The IQR spans from 0.01% to 0.03%, indicating a high right skew (3.18). Considering the threshold of 0.04% to indicate whether the concrete member displays a reactive behaviour, the field reactive binary variable indicates that approximately 23% of the exposed blocks are reactive.

6.7 Correlation analysis

Pearson correlation coefficients were computed to measure the linear relationship between variables. A correlation matrix was constructed to visualize these relationships, aiming to identify key indicators of the reactivity in each scenario (i.e. with and without SCMs). For the subset without SCMs (Figure 6. 5), the reactivity binary outcome is positively correlated with alkali content (0.22) and laboratory outcome expansion (0.22), confirming their association with higher field reactivity. Conversely, relative humidity shows a negative correlation with field's last expansion (-0.27). Additionally, the positive correlation between field temperature and last measured expansion (0.26) suggests that higher temperatures contribute to greater field expansion. Furthermore, the positive correlation between laboratory and field last measured expansion (0.46) indicates that generally higher laboratory expansions are associated with greater field expansions.

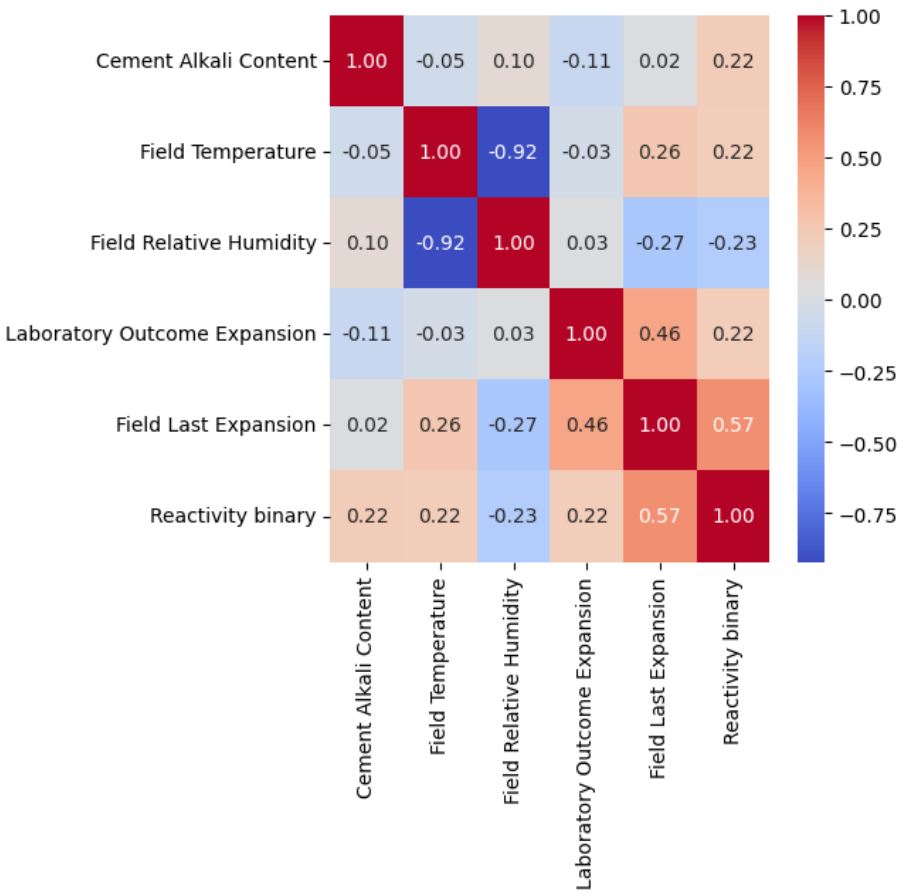


Figure 6. 5 – Correlation matrix for the numeric variables of the subset without SCMs.

In the subset with SCMs (Figure 6. 6), stronger relationships are observed. For instance, higher field expansion is associated with higher field temperature (0.59). However, relative humidity shows a negative correlation with the field's last measured expansion (-0.55). The effect of SCMs is evident, with high negative correlations with alkali content (-0.78) and moderate negative correlation with laboratory outcome expansion (-0.26), as anticipated.

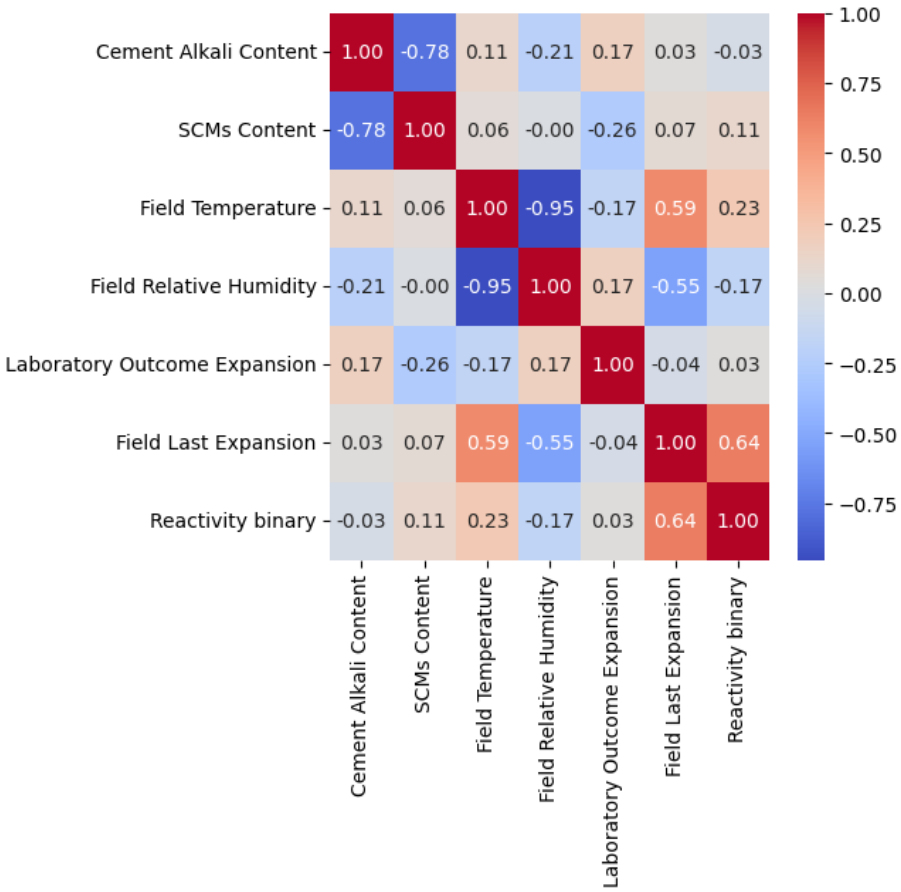


Figure 6. 6 – Correlation matrix for the numeric variables of the subset with SCMs.

6.8 Probabilistic modelling: Bayes theorem and Beta distribution

Bayesian inference is generally employed to fit a probability model to a set of data, summarizing results as probability distributions on parameters [39,40]. The Bayesian data analysis process involves three key steps: (i) setting up a full probability model (i.e., prior distribution); (ii) conditioning on observed data (i.e., likelihood function acting on the posterior distribution); and, (iii) evaluating the fit of the model [39]. In the current context, Bayes' Theorem [40] was utilized to identify the probability of the exposed concrete members being reactive giving the current test outcomes (i.e., AMBT and CPT) and thresholds [11].

In the current study, Bayesian inference was based on a data-driven prior derived from the observed data. The likelihood function was calculated by dividing the number of reactive

samples by the total number of reactive samples, while the marginal likelihood represented the ratio of total samples in each category to the total number of samples. The posterior probability was then computed using Bayes' theorem, incorporating the prior, likelihood, and marginal likelihood. To account for uncertainty, confidence intervals around the posterior probability were estimated using a neutral or non-informative beta distribution (i.e., Beta(1,1)) based on the observed counts of reactive and non-reactive blocks, as seen in Figure 6. 7.

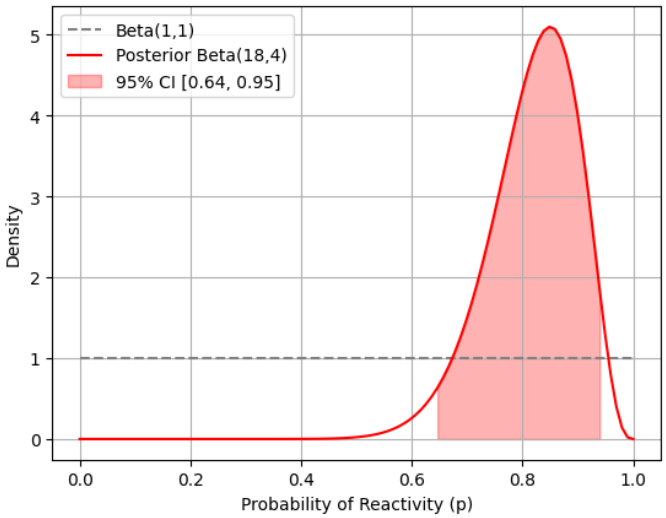


Figure 6. 7 – Beta and posterior distribution with CI.

Prior probability

The prior probability $P(R)$ represents the initial belief about the probability that the dependent variable, i.e., whether the block is reactive (i.e., YES | NO), given the conditional variables (i.e., level of reactivity, environment and alkali content of the binder). As seen in Equation 6.1, this probability is calculated as the total number of reactive blocks ($N_{reactive}$) divided by the total number of blocks (N_{total}).

$$P(R) = \frac{N_{reactive}}{N_{total}} \quad \text{(Equation 6. 17)}$$

Likelihood function

The likelihood function $P(L_i|R)$ incorporates the probability of the observed data given the conditional parameters. In the current case, the likelihood is expressed as the number of reactive blocks at a specific condition ($N_{L_i}^{reactive}$) divided by the total number of reactive blocks Equation 6.6. This approach provides a simple estimation of the likelihood based on the collected data.

$$P(L_i|R) = \frac{N_{L_i}^{reactive}}{N_{reactive}} \quad \text{(Equation 6. 18)}$$

Marginal likelihood and posterior probability

The marginal likelihood $P(L_i)$, representing the overall probability of observing a specific condition (L_i) is assessed as indicated in Equation 6.6, by the division of the total number of blocks at a specific condition ($N_{L_i}^{total}$) over the total number of blocks (N_{total}). The marginal likelihood normalizes the posterior distribution, ensuring that the total probability is equal to 1. By combining the prior and likelihood using Bayes' Theorem, the posterior probability (($P(R|L_i)$) is then computed, the posterior probability represents the probability of a block being reactive given the observed condition as per Equation 6.7.

$$P(L_i) = \frac{N_{L_i}^{total}}{N_{total}} \quad \text{(Equation 6. 19)}$$

$$P(R|L_i) = \frac{P(L_i|R) \cdot P(R)}{P(L_i)} = \frac{N_{L_i}^{reactive}}{N_{L_i}^{total}} \quad \text{(Equation 6. 20)}$$

Beta distribution and Confidence intervals

After observing the data, the parameters of a neutral and uniform Beta (i.e., Beta(1,1)) distribution are updated to model the posterior distribution [49]. Then, parameters α' and β', derived from the observed data are equivalent to the number of reactive and non-reactive blocks at a specific condition ($N_{L_i}^{reactive} + 1$; $N_{L_i}^{non-reactive} + 1$) as represented per

Equation 6.5 [50]. This posterior Beta distribution now models the probability of reactivity after observing the data. The probability density function (PDF) of the Beta distribution is represented in Equation 6.6, and the Beta function is defined in Equation 6.7.

$$Beta\ (\alpha', \beta') = Beta\ (N_{L_i}^{reactive} + 1, N_{L_i}^{non-reactive} + 1) \quad \text{(Equation 6. 21)}$$

$$f(t;\ \alpha, \beta) = \frac{1}{B(\alpha, \beta)} t^{\alpha-1}(1-t)^{\beta-1} \quad \text{(Equation 6. 22)}$$

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt \quad \text{(Equation 6. 23)}$$

Finally, the cumulative distribution function (CDF) of the Beta distribution gives the probability that a Beta-distributed random variable is less than or equal to a given value *x*. The inverse of CDF, known as the percentile point function (PPF) or the quantile function, allows for the computation of confidence intervals. For a 95% confidence interval, the values required are at 2.5th and 97.5% percentile of the distribution. Therefore, the lower and upper bounds are given as per Equation 6.8 and Equation 6.9, respectively.

$$CI_{lower} = Beta^{-1}(0.025; \alpha, \beta) \quad \text{(Equation 6. 24)}$$

$$CI_{upper} = Beta^{-1}(0.975; \alpha, \beta) \quad \text{(Equation 6. 25)}$$

In summary, this Bayesian approach, incorporating the Beta distribution allows for probabilistic modelling that relies on observed data and includes uncertainty quantification.

Model performance evaluation

Brier score measures how close the predicted probabilities are to the actual binary outcomes [51]. It ranges from 0 to 1, with 0 being the ideal score. The Brier score penalizes both the magnitude of prediction errors and how well-calibrated the model is. Therefore, considering

p_i as the predicted probability and y_i the actual binary outcome (i.e., 1 for field reactive and 0 for non-reactive), Equation 6.10 indicates the calculations for Brier score.

$$Brier\ Score = \frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2 \qquad \text{(Equation 6. 26)}$$

In this sense, when evaluating the reliability of laboratory tests in predicting field reactivity, two key parameters are considered: posterior probability and Brier score. For reactive cases, the most reliable test will show high posterior probability, indicating strong predictive power for reactive outcomes, and low Brier score, reflecting accurate and well-calibrated predictions. Conversely, for non-reactive cases, the most reliable test will display low posterior probability, suggesting the test correctly identifies non-reactive outcomes, while still maintaining low Brier score to ensure the accuracy and calibration of the predictions.

6.9 Results

6.9.1 Probability distribution for mixtures without SCMs

Probability per Reactivity Level

To evaluate the reactivity level of the test outcomes as established in ASTM C1778 [11], a comparison analysis has been conducted, as shown in Figure 6. 8. Outcomes from both tests, AMBT and CPT, were classified across different reactivity levels (R0, R1, R2, R3), where R0 is non-reactive, and R1 to R3 are reactive with increasing reactivity levels (i.e., moderately, highly and very highly reactive, respectively).

For the AMBT test without SCMs, the outcomes indicated varying probabilities that aggregates classified as reactive would indeed be reactive in the field. As the level of reactivity increased, the probability of being reactive in the field also increased. Moderate reliability was observed for R1, and high reliability for R2 and R3, with probabilities of 69%, 72%, and 84%, respectively. A similar pattern was observed for the CPT reactivity levels, where the probability of occurrence in the field increased with higher reactivity levels. The

R1, R2, and R3 levels showed probabilities of 59%, 72%, and 87%, respectively. This indicates moderate (R1), and high reliability (R2, R3).

It is expected that the probability of an aggregate classified as non-reactive is null or the lowest for being reactive in the field. However, both tests demonstrated limitations in accurately identifying non-reactive aggregates. The probability of an aggregate being reactive in the field, even if classified as non-reactive, was calculated as 41% and 61% for AMBT and CPT, respectively, indicating a significant rate of false negatives. Yet, the AMBT test had a lower false negative rate for non-reactive aggregates (R0) compared to the CPT test.

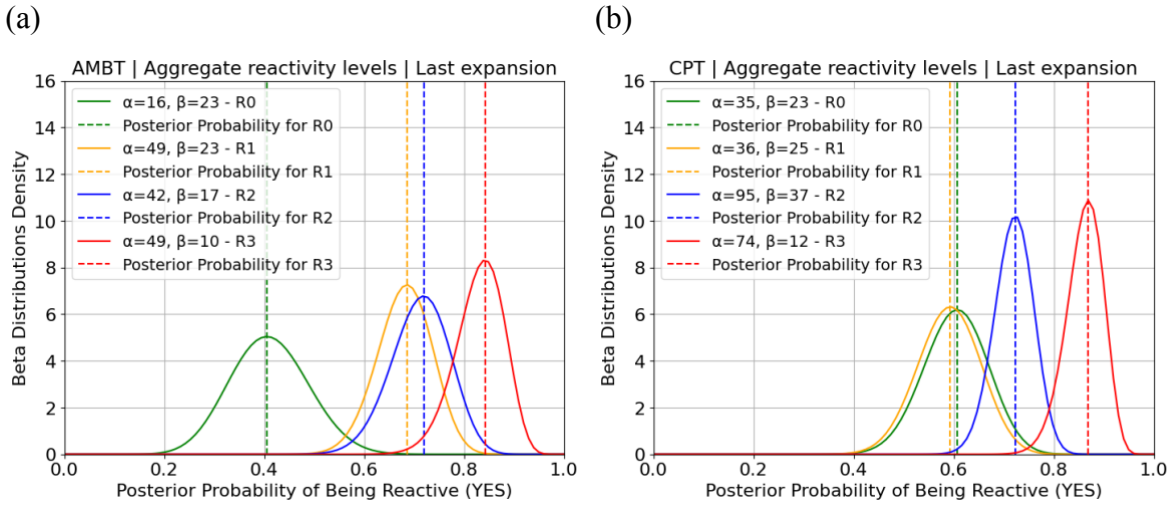


Figure 6. 8 – Posterior probability of a test outcome indicating reactivity in the field for different reactivity levels for cases without SCMs: (a) AMBT and (b) CPT.

Probability per Environment

Due to data availability, the comparison analysis has been conducted between two different environments: Ottawa, representing a colder scenario with an average temperature of 6.9°C and relative humidity of 75%, and Austin, representing a warmer environment with an average temperature of 21°C and relative humidity of 63%.

For the AMBT test, the probability that aggregates classified as reactive would indeed be reactive in the field was 59% for Ottawa and 78% for Austin, with moderate and high reliabilities, respectively (Figure 6. 9). This suggests that the AMBT test performs slightly

better in predicting reactivity in warmer environments. However, false positive rates for aggregates classified as non-reactive but reactive in the field were 20% for Ottawa and 44% for Austin, indicating very low and low performance in accurately identifying non-reactive aggregates. In the CPT test, probability rates for reactive predictions were observed as 59% for Ottawa and 77% for Austin. This indicates moderate and high reliability in predicting field reactivity for reactive aggregates in both environments, respectively. However, a notable difference was seen for non-reactive cases. The colder environment (Ottawa) showed a lower probability (i.e., 36%) compared to the warm environment (i.e., 66%), suggesting the CPT test is more likely to correctly classify non-reactive aggregates in colder environments compared to warmer ones.

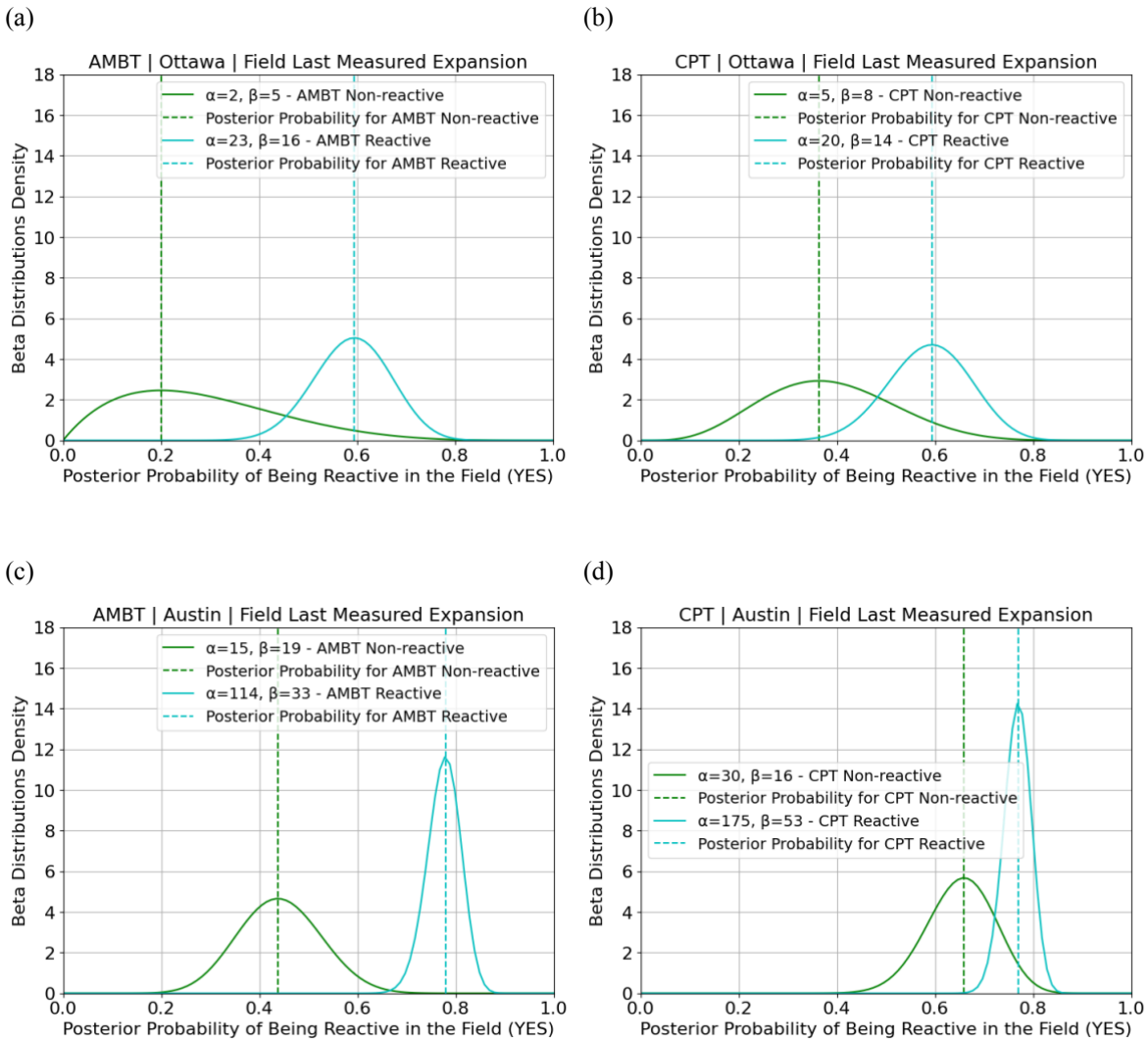


Figure 6. 9 – Posterior probability of a test outcome indicating reactivity in the field for different environments for cases without SCMs: (a) AMBT in Ottawa; (b) CPT in Ottawa; (c) AMBT in Austin; and, (d) CPT in Austin

Probability per Alkali loading

To evaluate the probability of occurrence based on concrete alkali content (Figure 6. 10), three main groups were selected for analysis based on data availability: low alkali content (less than 2.94 kg/m^3), moderate alkali content (2.94 kg/m^3 to 4.20 kg/m^3), and high alkali content (above 4.20 kg/m^3).

For AMBT without SCMs, results indicated that as alkali levels increase, the probability of reactivity occurrence in the field also increases. Reactive aggregates with low alkali content

showed a low likelihood of field reactivity (i.e., 38%), indicating a potential misinterpretation for true positive cases. Moderate and high alkali content mixes showed better performance, with reactive aggregates having a 63% and 84% chance of reacting in the field, respectively. Non-reactive aggregates had a 0%, 8% and 58% probability of reactivity for low, moderate and high alkali content, respectively indicating the test capacity to capture non-reactive cases in conditions of lower alkali content. The CPT test follows a similar pattern, with increasing alkali content resulting in higher probabilities of field reactivity: 65%, 57%, and 82% for low, medium, and high alkali content, respectively. However, the test thresholds became less reliable for non-reactive aggregates as alkali content increased (i.e., 18%, 53%, 82% for low, moderate and high alkali contents).

Both AMBT and CPT tests showed that higher alkali content (i.e., >5.25kg/m^3) resulted in higher probabilities of field reactivity. Besides CPT being slightly more reliable than AMBT in identifying reactive cases at lower alkali levels, both tests behave similarly at higher alkali levels. However, AMBT performs better than CPT for identifying non-reactive cases at low and moderate alkali levels, but both tests become unreliable at high alkali content due to a rising false positive rate.

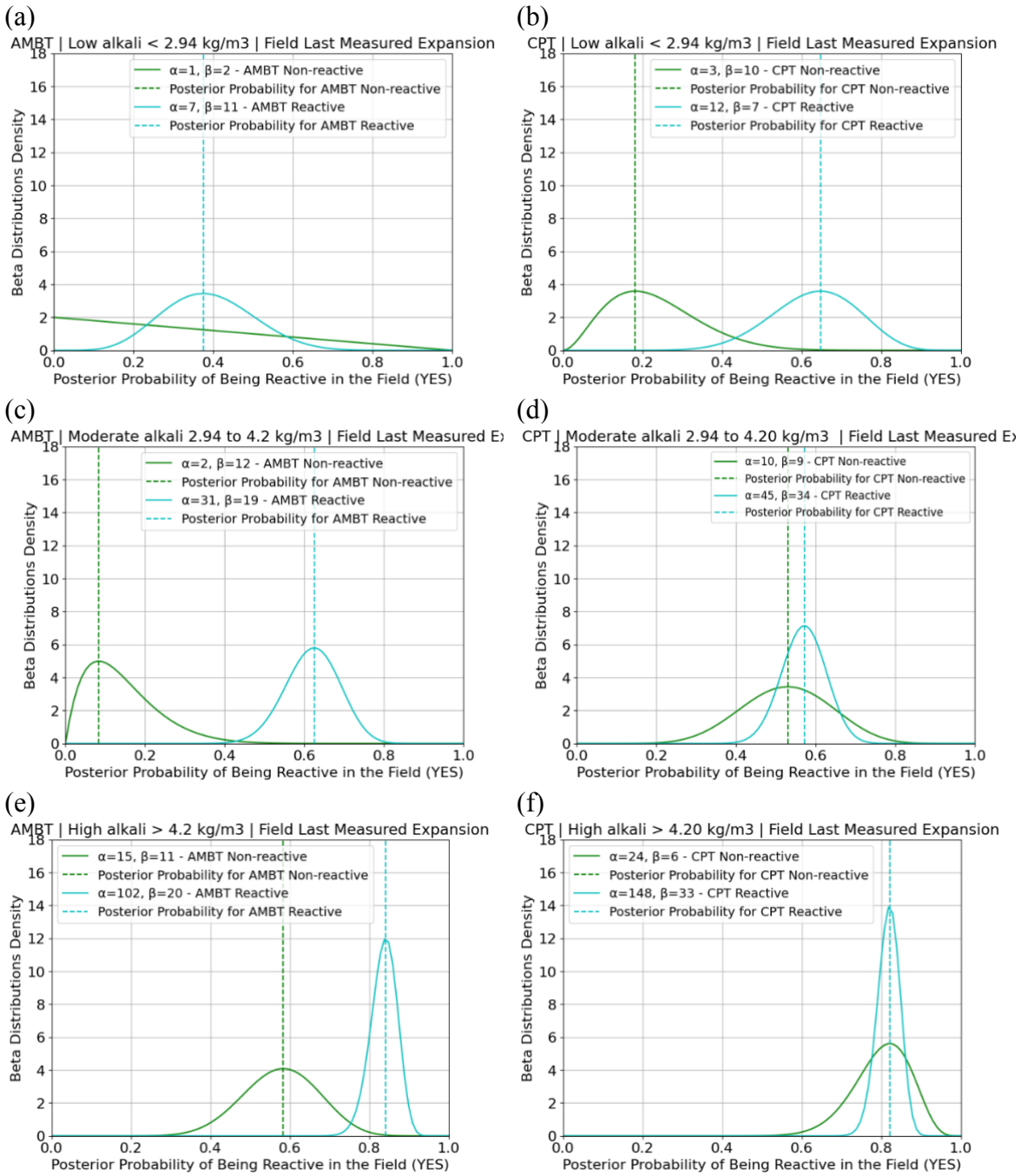


Figure 6. 10 – Posterior probability of a test outcome indicating reactivity in the field for different alkali content for cases without SCMs: (a) AMBT low alkali; (b) CPT low alkali; (c) AMBT medium alkali; (d) CPT medium alkali; (e) AMBT high alkali; and, (f) CPT high alkali.

6.9.2 Probability distribution for mixtures with SCMs

Probability per Alkali loading

To evaluate the probability of occurrence based on concrete alkali content in mixes containing SCMs (Figure 6. 11), three main groups were selected for analysis based on data availability: low alkali content (less than 2.94 kg/m^3), medium alkali content (2.94 kg/m^3 to 4.20 kg/m^3), and high alkali content (above 4.20 kg/m^3).

The AMBT test demonstrated high reliability for identifying non-reactive cases, with posterior probabilities of 10%, 27%, and 0% for low, moderate, and high alkali content, respectively. These low probabilities indicate that very few non-reactive cases were incorrectly identified as reactive, particularly at high alkali levels. However, for identifying reactive aggregates, the AMBT displayed moderate to low reliability. The posterior probabilities for reactive cases were 63% for low alkali content, 42% for moderate, and 50% for high, suggesting inconsistent performance as alkali levels varied.

In contrast, in the CPT test at low alkali content, the reliability was moderate for non-reactive aggregates, with a posterior probability of 18%, and higher for reactive aggregates, with a 65% probability. However, as alkali content increased, the reliability for non-reactive cases declined significantly, with 53% at moderate alkali levels and 82% at high alkali levels, indicating a substantial risk of misclassifying reactive cases as non-reactive. For reactive cases, the CPT's reliability improved with increasing alkali content, with posterior probabilities of 57% at moderate and 82% at high alkali levels, reflecting better performance for identifying reactive aggregates under higher alkali conditions.

In summary, both AMBT and CPT tests exhibit increasing reliability in identifying reactive aggregates with higher alkali content, though they differ in their ability to correctly identify non-reactive cases. AMBT shows strong performance at high alkali levels for non-reactive aggregates, with low probabilities of false positives, but its reliability in detecting reactive aggregates fluctuates across alkali contents. Conversely, CPT performs well in detecting reactive aggregates, particularly at higher alkali levels, but its reliability decreases for non-reactive aggregates, with increasing probabilities of false positives as alkali content rises. This inconsistency, especially at moderate alkali levels, highlights the need for further investigation into factors such as environmental conditions and the influence of

supplementary cementitious materials, which may contribute to the variations observed in the tests' outcomes.

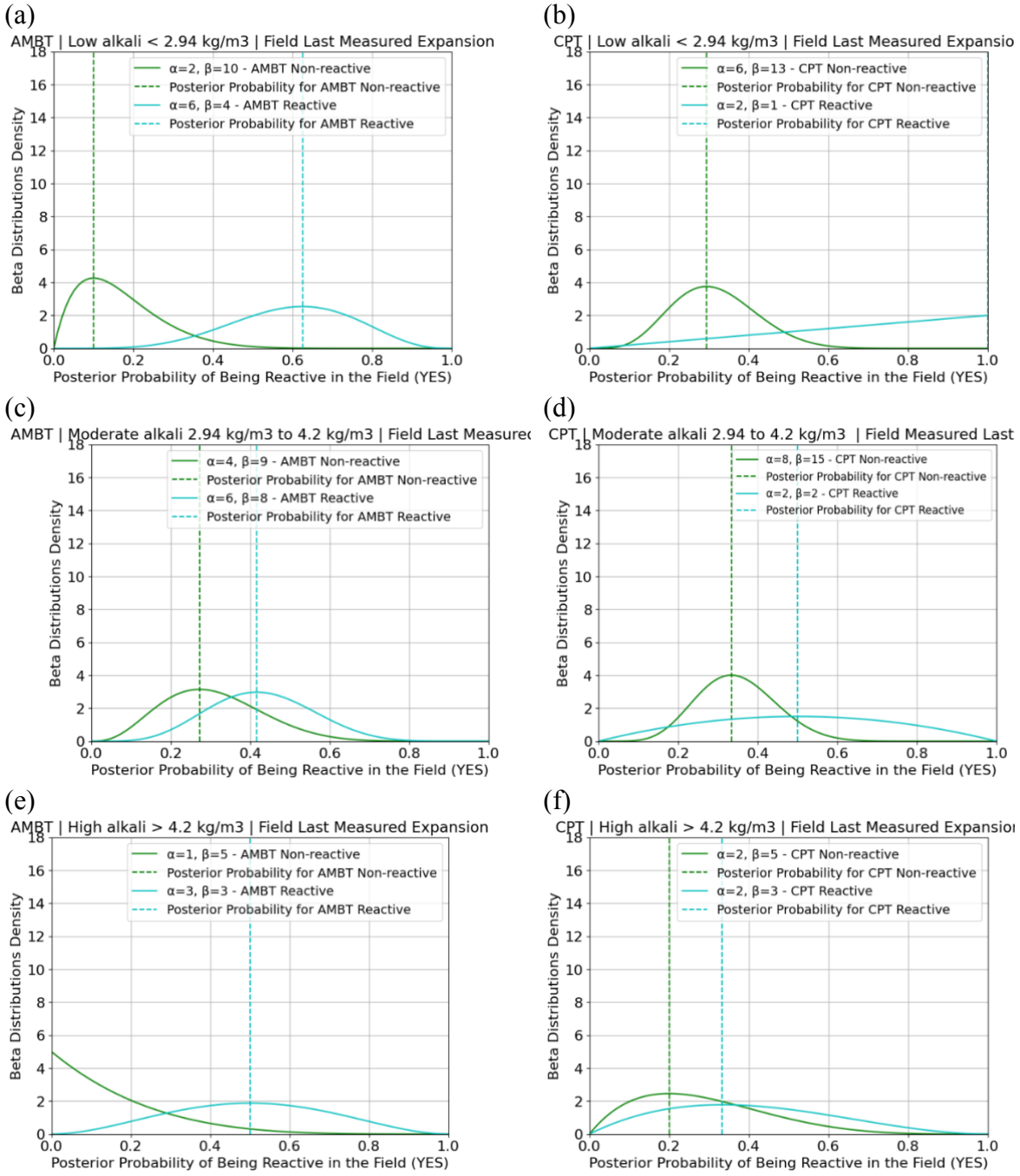


Figure 6. 11 – Posterior probability of a test outcome indicating reactivity in the field for different alkali content for cases with SCMs: (a) AMBT low alkali; (b) CPT low alkali; (c) AMBT medium alkali; (d) CPT medium alkali; (e) AMBT high alkali; and (f) CPT high alkali.

Probability per SCM type

To evaluate the probability of occurrence based on the SCM type, mixes containing Fly Ash and Slag were evaluated for non-reactive and reactive cases. For the AMBT test, results show that the presence of Fly Ash and Slag leads to the identification of non-reactive cases (i.e., 8% and 17%, respectively). While the test has a better performance in identifying reactive cases containing Slag (62.5%) when compared to Fly Ash (38%). Meanwhile, CPT indicates the probability of reactive cases for 25% and 100% for Fly Ash and Slag, respectively. These findings highlight the variability in test reliability depending on the type of SCM.

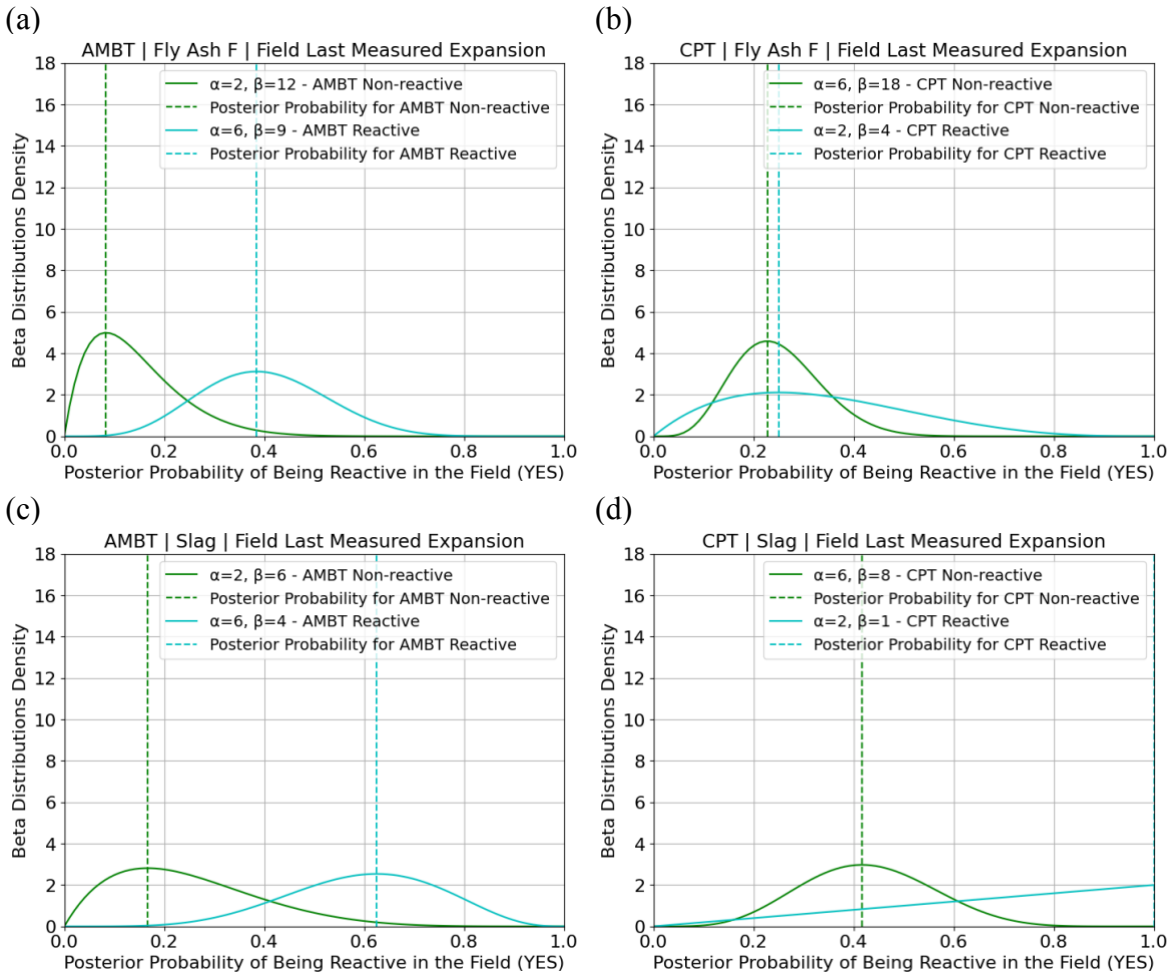


Figure 6. 12 – Posterior probability of a test outcome indicating reactivity in the field for different alkali content for cases with SCMs: (a) AMBT Fly Ash F; (b) CPT Fly Ash F; (c) AMBT Slag; and (d) CPT Slag.

Probability per Environment

Due to data availability, the comparison between two different environments is performed for mixes containing SCMs: Ottawa, representing a colder scenario with an average temperature of 6.9°C and relative humidity of 75%, and Austin, representing a warmer environment with an average temperature of 21°C and relative humidity of 63%.

For the AMBT test, the probability that a system with SCMs classified as reactive would be reactive in the field was 44% for Ottawa and 67% for Austin (Figure 6. 13). This suggests that the AMBT test performs better in predicting reactivity in warmer environments. Moreover, a lower rate for false positives is observed in the colder weather, with probabilities of 12% and 29% for Ottawa and Austin, respectively, in identifying non-reactive aggregates. In the CPT test, distinct probability rates for reactive predictions were observed, with 25% for Ottawa and 100% for Austin. This indicates high variations for the CPT test in predicting field reactivity for reactive cases in both environments in the presence of SCMs. However, the colder environment showed a lower probability (28%) compared to the warmer environment (38%), suggesting the CPT test is more likely to correctly classify non-reactive aggregates in colder environments compared to warmer ones.

Both AMBT and CPT tests show variability in their predictive reliability depending on the environment. While the AMBT test has slightly better performance in warmer environments, CPT has a similar performance in identifying reactive cases in both environments. Additionally, both tests showed better performance in colder environments when classifying systems with SCMs as non-reactive, with better performance for AMBT.

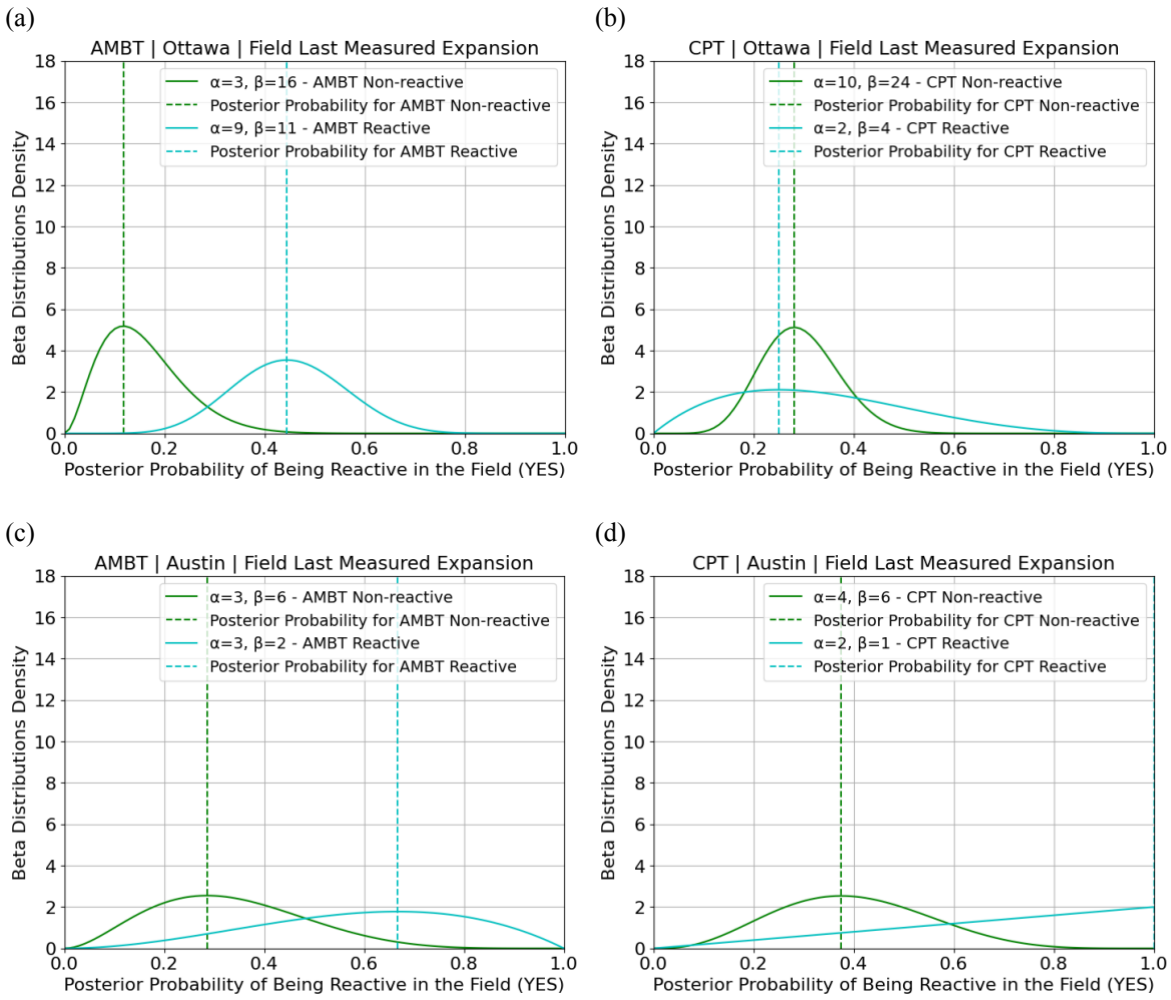


Figure 6. 13 – Posterior probability of a test outcome indicating reactivity in the field for different environments for cases with SCMs: (a) AMBT in Ottawa; (b) CPT in Ottawa; (c) AMBT in Austin; and, (d) CPT in Austin.

6.10 Discussion

Mixes without SCMs

The results from the probabilistic modelling and performance evaluation of AMBT and CPT tests reveal important insights into the factors influencing AAR in concrete. The key findings highlight the significance of reactivity levels, environmental conditions, and alkali loading in predicting field reactivity, aligning with the broader literature on AAR development [8,9,24,52].

The analysis shows that as the reactivity level increases, the probability of an aggregate being reactive in the field also rises for both AMBT and CPT tests. This trend, illustrated in Figure 6. 14a for mixes without SCMs, is consistent with previous research indicating that higher reactivity levels in laboratory tests correlate with increased field reactivity [19,53]. However, the broader confidence intervals for the AMBT test at lower reactivity levels suggest greater uncertainty compared to the CPT test, which exhibits a more consistent distribution across different reactivity levels.

Environmental conditions also play a crucial role in the reliability of these tests. Figure 6. 14b demonstrates that non-reactive aggregates are more reliably identified in colder environments, while warmer conditions result in higher probabilities of field occurrence, especially for CPT outcomes. This finding is supported by studies showing that temperature and humidity significantly impact AAR progression [45]. Warmer environments yield more consistent results, particularly for reactive aggregates, due to the narrower confidence intervals observed.

Alkali loading is another critical factor, with higher alkali content leading to increased probabilities of field reactivity for reactive aggregates, as shown in Figure 6. 14c. This relationship between alkali content and AAR development is well-documented in the literature [33,54]. The CPT test tends to show higher probabilities for non-reactive aggregates at elevated alkali levels, suggesting reduced consistency in identifying non-reactive cases. Conversely, the AMBT test, while showing broader confidence intervals for non-reactive aggregates, generally performs better at higher alkali content levels.

The integration of these findings with correlation analysis further underscores the importance of the alkali content, environment, and laboratory outcomes in predicting field reactivity. The positive correlation between reactivity and alkali content (0.25) supports the observed trend that higher alkali levels enhance field reactivity, aligning with existing studies [33]. Additionally, the correlation between field temperature and last expansion (0.30) reinforces the impact of warmer environments on AAR development, highlighting the need for temperature considerations in AAR assessments.

Moreover, the strong positive correlation between laboratory outcomes and the field's last measured expansion (0.45) indicates that laboratory test expansion largely contributes in

predicting field behaviour. This consistency is crucial for developing accurate predictive models, as evidenced by the higher reliability seen in the posterior probabilities for reactive aggregates in both AMBT and CPT tests, particularly at higher reactivity levels and alkali content.

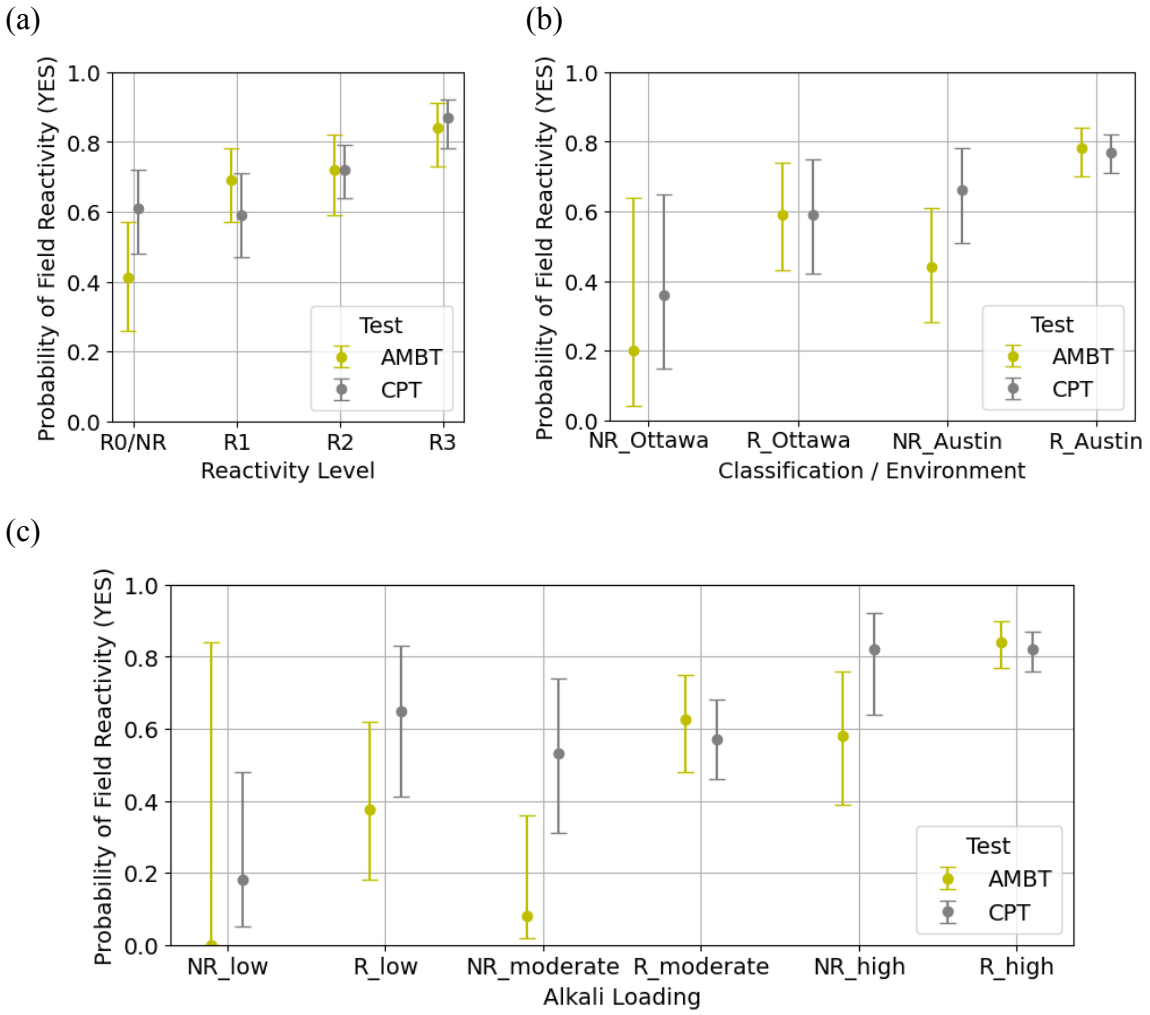


Figure 6. 14 – Summary of the posterior probabilities for cases without SCMs with a confidence interval of 95% for (a) reactivity level; (b) environment; and (c) alkali loading.

In summary, Figure 6. 15 presents the posterior probabilities for reactive and non-reactive cases without SCMs, indicating that the AMBT test generally shows lower probabilities of field reactivity compared to the CPT test. Despite similar posterior probabilities for identifying reactive cases (around 74% for AMBT for CPT), the AMBT shows lower

probabilities for non-reactive aggregates (i.e., 41%) compared with CPT (i.e., 61%), indicating better performance in the classification of non-reactive cases, and thus, avoiding false negative cases.

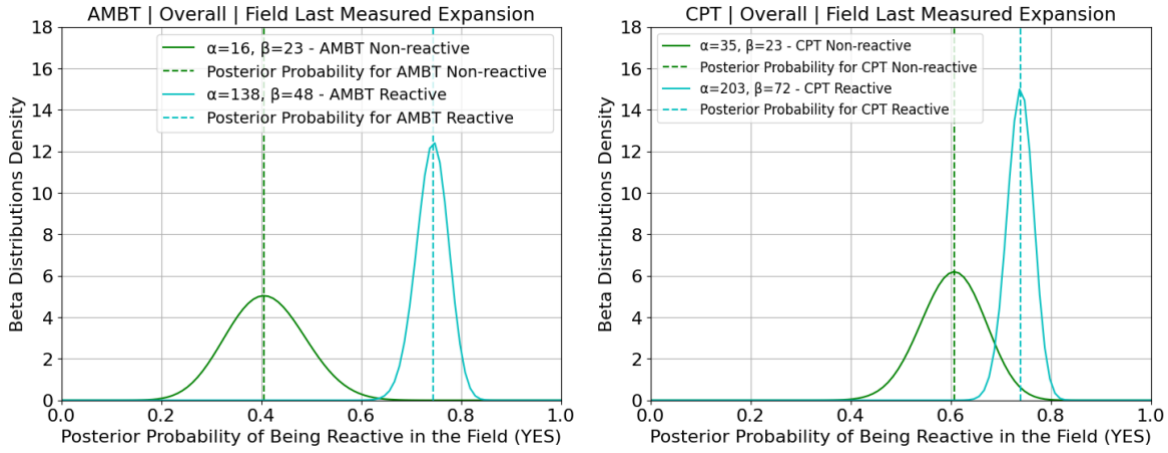


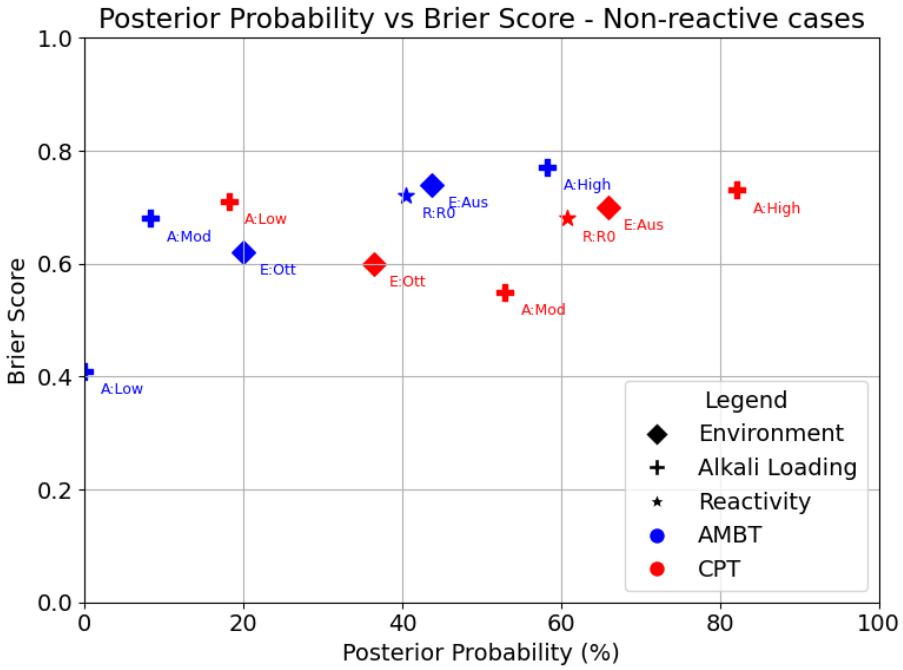
Figure 6. 15 – Summary of the posterior probabilities for cases without SCMs with a confidence interval of 95% for (a) AMBT and (b) CPT.

When analyzing the posterior probability and Brier score for both reactive and non-reactive mixtures without SCMs, Figure 6. 16 provides a summary of the results. For non-reactive cases, the optimal outcome is achieved by combining a low posterior probability with a low Brier score (i.e., lower left corner of Figure 6. 16a) to reflect the most reliable test. In this sense, AMBT tends to outperform in specific conditions, such as low and high alkali, as well as in warm and cold environments, enhancing slightly the test reliability. This is supported by the overall assessment, where AMBT shows a posterior probability of 41% and a Brier score of 0.71, compared to the CPT, which shows a posterior probability of 61% and a Brier score of 0.69.

Conversely, when evaluating the reactive cases, the optimal outcome is achieved by combining a high posterior probability with a low Brier score (i.e., lower right corner of Figure 6. 16b) to reflect the most reliable test. While there are specific conditions, such as moderate alkali levels and highly reactive aggregates where AMBT excels, and low alkali levels where CPT performs better, the overall reliability of both tests remains similar. This is further validated by the overall evaluation, with AMBT showing a posterior probability of

74% and a Brier score of 0.29, compared to CPT's posterior probability of 74% and a Brier score of 0.31.

(a)



(b)

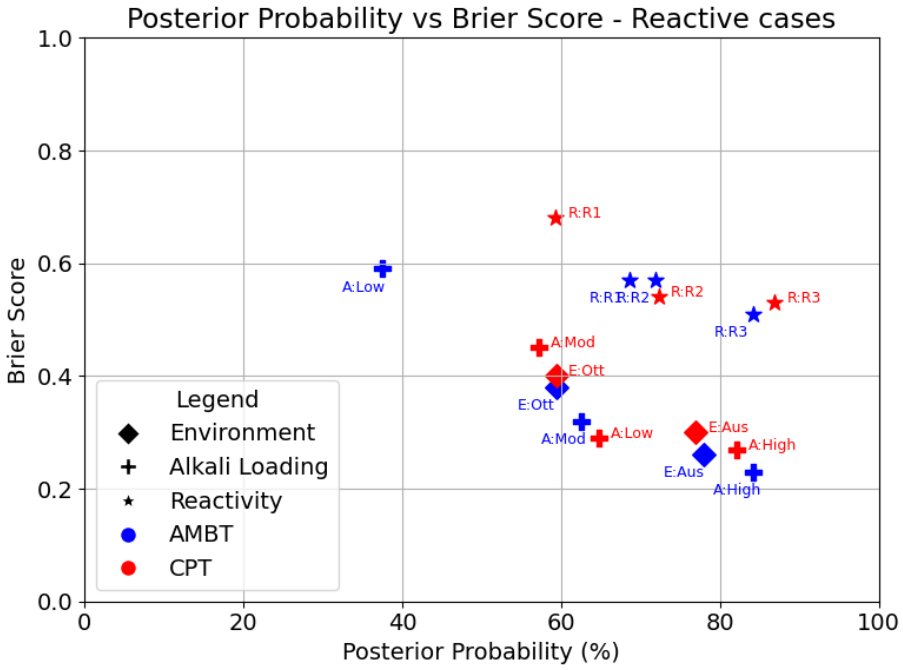


Figure 6. 16 – Posterior probabilities vs Brier score for cases without SCMs: (a) Reactive cases and (b) Non-reactive cases.

Mixes with SCMs

For mixes with SCMs, the trends remain similar. Figure 6. 17a-c illustrates that warmer environments and increased alkali content tend to contribute to higher probabilities of field reactivity. Notably, both AMBT and CPT tests show greater probabilities for mixes containing slag in comparison to those with fly ash (Figure 6. 17b). When compared to mixes without SCMs, mixes containing SCMs exhibit much broader confidence intervals, indicating increased variability in the test outcomes. This suggests that additional factors, such as the specific amount and type of SCMs used, may play a more significant role in influencing the results.

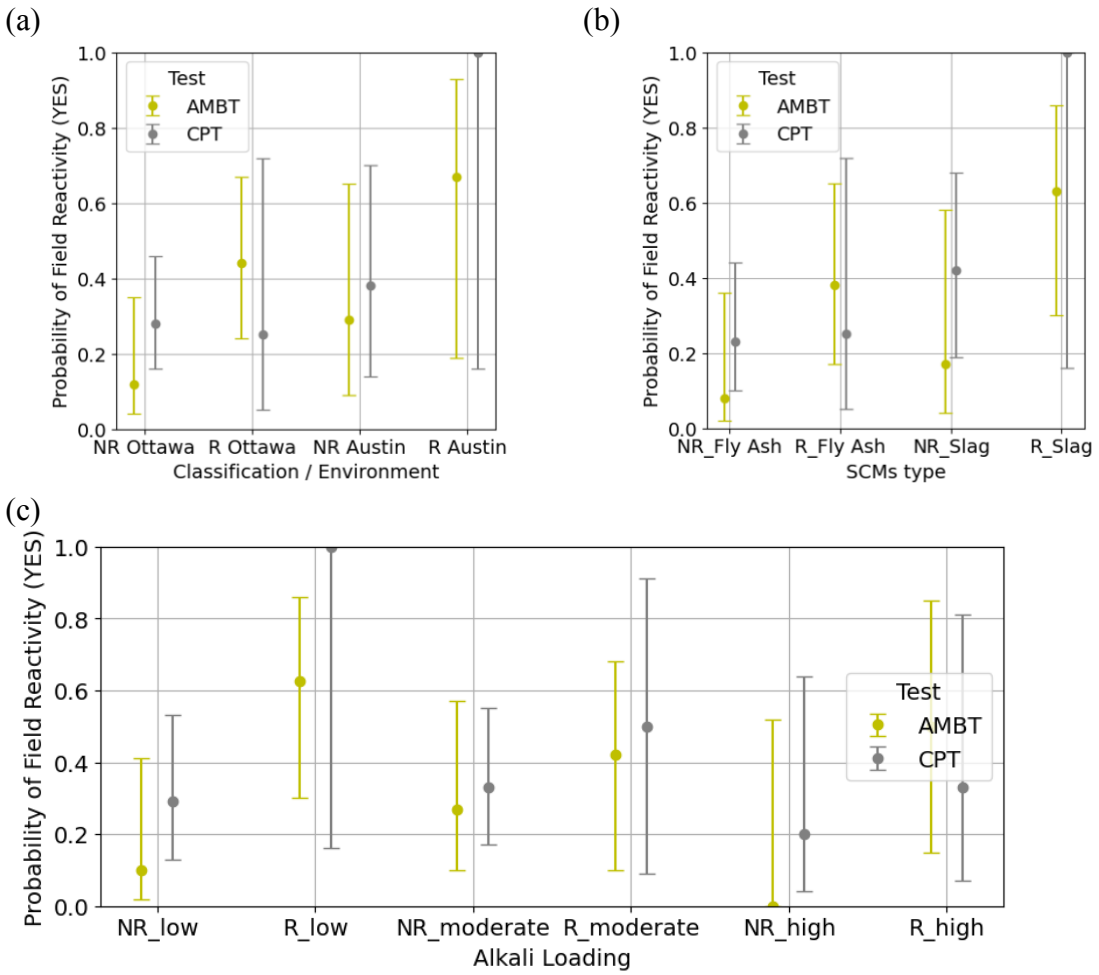


Figure 6. 17 – Summary of the posterior probabilities for cases with SCMs with a confidence interval of 95% for (a) Environment, (b) SCMs type, and (c) Alkali loading.

Figure 6. 18 provides a summary of the posterior probabilities for both reactive and non-reactive outcomes in tests involving SCMs. Despite the similar performance between AMBT and CPT with a posterior probability of 50%, the CPT test shows greater variability, as evidenced by its wider 95% confidence interval. For non-reactive cases, the AMBT test yields a lower posterior probability of 16% compared to 30% for CPT, suggesting that AMBT may offer more consistent performance in identifying non-reactive mixes, aligning with the observations for mixes without SCMs.

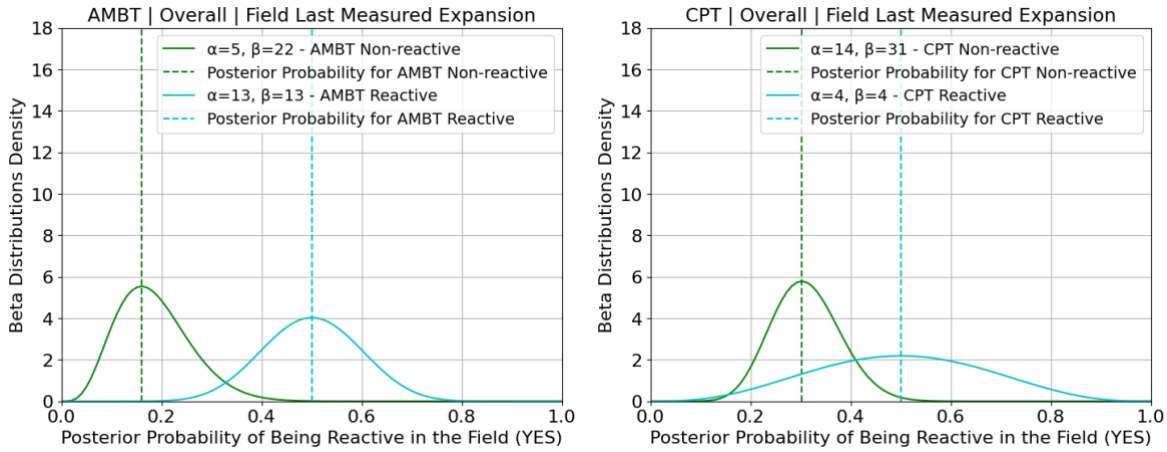


Figure 6. 18 – Summary of the posterior probabilities for cases with SCMs with a confidence interval of 95% for (a) AMBT and (b) CPT

When analyzing the posterior probability and Brier score for both reactive and non-reactive mixtures with SCMs, Figure 6. 19 provides a summary of the results. For non-reactive cases, the optimal outcome is achieved by combining a low posterior probability with a low Brier score (i.e., lower left corner of Figure 6. 19a) to reflect the most reliable test. Overall, AMBT tends to outperform with posterior probability for field reactivity of 16% and Brier 0.67 when compared to CPT with posterior of 30% and Brier score of 0.67.

Conversely, when evaluating the reactive cases, the optimal outcome is achieved by combining a high posterior probability with a low Brier score (i.e., lower right corner of Figure 6. 19b) to reflect the most reliable test. Yet, both tests show similar performance, which is further supported by the overall assessment, where CPT and AMBT demonstrate a posterior probability of 50% and a Brier score of 0.33. Therefore, if the goal is to avoid false

positive cases, meaning being classified as non-reactive in the laboratory but demonstrating reactivity in the field, AMBT becomes slightly more reliable than CPT for cases with SCMs.

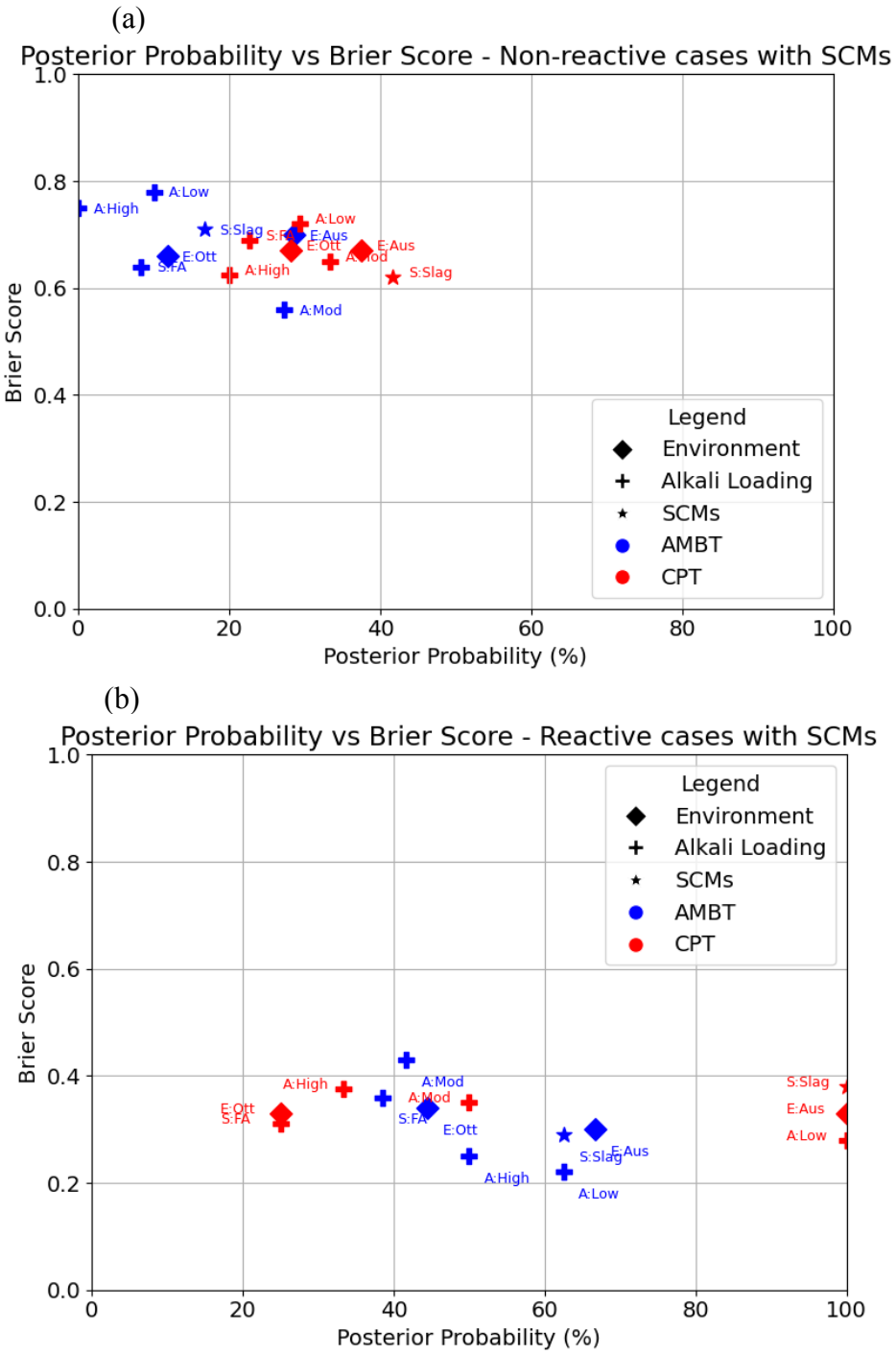


Figure 6. 19 – Posterior probabilities vs Brier score for cases with SCMs: (a) Reactive cases and (b) Non-reactive cases.

In summary, Table 6. 1 and Table 6. 2 provide an overview of the risks associated with accepting test outcomes under distinct environmental conditions (i.e., cold vs warm) and concrete alkali content levels (i.e., low, moderate, and high). The risks were determined based on the probability of correctly classifying reactive outcomes and minimizing false negatives for non-reactive outcomes. For non-reactive outcomes, risk levels were classified as very low (≤20%), low (20-40%), moderate (40-60%), and high (>60%) based on the probability of misclassification. For reactive outcomes, very low risk corresponds to a high probability of correct classification (>80%), low risk (70-80%), moderate risk (60-70%), and high risk (<60%), reflecting how reliably each test can identify reactivity in the field. As expected, tests performed on concrete that is exposed to cold environments displaying low alkali levels and SCMs tend to reduce risks, while exposure to warm environments with high alkali content and no SCMs increases the chances of misclassification, particularly for non-reactive outcomes.

Table 6. 1 – Summary of risks associated with test outcomes for mixes without SCMs.

Test	**Cases**	**Environment**		**Alkali content**		
		Cold	Warm	Low	Moderate	High
AMBT	Non-reactive	+	+++	+	+	+++
	Reactive	++++	++	++++	+++	++
CPT	Non-reactive	++	++++	+	+++	++++
	Reactive	++++	++	+++	++++	++

+: very low risk; ++: low risk; moderate risk: +++; high risk: ++++

Table 6. 2 – Summary of risks associated with test outcomes for mixes with SCMs.

Test	**Cases**	**Environment**		**Alkali content**		
		Cold	Warm	Low	Moderate	High
AMBT	Non-reactive	+	+	+	++	+
	Reactive	++++	+++	++++	+++	+++
CPT	Non-reactive	++	+++	++	++	++
	Reactive	++++	+*	++++	+*	+*

+: very low risk; ++: low risk; moderate risk: +++; high risk: ++++; *limited data available

In conclusion, the comprehensive analysis of the probabilistic distribution has provided valuable insights into AAR accelerated laboratory reliability in addressing discrepancies observed with field performance. The analysis highlights the need to consider multiple factors, including alkali content, environmental conditions, and laboratory outcomes, to

accurately predict the likelihood of field reactivity. Furthermore, the analysis was conducted based on current thresholds—0.04% at 365 and 730 days for mixes without and with SCMs in CPT, and 0.1% at 14 and 28 days for mixes without and with SCMs in AMBT. A further review of these threshold values and test durations could enhance the reliability of testing procedures.

6.11 Conclusion

This study provides a comprehensive framework for assessing the risk of AAR in new concrete structures by integrating laboratory and field data with statistical and probabilistic modelling. The main outcomes are:

- The probabilistic modelling demonstrates that as reactivity levels increase, the probability of field reactivity also increases for both AMBT and CPT test outcomes. This trend aligns with existing literature, highlighting the critical role of aggregate reactivity levels in predicting field behavior.
- Environmental conditions significantly impact the reliability of AAR predictions. Warmer environments generally yield higher probabilities of field reactivity, emphasizing the impact of temperature and relative humidity in AAR risk assessments.
- Higher alkali content in concrete mixtures corresponds to increased probabilities of field reactivity for reactive aggregates. This relationship underscores the importance of controlling alkali levels in concrete mix designs to mitigate AAR risks effectively.
- For mixes without SCMs, AMBT generally shows better performance in identifying non-reactive cases, with a posterior probability of 41% and a Brier score of 0.71, compared to CPT's posterior probability of 61% and a Brier score of 0.69, while both tests demonstrate similar reliability for reactive cases, with AMBT showing a posterior probability of 74% and a Brier score of 0.29, and CPT displaying 74% and 0.31 respectively.
- For mixtures containing SCMs, AMBT generally shows better performance in identifying non-reactive cases, with a posterior probability of 16% and a Brier score of 0.67, compared to CPT's posterior probability of 30% and a Brier score of 0.67, while both tests

demonstrate similar reliability for reactive cases, with AMBT and CPT showing a posterior probability of 50% and a Brier score of 0.33 respectively.

- Risks associated with test outcomes vary by environmental conditions and alkali content. Cold environments with low alkali and SCMs reduce misclassification risks, while warm environments with high alkali and no SCMs increase risks, especially for non-reactive cases.

6.12 References

[1] M.D.A. Thomas, K.J. Folliard, J.H. Ideker, North America (USA and Canada), in: Alkali-Aggreg. React. Concr. World Rev., CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[2] J. Lindgård, B. Grelk, B.J. Wigum, J. Trägårdh, K. Appelqvist, E. Hol, M. Ferreira, M. Leivo, Nordic Europe, in: Alkali-Aggreg. React. Concr. World Rev., CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[3] E.M.R. Fairbairn, South and Central America, in: Alkali-Aggreg. React. Concr. World Rev., CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[4] K. Yamada, T. Miyagawa, Japan, China and South-East Asia, in: Alkali-Aggreg. React. Concr. World Rev., CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[5] A. Shayan, S. Freitag, Australia and New Zealand, in: Alkali-Aggreg. React. Concr. World Rev., CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[6] A.K. Mullick, Indian Sub-Continent, in: Alkali-Aggreg. React. Concr. World Rev., CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[7] T. Kay, A.B. Poole, I. Sims, Middle East & North Africa, in: Alkali-Aggreg. React. Concr. World Rev., CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[8] B. Fournier, M. Berube, Alkali-aggregate reaction in concrete: a review of basic concepts and engineering implications, Can. J. Civ. Eng. 27 (2000) 167–191. https://doi.org/10.1139/l99-072.

[9] P.J. Nixon, I. Sims, eds., RILEM Recommendations for the Prevention of Damage by Alkali-Aggregate Reactions in New Concrete Structures, Springer Netherlands, Dordrecht, 2016. https://doi.org/10.1007/978-94-017-7252-5.

[10] I. Sims, P. Nixon, RILEM Recommended Test Method AAR-0: Detection of Alkali-Reactivity Potential in Concrete—Outline guide to the use of RILEM methods in assessments of aggregates for potential alkali-reactivity, Mater. Struct. 36 (2003) 472–479. https://doi.org/10.1007/BF02481527.

[11] ASTM C1778, Guide for Reducing the Risk of Deleterious Alkali-Aggregate Reaction in Concrete, ASTM International, 2022. https://doi.org/10.1520/C1778-22.

[12] M. Thomas, A. Dunster, P. Nixon, B. Blackwell, Effect of fly ash on the expansion of concrete due to alkali-silica reaction – Exposure site studies, Cem. Concr. Compos. 33 (2011) 359–367. https://doi.org/10.1016/j.cemconcomp.2010.11.006.

[13] B. Fournier, A. Bilodeau, N. Bouzoubaa, P.-C. Nkinamubanzi, Field and Laboratory Investigations on the Use of Fly Ash and Li-Based Admixtures to Prevent ASR in Concrete, in: United Kingdom, 2018.

[14] I. Borchers, J. Lindgård, C. Müller, Evaluation of laboratory test methods for assessing the alkali-reactivity potential of aggregates by field site tests, Mater. Constr. 72 (2022) e286. https://doi.org/10.3989/mc.2022.17221.

[15] J. Custodio, D. Costa, A.B. Ribeiro, A.S. Silva, Assessment of potential alkali-silica reactivity of aggregates for concrete, 4TH Int. Conf. Struct. Integr. ICSI 2021 37 (2022) 590–597. https://doi.org/10.1016/j.prostr.2022.01.127.

[16] J. Lindgård, P.J. Nixon, I. Borchers, B. Schouenborg, B.J. Wigum, M. Haugen, U. Åkesson, The EU "PARTNER" Project — European standard tests to prevent alkali reactions in aggregates: Final results and recommendations, Cem. Concr. Res. 40 (2010) 611–635. https://doi.org/10.1016/j.cemconres.2009.09.004.

[17] T. Drimalas, K.J. Folliard, J.H. Ideker, Findings from the University of Texas at Austin ASR Exposure Site after 20 Years, in: L.F.M. Sanchez, C. Trottier (Eds.), Proc. 17th Int. Conf. Alkali-Aggreg. React. Concr., Springer Nature Switzerland, Cham, 2024: pp. 516–523. https://doi.org/10.1007/978-3-031-59349-9_59.

[18] S. Hayman, M. Thomas, N. Beaman, P. Gilks, Selection of an effective ASR-prevention strategy for use with a highly reactive aggregate for the reconstruction of concrete structures at Mactaquac generating station, Cem. Concr. Res. 40 (2010) 605–610. https://doi.org/10.1016/j.cemconres.2009.08.015.

[19] B. Fournier, R. Chevrier, A. Bilodeau, P.-C. Nkinamubanzi, N. Bouzoubaa, COMPARATIVE FIELD AND LABORATORY INVESTIGATIONS ON THE USE OF SUPPLEMENTARY CEMENTING MATERIALS (SCMs) TO CONTROL ALKALI-SILICA REACTION (ASR) IN CONCRETE, in: Brazil, 2016.

[20] C.A. MacDonald, Kingston Outdoor Exposure Site for ASR –29 Year Update, (2020).

[21] G.E. Blight, M.G. Alexander, Alkali-Aggregate Reaction and Structural Damage to Concrete, 2011.

[22] B. Fournier, M.-A. Bérubé, Alkali–aggregate reaction in concrete: a review of basic concepts and engineering implications, 27 (2000).

[23] A. Leemann, M. Góra, B. Lothenbach, M. Heuberger, Alkali silica reaction in concrete - Revealing the expansion mechanism by surface force measurements, Cem. Concr. Res. 176 (2024) 107392. https://doi.org/10.1016/j.cemconres.2023.107392.

[24] M.D.A. Thomas, B. Fournier, K.J. Folliard, Alkali-Aggregate Reactivity (AAR) Facts Book, Book (2013) 224.

[25] P.E. Grattan-Bellew, T. Katayama, So-Called Alkali-Carbonate Reaction (ACR), in: Alkali-Aggreg. React. Concr., CRC Press, 2017: p. 26. https://doi.org/10.1201/9781315708959.

[26] J. Lindgard, M.D.A. Thomas, E.J. Sellevold, B. Pedersen, O. Andic-Cakir, H. Justnes, T.F. Ronning, Alkali-silica reaction (ASR)-performance testing: Influence of specimen pre-treatment, exposure conditions and prism size on alkali leaching and prism expansion, Cem. Concr. Res. 53 (2013) 68–90. https://doi.org/10.1016/j.cemconres.2013.05.017.

[27] ASTM C1260, Test Method for Potential Alkali Reactivity of Aggregates (Mortar-Bar Method), ASTM International, 2022. https://doi.org/10.1520/C1260-22.

[28] R.E. Oberholster, G. Davies, An accelerated method for testing the potential alkali reactivity of siliceous aggregates, Cem. Concr. Res. 16 (1986) 181–189. https://doi.org/10.1016/0008-8846(86)90134-1.

[29] J. Lindgård, Ö. Andiç-Çakır, I. Fernandes, T.F. Rønning, M.D.A. Thomas, Alkali–silica reactions (ASR): Literature review on parameters influencing laboratory performance testing, Cem. Concr. Res. 42 (2012) 223–243. https://doi.org/10.1016/j.cemconres.2011.10.004.

[30] S. Multon, M. Cyr, A. Sellier, P. Diederich, L. Petit, Effects of aggregate size and alkali content on ASR expansion, Cem. Concr. Res. 40 (2010) 508–516. https://doi.org/10.1016/j.cemconres.2009.08.002.

[31] ASTM C1293, Test Method for Determination of Length Change of Concrete Due to Alkali-Silica Reaction, (2024). https://doi.org/10.1520/C1293_C1293M-23A.

[32] J. Lindgård, T. Østnor, B. Fournier, Ø. Lindgård, T. Danner, G. Plusquellec, K. De Weerdt, Determining alkali leaching during accelerated ASR performance testing and in field exposed cubes using cold water extraction (CWE) and μXRF, MATEC Web Conf. 199 (2018) 03004. https://doi.org/10.1051/matecconf/201819903004.

[33] B. Fournier, J. Lindgård, B.J. Wigum, I. Borchers, Outdoor exposure site testing for preventing Alkali-Aggregate Reactivity in concrete – a review., MATEC Web Conf. 199 (2018) 03002. https://doi.org/10.1051/matecconf/201819903002.

[34] A.H.-S. Ang, W.H. Tang, A.H.-S. Ang, Probability concepts in engineering: emphasis on applications in civil & environmental engineering, 2nd ed, Wiley, New York, 2007.

[35] K.M. Ramachandran, C.P. Tsokos, Descriptive Statistics, in: Math. Stat. Appl. R, Elsevier, 2015: pp. 1–52. https://doi.org/10.1016/B978-0-12-417113-8.00001-1.

[36] O.C. Ibe, Introduction to Descriptive Statistics, in: Fundam. Appl. Probab. Random Process., Elsevier, 2014: pp. 253–274. https://doi.org/10.1016/B978-0-12-800852-2.00008-0.

[37] B.R. Martin, Statistics, Experiments, and Data, in: Stat. Phys. Sci., Elsevier, 2012: pp. 1–20. https://doi.org/10.1016/B978-0-12-387760-4.00001-9.

[38] O.C. Ibe, Introduction to Inferential Statistics, in: Fundam. Appl. Probab. Random Process., Elsevier, 2014: pp. 275–305. https://doi.org/10.1016/B978-0-12-800852-2.00009-2.

[39] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, Bayesian Data Analysis Third edition (with errors fixed as of 13 February 2020), (n.d.).

[40] LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S, Philos. Trans. R. Soc. Lond. 53 (1763) 370–418. https://doi.org/10.1098/rstl.1763.0053.

[41] T. Drimalas, K.J. Folliard, A. Parashar, M.D.A. Thomas, A. Ghanizadeh, A.M. Hossack, B. Fournier, National Cooperative Highway Research Program, Transportation Research Board, National Academies of Sciences, Engineering, and Medicine, Alkali-Silica Reactivity Potential and Mitigation: Test Methods and State of Practice, Transportation Research Board, Washington, D.C., 2023. https://doi.org/10.17226/27435.

[42] J. Lindgård, T.F. Rønning, M.D.A. Thomas, B. Fournier, A.S. Silva, ASR - PERFORMANCE TESTING: MAIN FINDINGS IN THE NORWEGIAN COIN PROJECT, in: Sao Paulo – Brazil, 2016.

[43] B. Fournier, A. Ferro, V. Sivasundaram, CANMET/Industry Research Consortium on Alkali-Aggregate Reactivity in Concrete, EPRI and CANMET, Ontario, Canada, 2001. https://www.epri.com/research/products/1004031.

[44] J. Custódio, J. Lindgård, B. Fournier, A. Santos Silva, M.D.A. Thomas, T. Drimalas, J.H. Ideker, R.-P. Martin, I. Borchers, B. Johannes Wigum, T.F. Rønning, Correlating field and laboratory investigations for preventing ASR in concrete – The LNEC cube study (Part I – Project plan and laboratory results), Constr. Build. Mater. 343 (2022) 128131. https://doi.org/10.1016/j.conbuildmat.2022.128131.

[45] B. Fournier, J.H. Ideker, K.J. Folliard, M.D.A. Thomas, P.-C. Nkinamubanzi, R. Chevrier, Effect of environmental conditions on expansion in concrete due to alkali–silica reaction (ASR), Mater. Charact. 60 (2009) 669–679. https://doi.org/10.1016/j.matchar.2008.12.018.

[46] I. Fernandes, A. Leemann, B. Fournier, E. Menendez, J. Lindgård, I. Borchers, J. Custódio, PARTNER project post-documentation study. Condition assessment of field exposure site cubes. Results of microstructural analyses, Cem. Concr. Res. 162 (2022) 107006. https://doi.org/10.1016/j.cemconres.2022.107006.

[47] K.J. Folliard, R. Barborak, T. Drimalas, L. Du, S. Garber, J. Ideker, T. Ley, S. Williams, M. Juenger, B. Fournier, M.D.A. Thomas, Preventing ASR/DEF in New Concrete: Final Report (0-4085-5), Texas Department of Transportation and the Federal Highway Administration, 2006.

[48] B.J. Wigum, L.T. Pedersen, B. Grelk, J. Lindgård, State-of-the art report: Key parameters influencing the alkali aggregate reaction, SINTEF Building and Infrastructure, 2006. https://www.sintef.no/globalassets/upload/byggforsk/partner/report-2.1-final-a06018.pdf.

[49] S. Sinharay, Continuous Probability Distributions, (n.d.).

[50] M. Shaw, Use of Bayes' theorem and the beta distribution for reliability estimation purposes, Reliab. Eng. Syst. Saf. 31 (1991) 145–153. https://doi.org/10.1016/0951-8320(91)90115-N.

[51] D.A. Redelmeier, D.A. Bloch, D.H. Hickam, Assessing predictive accuracy: How to compare brier scores, J. Clin. Epidemiol. 44 (1991) 1141–1146. https://doi.org/10.1016/0895-4356(91)90146-Z.

[52] J. Lindgard, E.J. Sellevold, M.D.A. Thomas, B. Pedersen, H. Justnes, T.F. Ronning, Alkali-silica reaction (ASR)-performance testing: Influence of specimen pre-treatment, exposure conditions and prism size on concrete porosity, moisture state and transport properties, Cem. Concr. Res. 53 (2013) 145–167. https://doi.org/10.1016/j.cemconres.2013.05.020.

[53] T.N. Nguyen, L.F.M. Sanchez, J. Li, B. Fournier, V. Sirivivatnanon, Correlating alkali-silica reaction (ASR) induced expansion from short-term laboratory testings to long-term field performance: A semi-empirical model, Cem. Concr. Compos. 134 (2022) 104817. https://doi.org/10.1016/j.cemconcomp.2022.104817.

[54] D. Hooton, B. Fournier, Long-term alkali–silica mitigation of high-alkali concrete with cement replacements, Proc. Inst. Civ. Eng. - Constr. Mater. 175 (2022) 125–136. https://doi.org/10.1680/jcoma.21.00049.

[55] K. Yeturu, Machine learning algorithms, applications, and practices in data science, in: Handb. Stat., Elsevier, 2020: pp. 81–206. https://doi.org/10.1016/bs.host.2020.01.002.

[56] H. Belyadi, A. Haghighat, Supervised learning, in: Mach. Learn. Guide Oil Gas Using Python, Elsevier, 2021: pp. 169–295. https://doi.org/10.1016/B978-0-12-821929-4.00004-4.

[57] R. Bleszynski, R.D. Hooton, M.D.A. Thomas, C. Rogers, Durability of Ternary Blend Concrete with Silica Fume and Blast-Furnace Slag: Laboratory and Outdoor Exposure Site Studies, ACI Mater. J. 99 (2002). https://doi.org/10.14359/12329.

6.13 Appendix A - Description of the main variables used for the model development

Variable	Description
FIELD_cement_alkali_content(kg/m3)	Cement alkali content in field samples (kg/m3)
FIELD_Tested_aggregate	Type of aggregate tested in field samples
FIELD_Aggregate_type	Aggregate type (coarse/fine)
FIELD_Last_expansion%	Last recorded expansion in field samples (%)
FIELD_TEMPERATURE	Annual average temperature at field site (°C)
FIELD_RH	Annual average relative humidity at field site (%)
LAB_outcome_expansion%	Expansion outcome from laboratory tests (%)
Field_Reactive?	Reactivity status of field samples (Yes/No)
LAB_TEST_NAME	Name of laboratory test conducted

6.14 Appendix B - Descriptive statistics for numerical variables without SCMs

	Field Cement Alkali Content (kg/m^3)	Field Last Measured Expansion (%)	Field Temperature (C)	Field RH (%)	Laboratory outcome expansion (%)	Field Reactive Binary (>0.04%)
mean	4.68	0.36	18.84	65.27	65.27	0.69
std	0.93	0.41	4.89	5.10	5.10	0.46
min	1.67	-0.02	5.60	63.00	63.00	0.00
25%	3.99	0.02	21.00	63.00	63.00	0.00
50%	5.25	0.20	21.00	63.00	63.00	1.00
75%	5.25	0.53	21.00	63.00	63.00	1.00
max	5.50	1.36	21.00	87.28	87.28	1.00

6.15 Appendix C - Descriptive analysis of the aggregates evaluated without SCMs

Tested aggregate	Aggregate type	Frequency	Proportion
Jobe - El Paso, Texas	Fine	135	15.7
Placitas - Albuquerque, New Mexico	Coarse	111	12.9
Wright - Robstown, Texas	Fine / Coarse	96 (84/12)	11.1
Spratt, ON	Coarse	87	10.1
Sudburry, ON	Coarse	62	7.2
Ottersbo, Norway	Coarse	54	6.3
Maryland - North East, MD	Coarse	30	3.5
New Ulm, MN	Coarse	29	3.4
Pioneer - Amarillo, Texas	Fine	22	2.6
Capitol - Austin, Texas	Fine	20	2.3
DDS - Cleveland, Texas	Fine	19	2.2
Daleville - Moscow, PA	Coarse	18	2.1
North Garden, VA	Coarse	18	2.1
Helotes, Texas	Coarse	16	1.9
Stone Cold - Victoria, Texas	Coarse	16	1.9
LG Everist - Dell Rapids, SD	Coarse	16	1.9
Cheyenne, WY	Coarse	15	1.7
Eagle Lake, Texas	Coarse	14	1.6
Fordyce - Mission, Texas	Fine	12	1.4
Hylas - Rockville, VA	Coarse	12	1.4
Richard Spur - Elgin, OK	Coarse	12	1.4
Hanson Little-River - Ashtown, AR	Coarse	10	1.2
Nova Scotia	Coarse	6	0.7
Alberta	Coarse	6	0.7
Quebec	Coarse	6	0.7
New Brunswick	Coarse	5	0.6
El Indio, Texas	Coarse	5	0.6
Stone Cold - Victoria, Texas + Wright - Robstow...	Fine + Coarse	4	0.5
San Antonio, Texas (NR)	Fine / Coarse / Fine + Coarse	3 (1/1/1)	0.3
Omaha, NE	Fine	2	0.2

6.16 Appendix D - Descriptive statistics for numerical variables with SCMs

	Field Cement Alkali Content (kg/m^3)	Field Last Measured Expansion (%)	Field Temperature (C)	Field RH (%)	Laboratory outcome expansion (%)	SCMs amount (%)	Field Reactive Binary (>0.04%)
mean	3.36	0.06	9.49	73.14	0.08	28.73	0.23
std	1.01	0.14	5.39	4.89	0.10	17.01	0.42
min	0.79	-0.01	6.90	63.00	-0.01	7.50	0.00
25%	2.73	0.01	6.90	75.00	0.02	12.50	0.00
50%	3.41	0.02	6.90	75.00	0.03	30.00	0.00
75%	4.16	0.03	6.90	75.00	0.12	40.00	0.00
max	5.47	0.60	21.00	80.50	0.47	76.00	1.00

6.17 Appendix E - Descriptive analysis of the aggregates evaluated with SCMs

Tested aggregate	Aggregate type	Frequency	Proportion
Spratt, ON	Coarse	36	22.9%
Sudburry, ON	Coarse	35	22.3%
Wright - Robstown, Texas	Fine	28	17.8%
New Brunswick	Coarse	22	14%
Alberta	Coarse	18	11.5%
Nova Scotia	Coarse	10	6.4%
Placitas - Albuquerque, New Mexico	Coarse	4	2.5%
Quebec	Coarse	4	2.5%

6.18 Appendix F - Descriptive analysis of the SCMs evaluated in mixes containing SCMs

SCM	Frequency	Proportion
Fly ash F	62	39.5%
Silica Fume	40	25.5%
Slag	27	17.2%
Fly ash C	11	7.0%
Slag + Silica Fume	5	3.2%
Metakaolin	3	1.9%
Ultra fine fly ash	3	1.9%
Fly ash C + Silica Fume	3	1.9%
Fly ash C + Ultra fine fly ash	3	0.6%

6.19 Appendix G - Probability modelling outcomes for AMBT Test results without SCMs

AMBT Results

	Reactivity Level	Reactive (YES)	Non-Reactive (NO)	Total	Prior Probability	Likelihood	Marginal Likelihood	Posterior Probability	CI Lower	CI Upper	Brier
Overall	NR	15	22	37	68.78	9.87	16.74	40.54	26.31	56.61	0.72
	R	137	47	184	68.78	90.13	83.26	74.46	67.69	80.21	0.28
Reactivity Level	R0	15	22	37	68.78	9.87	16.74	40.54	26.31	56.61	0.72
	R1	48	22	70	68.78	31.58	31.67	68.57	56.92	78.24	0.57
	R2	41	16	57	68.78	26.97	25.79	71.93	59.10	81.91	0.57
	R3	48	9	57	68.78	31.58	25.79	84.21	72.58	91.41	0.51
Environment	NR_Ottawa	1	4	5	54.76	4.35	11.90	20.00	4.33	64.12	0.62
	R_Ottawa	22	15	37	54.76	95.65	88.10	59.46	43.39	73.69	0.38
	NR_Austin	14	18	32	71.75	11.02	18.08	43.75	28.11	60.78	0.74
	R_Austin	113	32	145	71.75	88.98	81.92	77.93	70.49	83.90	0.26
Alkali loading	NR_low	0	1	1	35.29	0.00	5.88	0.00	1.26	84.19	0.41
	R_low	6	10	16	35.29	100.00	94.12	37.50	18.44	61.67	0.59
	NR_medium	1	11	12	51.67	3.23	20.00	8.33	1.92	36.03	0.68
	R_medium	30	18	48	51.67	96.77	80.00	62.50	48.29	74.80	0.32
	NR_high	14	10	24	79.86	12.17	16.67	58.33	38.67	75.60	0.77
	R_high	101	19	120	79.86	87.83	83.33	84.17	76.57	89.60	0.23

6.20 Appendix H - Probability modelling outcomes for CPT Test results without SCMs

CPT Results

	Reactivity Level	Reactive (YES)	Non-Reactive (NO)	Total	Prior Probability	Likelihood	Marginal Likelihood	Posterior Probability	CI Lower	CI Upper	Brier
Overall	NR	34	22	56	71.73	14.41	17.02	60.71	47.57	72.44	0.68
	R	202	71	273	71.73	85.59	82.98	73.99	68.47	78.83	0.32
Reactivity Level	R0	34	22	56	71.73	14.41	17.02	60.71	47.57	72.44	0.68
	R1	35	24	59	71.73	14.83	17.93	59.32	46.54	70.93	0.68
	R2	94	36	130	71.73	39.83	39.51	72.31	64.04	79.27	0.54
	R3	73	11	84	71.73	30.93	25.53	86.90	78.02	92.49	0.53
Environment	NR_Ottawa	4	7	11	53.49	17.39	25.58	36.36	15.17	65.11	0.60
	R_Ottawa	19	13	32	53.49	82.61	74.42	59.38	42.14	74.52	0.40
	NR_Austin	29	15	44	75.19	14.29	16.30	65.91	51.05	78.13	0.70
	R_Austin	174	52	226	75.19	85.71	83.70	76.99	71.07	81.99	0.30
Alkali loading	NR_low	2	9	11	46.43	15.38	39.29	18.18	5.49	48.41	0.71
	R_low	11	6	17	46.43	84.62	60.71	64.71	40.99	82.70	0.29
	NR_medium	9	8	17	56.38	16.98	18.09	52.94	30.76	73.98	0.55
	R_medium	44	33	77	56.38	83.02	81.91	57.14	45.98	67.61	0.45
	NR_high	23	5	28	82.13	13.53	13.53	82.14	64.23	92.01	0.73
	R_high	147	32	179	82.13	86.47	86.47	82.12	75.84	87.03	0.27

6.21 Appendix I - Probability modelling outcomes for AMBT Test results with SCMs

AMBT Results

	Reactivity Level	Reactive (YES)	Non-Reactive (NO)	Total	Prior Probability	Likelihood	Marginal Likelihood	Posterior Probability	CI Lower	CI Upper	Brier
Overall	NR	4.00	21.00	25.00	32.65	25.00	51.02	16.00	6.55	34.87	0.67
	R	12.00	12.00	24.00	32.65	75.00	48.98	50.00	31.31	68.69	0.33
Environment	NR_Ottawa	1.00	11.00	12.00	24.00	16.67	48.00	8.33	1.92	36.03	0.64
	R_Ottawa	5.00	8.00	13.00	24.00	83.33	52.00	38.46	17.66	64.86	0.36
	NR_Austin	1.00	5.00	6.00	42.86	16.67	42.86	16.67	3.67	57.87	0.71
	R_Austin	5.00	3.00	8.00	42.86	83.33	57.14	62.50	29.93	86.30	0.29
Alkali loading	NR_low	2.00	15.00	17.00	28.57	20.00	48.57	11.76	3.58	34.71	0.66
	R_low	8.00	10.00	18.00	28.57	80.00	51.43	44.44	24.45	66.50	0.34
	NR_medium	2.00	5.00	7.00	40.00	50.00	70.00	28.57	8.52	65.09	0.70
	R_medium	2.00	1.00	3.00	40.00	50.00	30.00	66.67	19.41	93.24	0.30
	NR_high	1.00	9.00	10.00	33.33	16.67	55.56	10.00	2.28	41.28	0.78
	R_high	5.00	3.00	8.00	33.33	83.33	44.44	62.50	29.93	86.30	0.22
SCMs	NR_Flyash	3.00	8.00	11.00	34.78	37.50	47.83	27.27	9.92	57.19	0.56
	R_Flyash	5.00	7.00	12.00	34.78	62.50	52.17	41.67	19.22	68.42	0.43
	NR_Slag	0.00	4.00	4.00	25.00	0.00	50.00	0.00	0.51	52.18	0.75
	R_Slag	2.00	2.00	4.00	25.00	100.00	50.00	50.00	14.66	85.34	0.25

6.22 Appendix J - Probability modelling outcomes for CPT Test results with SCMs

CPT Results

	Reactivity Level	Reactive (YES)	Non-Reactive (NO)	Total	Prior Probability	Likelihood	Marginal Likelihood	Posterior Probability	CI Lower	CI Upper	Brier
Overall	NR	13.00	30.00	43.00	32.65	81.25	87.76	30.23	18.61	45.20	0.67
	R	3.00	3.00	6.00	32.65	18.75	12.24	50.00	18.41	81.59	0.33
Environment	NR_Ottawa	5.00	17.00	22.00	23.08	83.33	84.62	22.73	10.23	43.70	0.69
	R_Ottawa	1.00	3.00	4.00	23.08	16.67	15.38	25.00	5.27	71.64	0.31
	NR_Austin	5.00	7.00	12.00	46.15	83.33	92.31	41.67	19.22	68.42	0.62
	R_Austin	1.00	0.00	1.00	46.15	16.67	7.69	100.00	15.81	98.74	0.38
Alkali loading	NR_low	9.00	23.00	32.00	27.78	90.00	88.89	28.13	15.59	45.52	0.67
	R_low	1.00	3.00	4.00	27.78	10.00	11.11	25.00	5.27	71.64	0.33
	NR_medium	3.00	5.00	8.00	44.44	75.00	88.89	37.50	13.70	70.07	0.67
	R_medium	1.00	0.00	1.00	44.44	25.00	11.11	100.00	15.81	98.74	0.33
	NR_high	5.00	12.00	17.00	33.33	83.33	94.44	29.41	13.34	53.48	0.72
	R_high	1.00	0.00	1.00	33.33	16.67	5.56	100.00	15.81	98.74	0.28
SCMs	NR_Flyash	7.00	14.00	21.00	34.78	87.50	91.30	33.33	17.20	54.87	0.65
	R_Flyash	1.00	1.00	2.00	34.78	12.50	8.70	50.00	9.43	90.57	0.35
	NR_Slag	1.00	4.00	5.00	25.00	50.00	62.50	20.00	4.33	64.12	0.63
	R_Slag	1.00	2.00	3.00	25.00	50.00	37.50	33.33	6.76	80.59	0.38

Chapter 7: Enhancing alkali-aggregate reaction (AAR) predictive models using machine learning to integrate laboratory and field data

Ana Bergmann and Leandro Sanchez

Abstract:

Alkali-aggregate reaction (AAR) is a major durability issue that compromises concrete structures worldwide, making prevention the most effective strategy, as remediation after initiation remains difficult and uncertain. Standard test methods, such as the Accelerated Mortar Bar Test (AMBT) and the Concrete Prism Test (CPT), while widely used to assess the potential reactivity of aggregates and preventative measures against AAR, have shown limited accuracy. This indicates that the thresholds established by these tests may not fully capture the complexities of real-world environments. To address this gap, this study applies machine learning (ML) techniques, including decision tree classifiers and logistic regression, to appraise and adjust the thresholds of AMBT and CPT. The proposed flexible coupled threshold-time (FCTT) approach evaluates thresholds probabilistically and dynamically, integrating expansion levels, test duration, environmental factors, and alkali content to improve AAR occurrence prediction. This approach enhances the reliability of AAR risk assessments, thereby improving the long-term durability of concrete structures.

Keywords: Alkali-Aggregate Reaction, Machine Learning, Accelerated Mortar Bar Test, Concrete Prism Test, Threshold Calibration, Field Conditions

7.1 Introduction

Alkali-aggregate reaction (AAR) represents a significant challenge to the durability of concrete structures, causing significant distress and premature deterioration worldwide. This internal swelling reaction, predominantly manifesting as alkali-silica reaction (ASR), leads to expansive gels forming that induce cracking and compromise structural integrity [1,2]. Various laboratory tests have been developed to assess and mitigate the risks associated with AAR, with the Accelerated Mortar Bar Test (AMBT) and Concrete Prism Test (CPT) being

the most commonly used [3–5]. However, a critical challenge remains, since the thresholds established by these tests often fail to accurately predict long-term field performance [2,6].

Thresholds in AMBT and CPT are intended to classify aggregates and concrete mixes based on their reactivity and potential for causing AAR. However, these thresholds are often set based on accelerated conditions that do not fully capture the complexities of real-world environments [5,7,8]. Factors such as alkali concentration, moisture availability, and temperature can vary significantly between laboratory and field conditions, leading to discrepancies in the predictive accuracy of these tests. For instance, alkali leaching, which is controlled in laboratory settings, can be a significant factor in the field, affecting the long-term expansion of concrete [5,9]. Additionally, the recycling of alkalis during tests and the variable effects of temperature on alkali mobility and concentration further complicate the determination of accurate thresholds [8].

These inconsistencies suggest that the current thresholds used in AMBT and CPT may be insufficient in predicting AAR field development, depending on the specific environmental conditions. As a result, there is a need to re-evaluate these thresholds to better reflect the long-term performance of concrete given specificities such as environmental contexts, mix design (i.e., alkali content), and aggregate mineralogy. This process requires a comprehensive understanding of the factors influencing AAR and their interactions under field conditions, which is often beyond the scope of traditional testing methods [10].

Recent advancements in machine learning (ML) offer a powerful tool for addressing these challenges. By integrating vast amounts of laboratory and field performance data, ML models can identify patterns and relationships that are not apparent through conventional analysis. Therefore, this study applies ML techniques to propose a flexible coupled threshold-time (FCTT) approach for AMBT and CPT outcomes, ensuring they align more closely with real-world performance over extended periods. By developing models that account for the complex interactions between environmental factors and concrete composition, this research aims to enhance the reliability of AAR risk assessments. Ultimately, the FCCT approach will establish a more robust, risk-based framework for preventing AAR-induced damage in varied exposure conditions of concrete structures, enhancing their durability and safety.

7.2 Background

7.2.1 Laboratory tests and field performance to appraise AAR potential

Once initiated, addressing AAR is both complex and uncertain, making prevention the most effective strategy to avoid such degradation. Therefore, the need to accurately predict and mitigate AAR has driven the development of various laboratory testing methods over the decades. Among these, the AMBT and CPT have emerged as the most widely utilized methods for assessing the reactivity potential of aggregates [4,9]. The AMBT is performed by immersing mortar bars (i.e., 25x25x285 mm^3) in a high-alkali solution (1M NaOH) at 80°C, with expansions measured over 14 to 28 days for mixtures without and with SCMs, respectively [11,12]. This accelerated method yields rapid results and is often recommended for preliminary screenings of aggregate reactivity. Aggregates are considered reactive in the AMBT if expansion surpasses the 0.1% threshold [13]. In contrast, the CPT employs concrete prisms (i.e., 75x75x250 mm^3) and monitors expansion over a longer duration, typically 1 to 2 years for mixtures without and with SCMs, under conditions of moderate temperature (i.e., 38°C) and high relative humidity (i.e., 95%) [14]. The reactivity threshold in CPT is generally set at 0.04% expansion [13].

Despite their widespread use, these tests have notable limitations that can lead to significant discrepancies between laboratory predictions and actual field performance. The AMBT, while offering quick results, often yields conservative or exaggerated reactivity classifications. This method can result in false positives, due to its harsh testing conditions, using mortar instead of concrete, and crushing process of the aggregates which can expose unstable minerals and exaggerate expansion results [5,15–18]. On the other hand, the CPT, while considered more accurate due to its longer duration and use of concrete rather than mortar, is still susceptible to alkali leaching and other factors that may not be representative of actual service conditions, potentially leading to underestimation of reactivity [3].

Additionally, the thresholds established for both AMBT and CPT, which classify aggregates and concrete mixes based on their expansion rates, have been established through a combination of empirical studies and consensus, which may not fully capture the complexities of real-world environments. Variables such as moisture availability,

environmental temperature fluctuations, and the presence of external alkali sources can significantly alter the long-term behavior of concrete structures in the field [9,19]. For instance, alkali leaching during CPT can reduce the alkali content within the specimen, leading to lower expansions than might be observed under field conditions where leaching is less controlled [10]. Similarly, the high temperature in AMBT can accelerate reactions that would normally occur over much longer periods, potentially skewing the results [15].

Field studies further highlight these discrepancies, revealing notable differences between laboratory predictions and actual field performance. In Ontario (MTO) sites, for instance, the AMBT correlated better with field data than the CPT, but both tests underpredicted expansions in some cases [20]. Similarly, in Iceland, certain mixtures exhibited higher expansion in outdoor conditions than in CPT, particularly with reactive gravel and boosted alkali levels, suggesting that the accelerated testing conditions may not fully capture field behavior [21]. At the UK's BRE sites, field-exposed blocks showed more expansion than the CPT predicted, while AMBT provided a better estimate for the required fly ash dosage, though it struggled to account for the influence of cement alkalis [22]. The CANMET study in Canada highlighted that CPT reliably predicted SCM performance for short-term exposure but underpredicted long-term field behavior, primarily due to alkali leaching and environmental factors [4,23]. The PARTNER project in Europe further demonstrated that RILEM methods were effective in identifying the reactivity of typical aggregates but less reliable for slow-reacting types, with ASR progressing more rapidly in warmer climates [24]. Similarly, field testing in Texas revealed that both AMBT and CPT often underestimated the amount of SCMs required in high-alkali conditions, emphasizing the need for more comprehensive testing protocols that consider environmental variability [25].

Given these challenges and the availability of long-term exposed concrete blocks, there is a growing recognition of the need to re-evaluate the test thresholds to better reflect actual field performance. Traditional methods often fail to account for the nuanced interactions between environmental conditions, aggregate mineralogy, and mix design that influence AAR development. Consequently, the predictive accuracy of these tests in real-world scenarios remains limited.

7.2.2 Analytical methodologies

Uncertainty is an inherent aspect of engineering, especially when predicting material behavior, arising from both the inherent randomness of natural processes (i.e., aleatory uncertainty) and limitations in knowledge or predictive models (i.e., epistemic uncertainty) [26]. These uncertainties are managed through a combination of statistical, probabilistic, and recently increasingly artificial intelligence (AI) driven approaches. Statistical methods remain fundamental in quantifying the likelihood of various outcomes and assessing associated risks, helping engineers make more informed, data-driven decisions.

In summary, statistics can be understood as the science of data, providing tools to collect, summarize, organize, analyze, and model complex datasets [27,28]. These methods are essential in research for identifying patterns, testing hypotheses, and making predictions based on empirical data. Broadly, statistical methods can be categorized into two general branches, including descriptive statistics, which focuses on summarizing and organizing data, and inferential statistics, which extends conclusions beyond the immediate dataset, using techniques such as hypothesis testing, regression analysis and probability theory [28,29].

Machine learning (ML) is based on a multidisciplinary concept, combining artificial intelligence, probability and statistics, computational complexity theory, control theory, information theory, philosophy, psychology, neurobiology, and so on [30]. Therefore, when integrated with ML and AI, statistical methods become even more powerful. By leveraging AI-driven tools such as decision trees (DTs) and logistic regression, uncertainties can be managed more effectively, enhancing the predictive accuracy of models. Decision trees, for instance, are supervised learning models that split data based on key variables, allowing for the development of decision rules to predict outcomes [31]. They are particularly useful for tasks where clear thresholds or decision points need to be identified, such as predicting alkali-aggregate reaction (AAR) behavior. On the other hand, logistic regression is a classification machine learning model used to predict the probability of a binary outcome, such as the likelihood of field reactivity [31,32]. Logistic regression provides a probability-based estimate for an outcome, enabling the modelling of complex relationships between variables while accounting for uncertainty. Both DTs and logistic regression offer flexible,

interpretable frameworks for analyzing engineering data, helping to enhance the decision-making process where accurate predictions are critical.

In this context, ML offers a promising approach to enhance the predictive capabilities of AAR assessments. By integrating extensive laboratory and field data, ML models can analyze complex, non-linear relationships that traditional statistical methods might overlook. Therefore, this study aims to leverage ML techniques to develop a flexible coupled threshold-time (FCTT) approach for interpreting AMBT and CPT results, allowing for a more representative and dynamic reflection of long-term field performance. By strengthening the reliability of AAR prediction, this approach aims to establish a more robust, risk-based framework for preventing AAR-induced damage in varied exposure conditions of concrete structures, enhancing their durability and safety.

7.3 Scope of the work

This study aims to enhance the prediction of AAR occurrence in the field by integrating machine learning techniques with traditional laboratory test outcomes, specifically the AMBT and CPT. The research relies on a compiled comprehensive database of laboratory and field data, focusing on key variables such as expansion measurements over time, alkali content, SCMs presence and content, and environmental conditions.

A detailed analysis is conducted to evaluate the reliability of laboratory test outcomes in predicting field reactivity, with a particular focus on the standard expansion thresholds of 0.1% for AMBT and 0.04% for CPT. The study highlights the variability of these thresholds and identifies the conditions the tests are capable of predicting long-term field performance. Yet, to improve these predictions, a machine learning model, particularly decision trees (DT), is employed to optimize critical thresholds and decision rules, including testing age. Recognizing the need for a dynamic approach, logistic regression is also applied to develop a flexible coupled threshold-time (FCTT) method for interpreting AMBT and CPT results. This framework provides a more representative approach for predicting long-term field performance across varied conditions, thereby enhancing the durability and safety of concrete structures.

7.4 Methodology

The study focuses on developing a flexible coupled threshold-time (FCTT) approach for laboratory test outcomes to accurately predict field reactivity. By computing data and assessing probabilities of occurrence, it allows for a risk-based framework for preventing AAR-induced damage in varied exposure conditions of concrete structures. To achieve this, the methodology includes data preparation, performance metric analysis, decision tree modelling, and logistic regression.

7.4.1 Data Preparation

The dataset for this study was sourced from an established database [33,34], which contains comprehensive laboratory test results and corresponding field reactivity data. These data were derived from published sources, including journal articles, conference papers, and technical reports, spanning from 2001 to 2024, with field exposure results dating back to the early 1990s [15,23–25,35–42]. To ensure consistency and minimize errors, only datasets that provided detailed information linking laboratory and field conditions were included. It is important to note that concrete members exposed in the field are referred to as "blocks."

The inclusion criteria focused on datasets with thorough descriptions of testing protocols and outcomes. Only results following standardized laboratory methods were selected, concentrating on the AMBT and CPT, as these are the most widely used worldwide. As a result, the dataset contains detailed information on variables such as cement alkali content, aggregate type, the presence and quantity of supplementary cementitious materials (SCMs), and expansion measurements over time. For field-exposed concrete blocks, additional data on environmental factors, including location, temperature, and humidity, were also collected. Figure 7. 1 illustrates the information of laboratory and field data included in this dataset.

Laboratory	Field
• Test condition: Temperature, solution concentration, and relative humidity. • Sample details: Geometry, dimensions, volume, and surface area. • Aggregate details: Type (i.e., fine or coarse) and source (i.e., country or region of origin). • SCMs: Type and amount. • Expansion measurements: Laboratory outcomes (i.e., expansion at a given time), last measured expansion, age of last measured expansion and expansion over time.	• Sample details: Geometry, dimension, volume, surface area, and reinforcement. • Aggregate: Type and origin. • Supplementary cementitious materials (SCMs): Type and amount. • Alkali content: Amount. • Admixtures: Type and amount. • Environment conditions: Annual average temperature and relative humidity. • Expansion measurements: Last measured expansion, age of last measured expansion and expansion over time.

Figure 7. 1 – Database information for laboratory and field entries.

An example entry illustrating the detailed data collected is provided below:

- Laboratory: Aggregate tested: Spratt, ON, Aggregate type: Coarse, SCMs: None, SCMs amount: 0%, Sample shape: prism, Sample surface area (m^2): 0.10125, Sample volume (m^3): 0.0016875, Test name: CPT, Test temperature (^oC): 38, Solution concentration (Mol): N/A, External humidity (%): 95, Laboratory outcome expansion (%): 0.184, Laboratory last measured expansion (%): 0.225, Laboratory age of last recorded expansion (days): 2190, Expansion over time.
- Field: Mix-design name: M25-SP+, Aggregate tested: Spratt, ON, Aggregate type: Coarse, Cement alkali content (kg/m^3): 5.225, SCMs: None, SCMs amount: 0%, Admixture: Air entrained: Sample shape: beam, Sample surface area (m^2): 1.44, Sample volume (m^3): 0.112, Reinforcement: none, Environment: Ottawa, Field annual average temperature (^oC): 6.9, Field annual average relative humidity (%): 75, Field last measured expansion (%): 0.349, Field age of last recorded expansion (days): 6935, Expansion over time.

As previously mentioned, the dataset primarily reflects results from AMBT [11] and CPT [14] tests, which are the focus of this study. The classification of aggregate reactivity for each test follows ASTM C1778 standard protocol [13], with thresholds of 0.10% and 0.04% for AMBT and CPT at 14 and 365 days, respectively.

For the binary classification of field reactivity (i.e., YES | NO), a threshold of 0.04% was chosen, as this expansion level typically corresponds to visible external cracking in concrete. Additionally, when not indicated, the last recorded expansion data was adopted as the reference for the field's ultimate expansion in this analysis, reflecting the entire history of the reaction for each member under evaluation.

Additionally, data filtering focused on selecting the specific test results (i.e., AMBT and CPT) for mixes with and without the SCMs, to ensure distinguishing their effects on the analysis. Samples with field reinforcement, and those containing admixtures other than air entrainment were excluded from this analysis. Additional variables representing field conditions (i.e., temperature), mixes (i.e., alkali content) and mitigation options (i.e., SCMs amount (%)) were integrated into the analysis to account for their influence on AAR development.

7.4.2 Performance metrics

When assessing a model's performance or the effectiveness of laboratory tests such as the AMBT and CPT in predicting AAR outcomes, performance metrics are critical to consider. Accuracy, for instance, represents the overall correctness of the test. It is calculated as the ratio of correct predictions (i.e., both true positives and true negatives) to the total number of predictions, offering a measure of how often the laboratory test outcomes align with actual field performance. Precision, on the other hand, measures the accuracy of positive predictions. It is defined as the proportion of true positive results (i.e., correctly identified reactive cases) relative to the total number of positive predictions (i.e., true positives + false positives). A high precision indicates a low false positive rate, essential when false positives are significant.

Additionally, recall or sensitivity assesses the test's ability to correctly identify true positives. It is calculated by comparing the number of true positives to the sum of true positives and false negatives. High sensitivity is particularly decisive in minimizing the risk of missing reactive cases, as it relates to the test's ability to correctly identify reactivity. Finally, the F1-score, the harmonic mean of precision and recall, provides a balanced metric accounting for both false positives and false negatives, making it useful for uneven class distributions or when both precision and recall are equally important.

Therefore, to improve the predictive performance of laboratory tests, the focus should be on reducing false negatives (i.e., missed reactive cases), while increasing the probability of correctly identifying reactive samples. This approach aims to enhance sensitivity (i.e., true positive rate) while ensuring that non-reactive cases are accurately identified, thereby improving specificity (i.e., true negative rate). In other words, the goal is to detect the most reactive cases without raising the false positive rate.

7.4.3 DT Model Development

DT, a type of supervised machine learning model, is particularly well-suited for tasks where the goal is to establish decision rules based on a set of input variables [31]. Following the structure presented in Figure 7. 2, DT can effectively define threshold rules, making them ideal for determining the optimal expansion levels and time points that predict reactivity. Each split in the tree aims to maximize the separation between reactive and non-reactive outcomes, thereby refining the predictive accuracy.

Along its structure, the root node represents the entire dataset, being the starting point of the tree and the most significant decision is made. Next, internal nodes represent the points where the data is further split based on additional characteristics, leading to further branching. Leaf nodes are then the terminal points providing the output or classification where no further splits occur, representing a final classification made by the model.

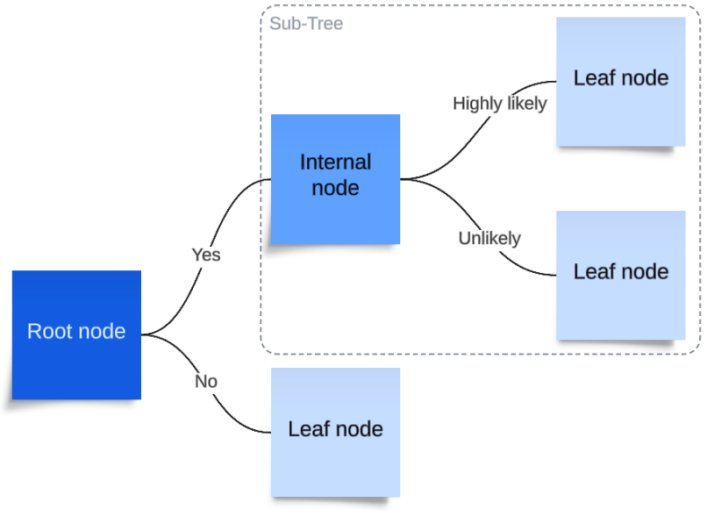


Figure 7. 2 – Decision tree structure, including root, internal and leaf nodes.

In this sense, the dataset was split into training and test sets using a 70-30 ratio, ensuring that the model's performance could be robustly evaluated on unseen data. Moreover, the model was trained using a recursive partitioning approach, where data was split based on expansion percentages at key time intervals. To prevent overfitting and ensure the model remained interpretable, the DT was constrained to a maximum depth of four. Additionally, a minimum of ten samples were required to split an internal node, and at least five samples were needed at a leaf node.

The decision tree was initially employed to identify and interpret the critical thresholds for predicting field reactivity. Its visualization helped extract decision rules that represent practical thresholds for laboratory test outcomes. To further refine these thresholds, a second decision tree model was developed, incorporating additional parameters such as alkali content (i.e., cement alkali content plus boosting alkali) and environmental factors (i.e., temperature and relative humidity).

This enhanced model allowed for a more comprehensive analysis of the factors influencing AAR risk, enabling the identification of specific thresholds and rules. The second decision tree provided deeper insights into the interactions between variables, aiming to improve the accelerated laboratory outcomes' predictive accuracy in estimating field reactivity based on laboratory results.

7.4.4 Logistic regression

Logistic regression is a classification machine learning model used to predict the probability of a binary outcome, such as the likelihood of field reactivity [31,32]. The logistic regression model is based on the logistic function Equation 7.1, which transforms input data into a probability value ranging between 0 and 1, suitable for modeling binary outcomes.

$$Logistic\ function = \frac{1}{1 + e^{-x}} \quad \text{(Equation 7. 1)}$$

In this context, a weight factor is used to predict the log odds of the probability that a given input belongs to a particular class (0 or 1). The log odds, or the logit function Equation 7.2 is defined as the natural logarithm of the odds ratio, in which P represents the probability of the event occurring.

$$odds = \frac{P}{P - 1} \ \rightarrow logit(P) = ln\left(\frac{P}{1 - P}\right) \quad \text{(Equation 7.2)}$$

Considering the logit function, the probability occurrence can be estimated as per Equation 7.3.

$$logit(P) = \ mx + b \ \rightarrow P(x) = \frac{1}{1 + e^{-(mx+b)}} \quad \text{(Equation 7.3)}$$

The maximum likelihood estimation (MLE) is the loss function in the logistic regression; it aims to find the parameter values that maximize the likelihood of the observed data. In this process, MLE optimizes the model parameters by interactively adjusting them to ensure the best fit between the model and the data. MLE is expressed by Equation 7.4, where y_i is the actual output and $\hat{y}_i$ is the predicted probability.

$$L(m) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i) \quad \text{(Equation 7.4)}$$

In logistic regression, each coefficient represents the change in the log odds of the outcome for a one-unit change in the predictor variable, while keeping the other variables constant. In that sense, this approach was employed to train the model to predict the occurrence of reactivity in the field (Equation 7.5) based on the accelerated laboratory test outcomes (i.e., aggregate reactivity levels or expansion, EXP), environmental conditions of exposed elements (i.e., temperature, TEMP and relative humidity, RH) and mix-design information (i.e., alkali content of the binder, ALK).

$$x=\beta_0+\beta_{TEMP}\cdot\text{TEMP}+\beta_{RH}\cdot\text{RH}+\beta_{EXP}\cdot\text{EXP}+\beta_{ALK}\cdot\text{ALK} \quad \text{(Equation 7.5)}$$

The probability of reactivity can, therefore, be calculate as per Equation 7.6.

$$P(x) = \frac{1}{1 + e^{-(\beta_0+\beta_{TEMP}\cdot\text{TEMP}+\beta_{RH}\cdot\text{RH}+\beta_{EXP}\cdot\text{EXP}+\beta_{ALK}\cdot\text{ALK})}} \quad \text{(Equation 7.6)}$$

7.5 Performance of laboratory tests for predicting AAR

As discussed in [43,44], the performance of AMBT and CPT under current standard testing conditions and thresholds reveals both strengths and limitations. For mixes without SCMs, CPT at 0.04% expansion at 365 days demonstrates reasonable performance, with an accuracy rate of 68%, effectively identifying true reactive cases. However, there is room for improvement in reducing false positives to enhance precision. In contrast, AMBT shows higher sensitivity (90%), meaning it is very effective in detecting true reactive cases, but with a lower precision (74%), indicating a higher rate of false positives compared to CPT. This results in more non-reactive cases being falsely classified as reactive, which highlights the need to improve its reliability.

For mixes containing SCMs, the performance of CPT at 0.04% expansion at 730 days drops significantly in terms of sensitivity, indicating challenges in accurately identifying true positive cases, and leading to an increased number of false negatives (missed reactive cases). While CPT achieves high specificity (91%), it struggles with reactive case detection. On the other hand, AMBT continues to show strong sensitivity (75%) in identifying reactive cases, but its precision remains limited, with half of its positive predictions being false positives, resulting in a lower specificity (64%).

These findings highlight the need to re-evaluate current thresholds, suggesting that the analysis should not be limited to a fixed laboratory interval, but incorporate a flexible analysis over time. Therefore, the figures below (Figure 7. 3, Figure 7. 4, Figure 7. 5, Figure 7. 6) reflect the evaluations performed by not only changing the threshold (i.e., laboratory expansion) but also the time effect on the performance metrics (i.e., accuracy, precision, recall, f-1 score). This analysis supports further decisions regarding the establishment of new approaches aiming to enhance the predictive capacity of current laboratory tests.

Without SCMs

As illustrated in Figure 7. 3, the trend for mixes without SCMs under AMBT shows that increasing accuracy corresponds with lower expansion values (i.e., up to 0.2%) and laboratory duration below 15 days. This pattern is similarly observed for recall and F1-score, both of which peak under similar conditions, indicating that shorter testing durations and lower expansion thresholds enhance the test's ability to correctly identify reactive cases while maintaining a balanced performance. However, precision (i.e., the accuracy of positive predictions) demonstrates an opposing behavior, improving with higher expansion values and longer testing times. This divergence suggests that while precision benefits from extended testing, potentially reducing false positives, it decreases accuracy and recall, which are critical for reliably detecting true positives.

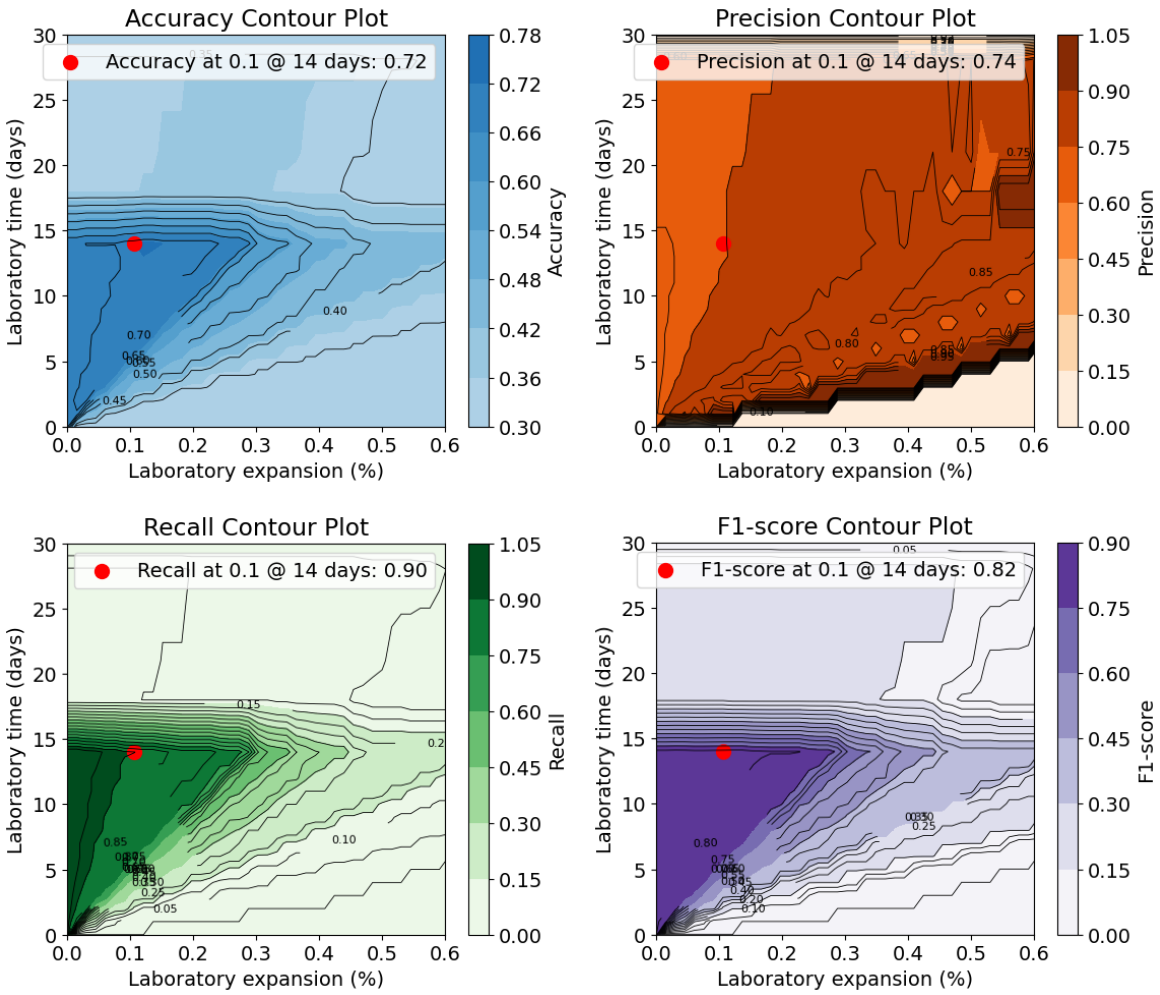


Figure 7. 3 – Performance evaluation of AMBT when varying both expansion threshold (i.e., laboratory expansion) and time (i.e., laboratory time) for mixes without SCMs.

Mixes without SCMs under the CPT, as shown in Figure 7. 4, also show that accuracy tends to increase with lower expansion values (i.e., up to 0.1%) and testing times up to 400 days. This pattern is consistently observed across other performance metrics, such as recall and F1-score, suggesting that during this period, the CPT is more effective at correctly identifying reactive cases, thereby enhancing the overall reliability of the test. However, precision behaves differently, tending to rise with higher expansion values and lower test duration. This indicates that while extended testing periods improve accuracy, recall, and F1-score, higher expansion thresholds and shorter testing duration may be necessary to maintain precision.

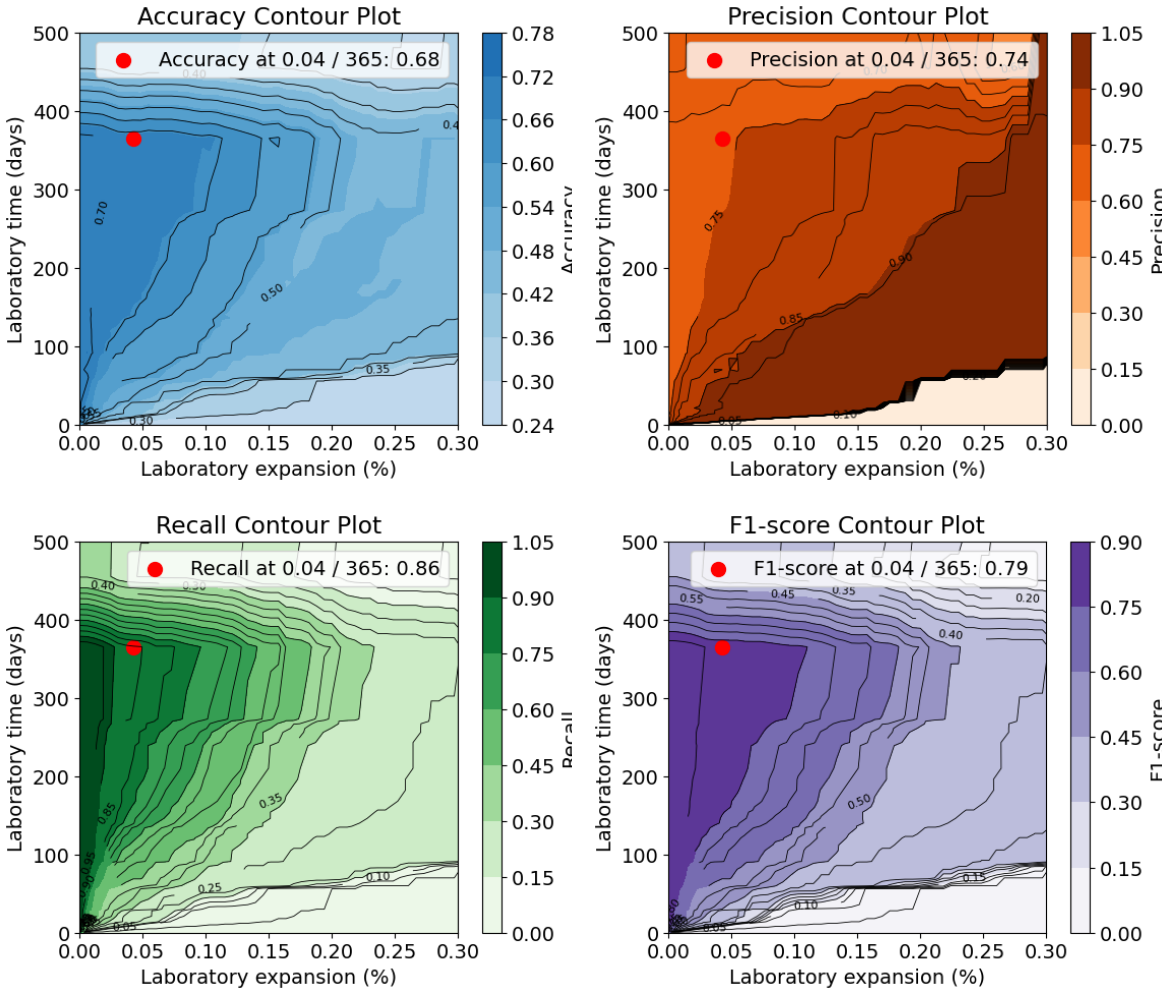


Figure 7. 4 – Performance evaluation of CPT when varying both expansion threshold (i.e., laboratory expansion) and time (i.e., laboratory time) for mixes without SCMs.

With SCMs

As illustrated in Figure 7. 5, the trend for mixes containing SCMs under AMBT reveals that accuracy is highly dependent on the combination of time and expansion. However, recall exhibits an opposite behavior, showing peak performance under conditions that do not align closely with those that optimize accuracy. This divergence indicates that a careful calibration analysis is necessary to identify the optimal scenario where accuracy can improve alongside the test's ability to detect all reactive cases, as measured by recall. Conversely, precision and F1-score demonstrate similar behaviors, both increasing slightly with higher expansion

values, though they remain relatively low overall. This suggests that while the test is somewhat effective at balancing true positive rates with false positives, its overall reliability in identifying reactive cases in mixes containing SCMs is limited.

When compared with the performance of AMBT without SCMs (Figure 7. 3), the presence of SCMs appears to complicate the predictive performance, leading to generally lower accuracy and recall, as well as similarly low precision and F1-scores. This comparison suggests that further segmentation is needed to better understand these trends, potentially by examining factors such as the types of SCMs, the amount of SCMs, alkali content, and environmental conditions.

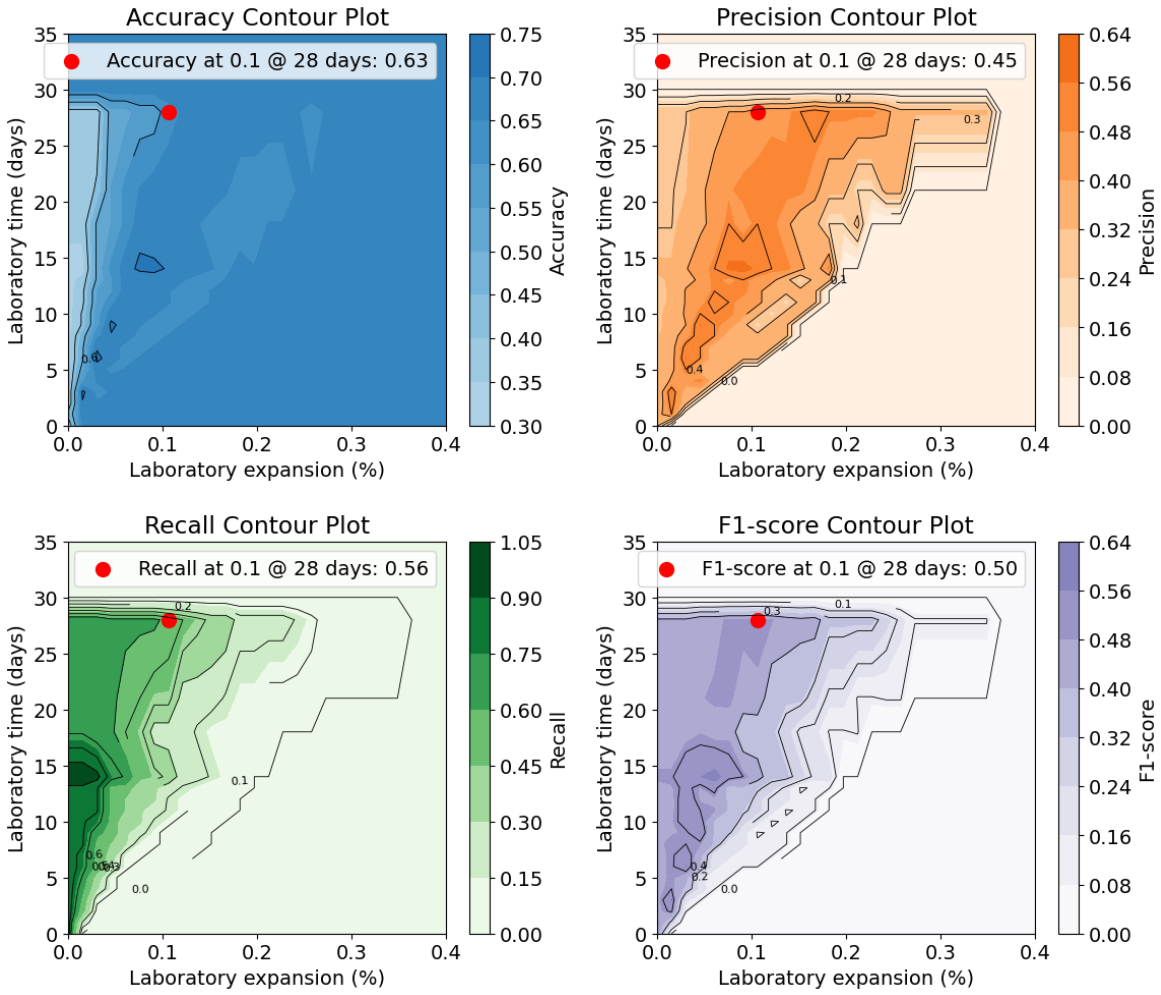


Figure 7. 5 – Performance evaluation of AMBT when varying both expansion threshold (i.e., laboratory expansion) and time (i.e., laboratory time) for mixes with SCMs.

Figure 7. 6 shows the trend for mixes containing SCMs under the CPT, indicating a variable performance on the accuracy along testing time and expansion, mostly improving with extended testing times, or with lower expansion values. This suggests that long-term testing is associated with higher accuracy, due to the slower reaction kinetics associated with SCMs. In contrast, recall exhibits better performance with lower expansion and shorter test duration. Precision, on the other hand, shows a consistent increase with both higher expansion values and longer test durations. This suggests that as the test runs longer and expansion increases, the ability to correctly identify true positives without falsely categorizing non-reactive cases as reactive improves, thereby reducing the likelihood of false positives. Finally, F1-score, which balances precision and recall, indicates better performance at lower expansion levels with extended testing durations. This indicates that while precision and recall may vary independently, their combined effectiveness in predicting reactivity is maximized under these specific conditions.

When compared to CPT without SCMs (Figure 7. 4), CPT with SCMs generally exhibits lower recall and precision, despite achieving reasonable accuracy with extended testing or lower expansion values. Additionally, the presence of SCMs complicates the predictive performance, similar to the trends observed with AMBT for mixes containing SCMs (Figure 7. 5), suggesting that further analysis is required. Specifically, exploring the different types and amounts of SCMs, and varying environmental conditions, in the reliability of these tests in predicting AAR occurrence in the field.

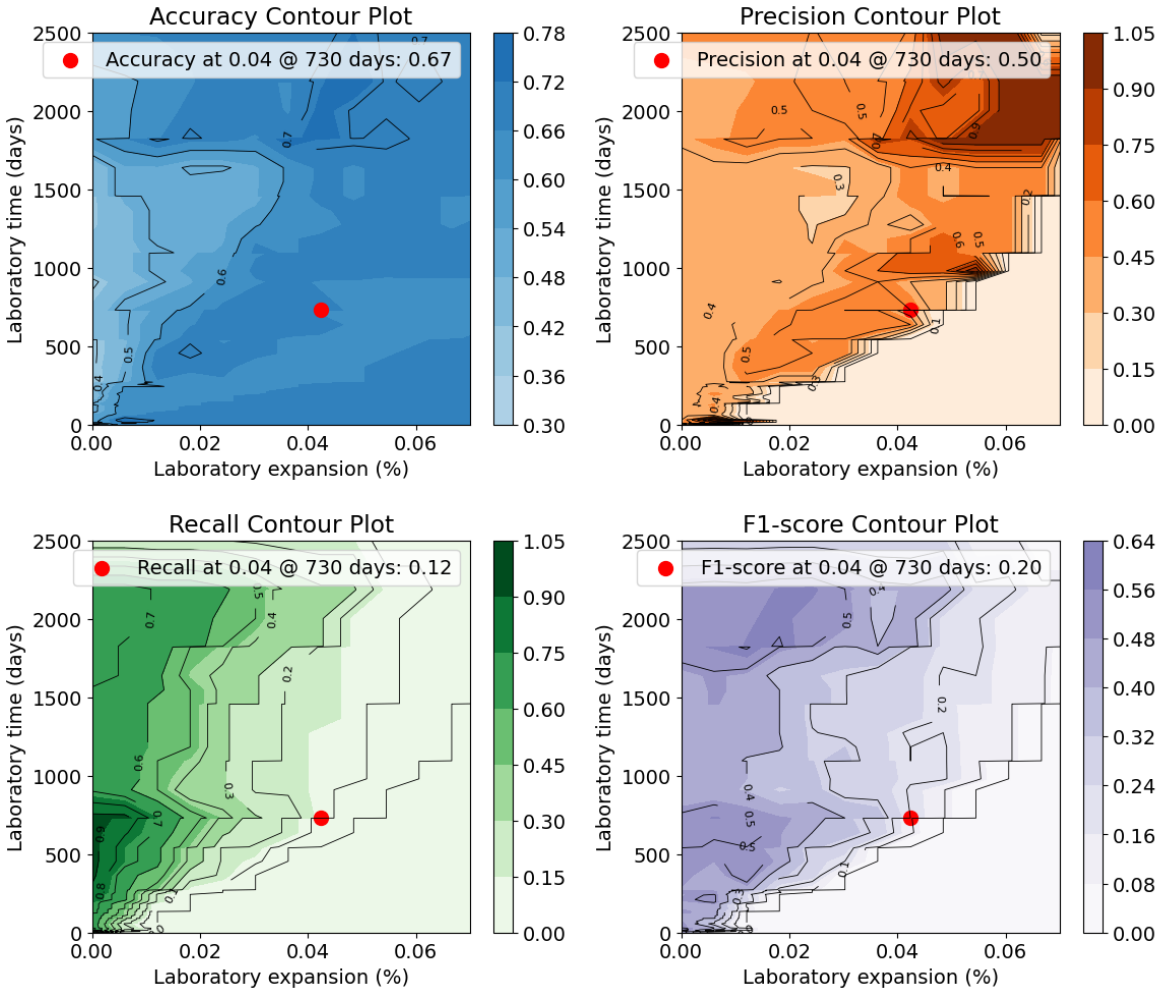


Figure 7. 6 – Performance evaluation of CPT when varying both expansion threshold (i.e., laboratory expansion) and time (i.e., laboratory time) for mixes with SCMs.

7.6 DT outcomes | Threshold analyses

The application of DT models to evaluate laboratory test outcomes for mixes with and without SCMs revealed distinct threshold rules. After employing a general model to split the categories based on expansion and test duration, a secondary analysis was employed incorporating alkali content and field temperature parameters. The results below include the models' prediction metrics and the set of rules observed.

Without SCMs

The DT model for mixes without SCMs represented by AMBT outcomes, observed in Figure 7. 7, defines a general threshold based on laboratory expansion at 10 days. The DT established that an expansion greater than 0.125% at 10 days is indicative of a potentially reactive aggregate. This model achieved an accuracy of approximately 60%, indicating a moderately reliable classification. It is worth mentioning the classification process considers different test periods and thresholds, indicating the complex influence of the reaction kinetics in splitting the outcomes into reactive and non-reactive.

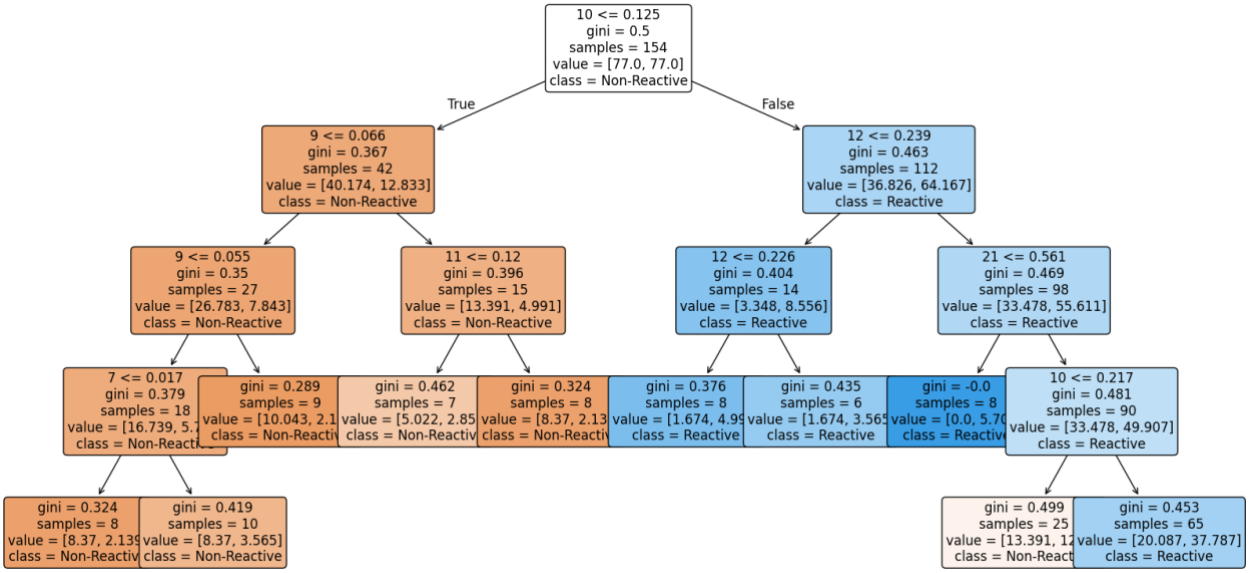


Figure 7. 7 – DT for mixes without SCMs and AMBT outcomes, establishing the general threshold rule regarding laboratory expansion and testing duration.

Aiming to incorporate additional parameters such as alkali content and field temperature to understand their impact on the classification process, a second DT model has been developed (Figure 7. 8). This model demonstrated a slight improvement in accuracy, reaching approximately 63%. The model identified critical conditions under which a mix is classified as either non-reactive or reactive. For instance, a mix is classified as non-reactive if the expansion at 10 days is $\leq$ 0.125%, regardless of the alkali content or the field temperature. However, if the expansion at 10 days is greater than 0.13%, a conditional about the alkali content (i.e., lower than 4.61 kg/m^3) can indicate a non-reactive behaviour in the field.

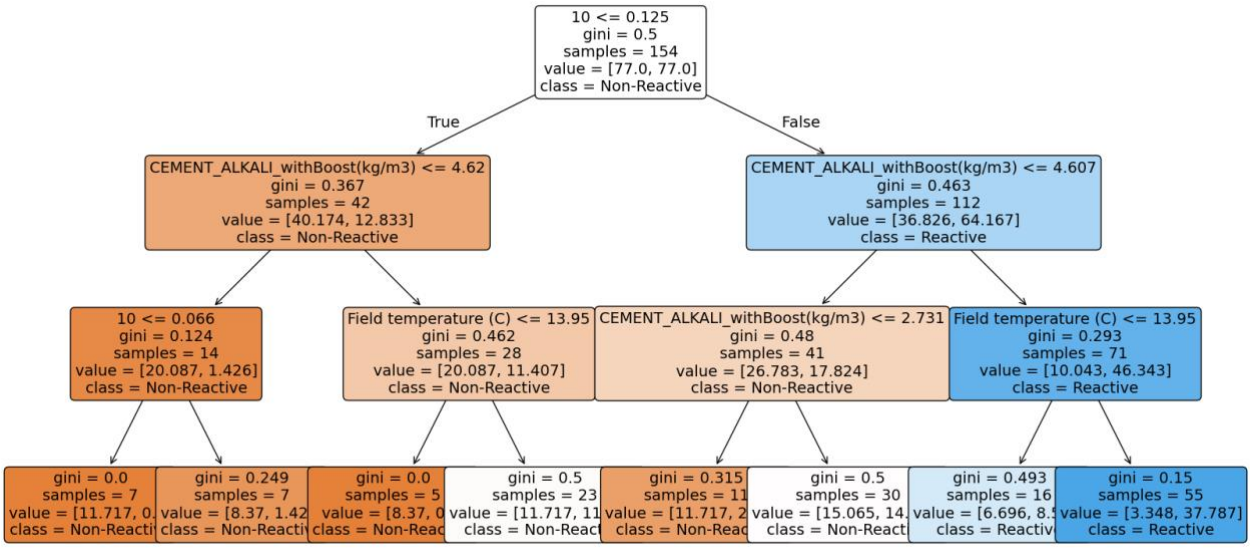


Figure 7. 8 – DT for mixes without SCMs and AMBT outcomes, establishing the general threshold rule regarding laboratory expansion and testing duration incorporating alkali content and field temperature parameters.

In summary, and incorporating previous knowledge about the ASR to eliminate irrelevant rules and small representative samples, the following classification applies to AMBT mixes without SCMs:

AMBT without SCMs <u>non-reactive</u>:

- Expansion at 10 days is ≤ 0.125%
- Expansion at 10 days is > 0.125% **and** Alkali content is ≤ 4.61 kg/m^3

AMBT without SCMs <u>reactive</u>:

Expansion at 10 days is > 0.125% **and** Alkali content is > 4.61 kg/m^3

Similarly, the DT model for CPT outcomes without SCMs, represented in Figure 7. 9, establishes a general threshold rule based on laboratory expansion. This model demonstrated limited accuracy, with an overall accuracy of 53% and identified a threshold of 0.115% expansion as critical for classifying mixes as reactive. Generally, if the expansion exceeds 0.115% at 136 days, there is a reactive potential in the field. More complex rules are also

observed based on different test periods (i.e., 1, 9, 12, 13, 60, 70, and 273 days) and expansion limits, highlighting the importance of the reaction kinetics.

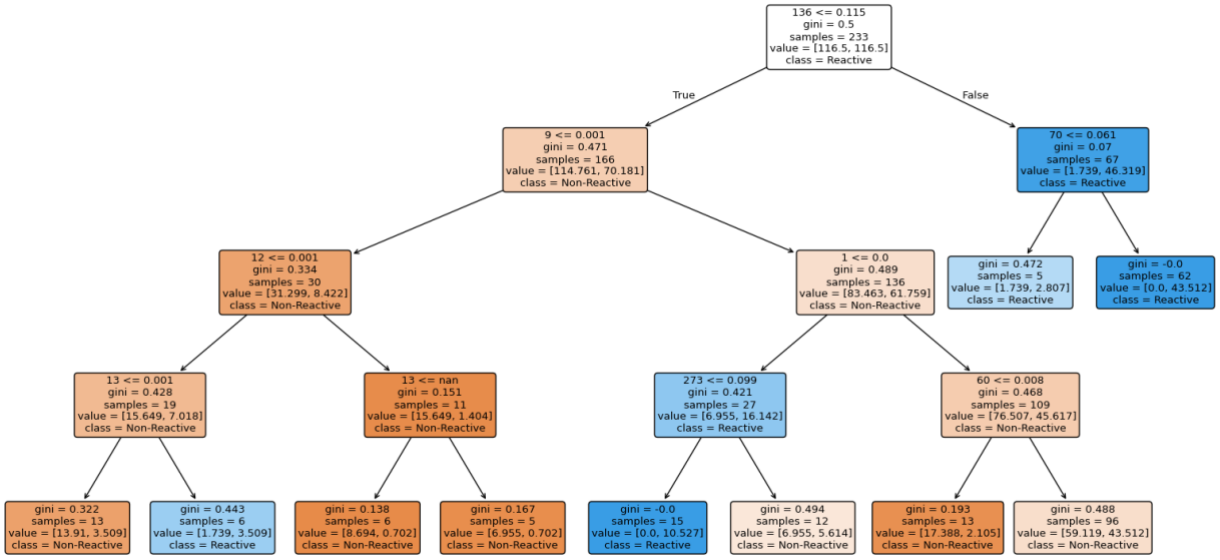


Figure 7. 9 – DT for mixes without SCMs and CPT outcomes, establishing the general threshold rule regarding laboratory expansion and testing duration.

Incorporating additional parameters such as the alkali content and field temperature (Figure 7. 10) over the defined period (i.e., 136 days), a secondary DT has been developed. This model demonstrated significantly improved accuracy, reaching 81%, and generally indicated a more complex set of rules in defining reactive and non-reactive cases. The results highlight the interdependencies between expansion thresholds, alkali content, and temperature, suggesting that these factors jointly influence reactivity occurrence.

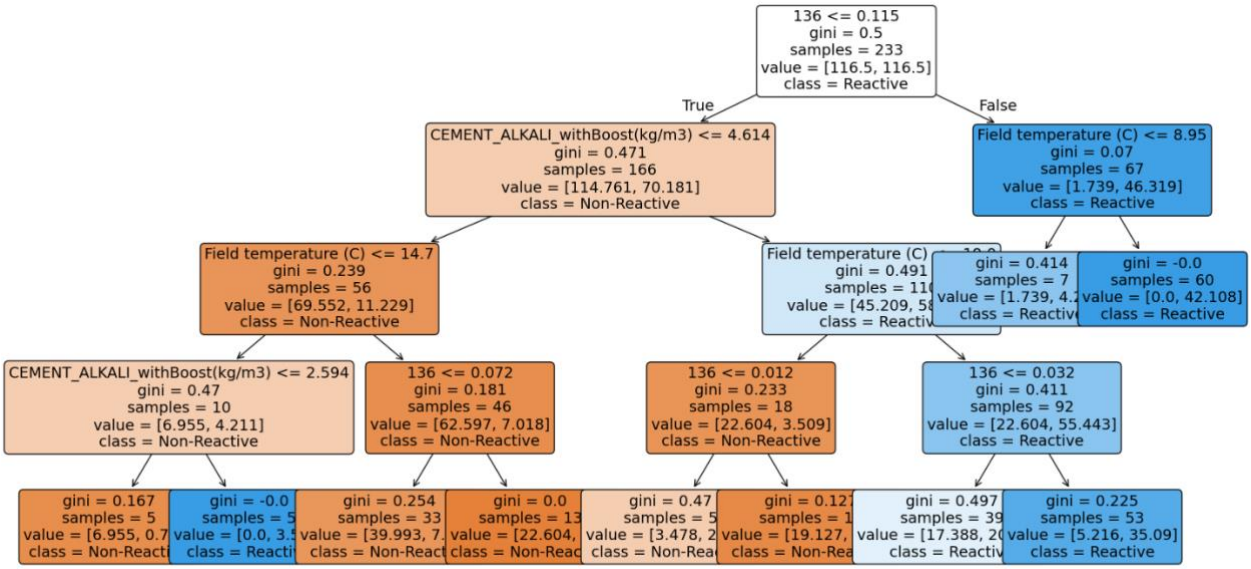


Figure 7. 10 – DT for mixes without SCMs and CPT outcomes, establishing the general threshold rule regarding laboratory expansion and testing duration incorporating alkali content and field temperature parameters.

In summary, and incorporating previous knowledge about the ASR to eliminate irrelevant rules and small representative samples, the following classification applies to CPT mixes without SCMs:

CPT without SCMs <u>non-reactive</u>:

- Expansion at 136 days is ≤ 0.115% and Alkali content ≤ 4.61kg/m3
- Expansion at 136 days is ≤ 0.115% and Alkali content > 4.61kg/m3 and Field temperature < 19°C

CPT without SCMs <u>reactive</u>:

- Expansion at 136 days is > 0.115%
- Expansion at 136 days is ≤ 0.115% and Alkali content > 4.61kg/m3 and Field temperature > 19°C

With SCMs

The DT model for mixes with SCMs for AMBT, as shown in Figure 7. 11, establishes a general threshold rule based on the laboratory expansion at 28 days. The DT identified that an expansion greater than 0.097% at 28 days is indicative of a reactive mix. This model achieved an accuracy of approximately 53%.

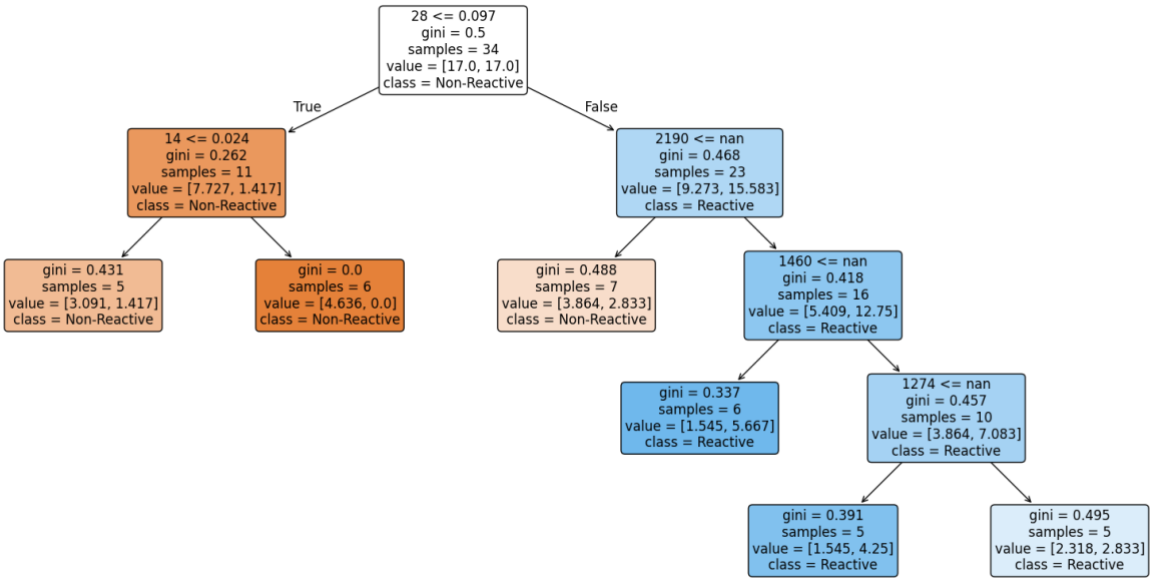


Figure 7. 11 – DT for mixes with SCMs and AMBT outcomes, establishing the general threshold rule regarding laboratory expansion and testing duration.

The second DT model incorporating additional parameters such as alkali content and field temperature is presented in Figure 7.12. This model showed a similar accuracy of approximately 53%. Yet, when limited to analyzing the outcomes from 28 days, a rule on alkali content is presented splitting reactive and non-reactive cases. Next, it states at 0.094% expansion level at 28 days to split between reactive and non-reactive.

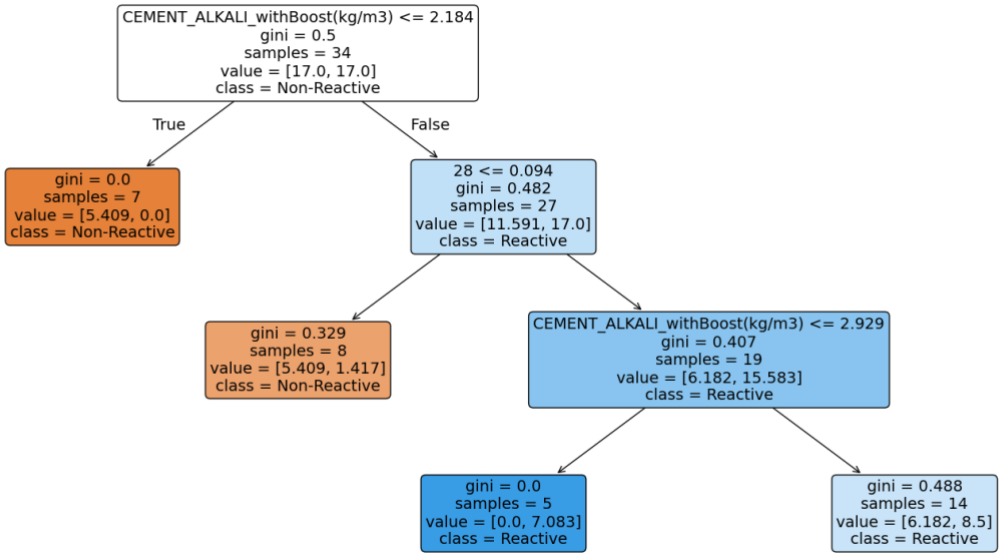


Figure 7. 12 – DT for mixes with SCMs and AMBT outcomes, establishing the general threshold rule regarding laboratory expansion and testing duration incorporating alkali content and field temperature parameters

In summary, the rules for mixes with SCMs for AMBT outcomes are suggested to be as follows:

AMBT with SCMs as non-reactive:

- Alkali content is ≤ 2.18 kg/m3
- Alkali content is > 2.18 kg/m3 and Expansion at 28 days is ≤ 0.094%.

AMBT with SCMs as reactive:

- Alkali content is > 2.18 kg/m3 and Expansion at 28 days is > 0.094%.

Similarly, the DT model for CPT outcomes with SCMs, represented in Figure 7. 13, establishes a general threshold rule based on laboratory expansion at 455 days. The DT identified that an expansion greater than 0.014% at 455 days indicates a reactive mix. The model demonstrated an overall accuracy of approximately 47% in classifying the testing data.

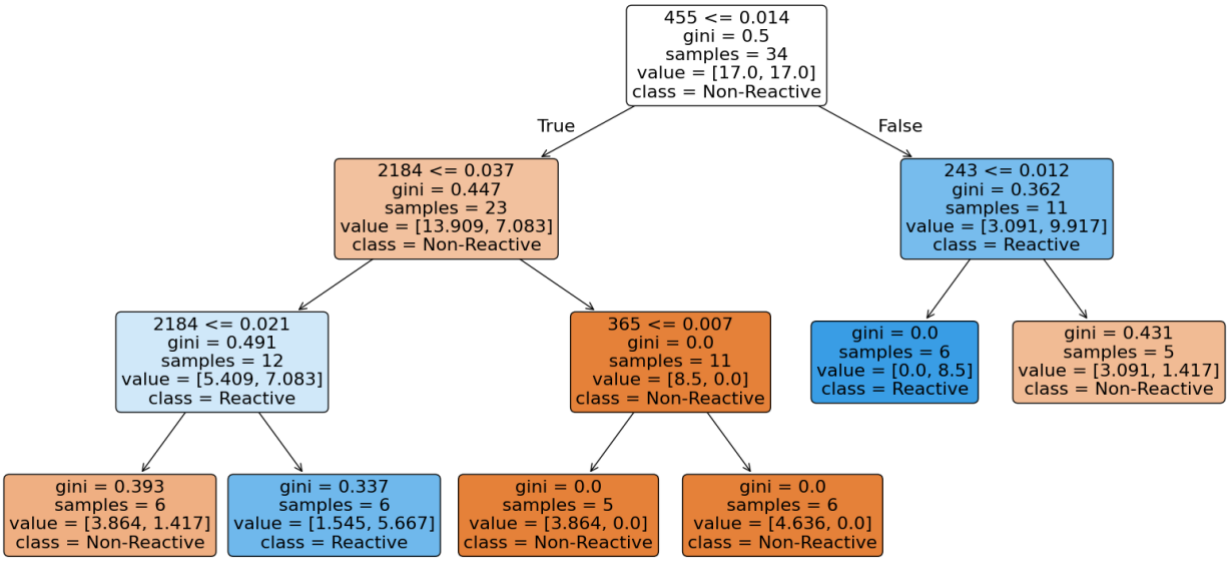


Figure 7. 13 – DT for mixes with SCMs and CPT outcomes, establishing the general threshold rule regarding laboratory expansion and testing duration.

A second DT model incorporates additional parameters such as alkali content and field temperature (Figure 7. 14). This model didn't show an increase in accuracy, reducing to approximately 33%. Similarly to what was observed for AMBT, a general rule is stated for alkali content and next for the expansion and test duration.

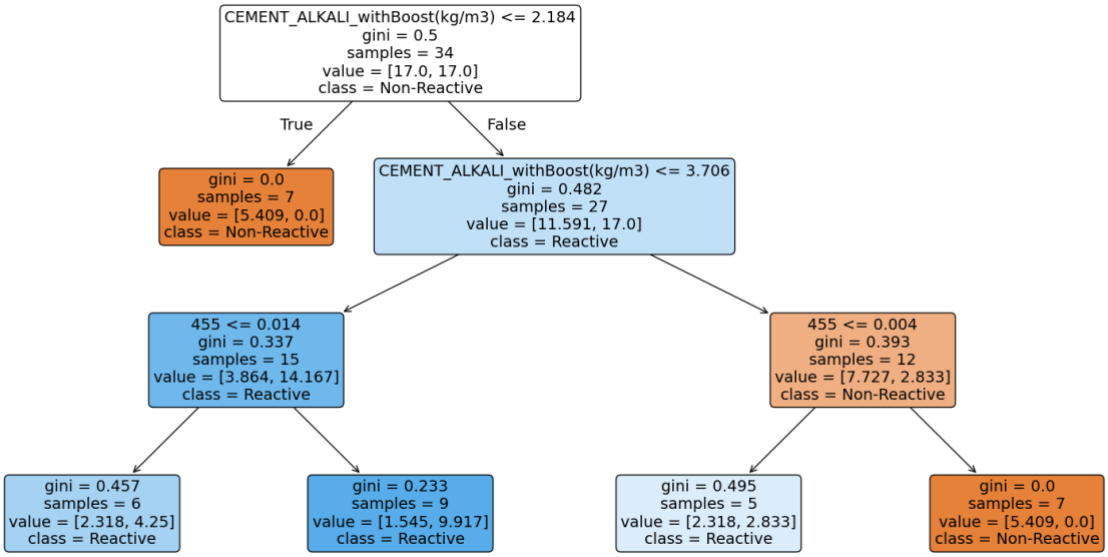


Figure 7. 14 – DT for mixes with SCMs and CPT outcomes, establishing the general threshold rule regarding laboratory expansion and testing duration incorporating alkali content and field temperature parameters

In summary, the classification for CPT mixes containing SCMs is primarily based on alkali content. Specifically, mixes with alkali content ≤ 2.18 kg/m³ are classified as non-reactive, while those with alkali content between > 2.18 kg/m³ and ≤ 3.71 kg/m³ are considered reactive. Consequently, the general rule for predicting reactivity will be applied when expansion exceeds 0.014% at 455 days.

CPT with SCMs as non-reactive:

- Expansion ≤ 0.014% at 455 days

CPT with SCMs as reactive:

- Expansion > 0.014% at 455 days

These results underscore the importance of considering environmental conditions (i.e., temperature) and mix characteristics (i.e., alkali content) alongside laboratory expansion results when assessing the AAR field reactivity potential. Yet, demanding a thorough assessment of each variable for incorporating such findings into a general framework.

7.7 Probabilistic estimation of field reactivity occurrence

As observed, predicting the occurrence of AAR in the field based on laboratory test outcomes is a complex task due to the numerous interacting variables and the variability in test durations. While DT analysis has been essential in identifying critical points within the dataset, it also highlights the limitations of using simplistic models for field predictions. To address these challenges and better understand the impact of each variable on AAR, a logistic regression model has been employed. This model allows for probabilistic estimation of AAR occurrence in field conditions, taking into account environmental factors, mix design parameters, and the expansion over time of accelerated laboratory tests. The ultimate goal is to establish a fexible threshold-time approach that can be systematically applied to reduce the risk of AAR across diverse concrete structure scenarios. This model has been specifically designed to include variables that are most relevant to AAR development, such as environmental conditions (i.e., temperature (TEMP), relative humidity (RH)), laboratory expansion (EXP), and field alkali content (ALK). The model is defined as per Equation 7.7 and the probability of reactivity can, therefore, be calculate as per Equation 7.8.

$$x=\beta_0+\beta_{TEMP}\cdot\mathrm{TEMP}+\beta_{RH}\cdot\mathrm{RH}+\beta_{EXP}\cdot\mathrm{EXP}+\beta_{ALK}\cdot\mathrm{ALK} \quad \text{(Equation 7.7)}$$

$$P(x)=\frac{1}{1+e^{-(\beta_0+\beta_{TEMP}\cdot\mathrm{TEMP}+\beta_{RH}\cdot\mathrm{RH}+\beta_{EXP}\cdot\mathrm{EXP}+\beta_{ALK}\cdot\mathrm{ALK})}} \quad \text{(Equation 7.8)}$$

It is worth mentioning that given the variability in the reporting times of different test durations from the database, linear interpolation was performed to fill in the gaps, ensuring that all test durations could be consistently analyzed. Additionally, the analysis was performed only on mixes without SCMs given the data availability. Moreover, the binary outcome (i.e., YES or NO) was based on the last recorded expansion of field blocks, meaning an average of 8.3 and 6.2 years for AMBT and CPT, respectively. Yet, they are explored in time ranges in section 7.8. The outcomes of the logistic regression model are provided for both the AMBT and the CPT without the use of SCMs, as observed in Table 7. 1 and

Table 7. 2.

Table 7. 1 – Overall coefficients for the AMBT regression model.

Age (days)	β_0	β_{TEMP}	β_{RH}	β_{EXP}	β_{ALK}	Accuracy
5	0.276	0.008	-0.085	13.156	0.983	0.77 (44)
7	0.012	0.014	-0.087	10.151	1.029	0.77 (44)
9	0.047	0.011	-0.087	8.013	1.024	0.75 (44)
10	0.127	0.008	-0.087	7.204	1.018	0.75 (44)
11	0.024	0.009	-0.085	6.520	1.009	0.75 (44)
12	-0.200	0.010	-0.082	5.992	1.008	0.75 (44)
13	-0.029	0.007	-0.083	5.478	0.997	0.75 (44)
14	0.065	0.006	-0.083	5.068	0.987	0.75 (44)

As observed, EXP emerges as the most significant predictor of AAR probability in AMBT, with consistently positive coefficients across all ages. In fact, EXP has a substantial impact, with coefficients starting at 13.156 at 5 days and decreasing to 5.068 by 14 days. This strong relationship indicates that the degree of expansion observed in laboratory conditions is a critical indicator of AAR potential in the field, with early higher expansions strongly correlating with higher AAR risk.

ALK consistently exhibits a positive correlation with AAR probability in the AMBT, underscoring the critical role of alkalis in promoting AAR. The ALK coefficient is relatively stable, ranging from 0.983 to 1.029. This stability suggests that alkali content has a sustained influence on AAR risk, regardless of the test duration. The consistent positive impact of ALK across all ages highlights the need for stringent control of alkali levels in concrete to mitigate AAR.

In the AMBT model, temperature exhibits a generally positive correlation with the likelihood of AAR occurrence. This aligns with the understanding that higher temperatures accelerate the chemical reactions associated with AAR, leading to increased expansion and cracking in concrete [24]. Finally, RH generally has a negative correlation also around zero with AAR probability. This trend suggests that lower humidity may have a minimal effect or even reduce the likelihood of AAR, given the combination of the other parameters.

Table 7. 2 – Overall Coefficients for the CPT regression model.

Age (days)	β_0	β_{TEMP}	β_{RH}	β_{EXP}	β_{ALK}	Accuracy
28	0.413	0.000	-0.053	30.014	0.736	0.78 (67)
90	0.420	-0.017	-0.071	15.654	0.975	0.78 (67)
121	-0.717	0.000	-0.064	13.921	1.033	0.79 (67)
154	1.857	-0.018	-0.089	10.408	0.904	0.79 (67)
182	1.860	-0.016	-0.089	9.308	0.887	0.79 (67)
210	1.480	-0.012	-0.084	8.887	0.883	0.79 (67)
243	1.259	-0.009	-0.081	8.466	0.870	0.78 (67)
270	1.064	-0.007	-0.078	7.713	0.859	0.78 (67)
365	1.009	0.002	-0.071	6.191	0.749	0.82 (67)

In the CPT model, laboratory expansion (EXP) plays a dominant role, with coefficients ranging from 30.014 at 28 days to 6.191 at 365 days. The strong positive correlation between EXP and AAR probability underscores the importance of laboratory expansion as a primary predictor of AAR, particularly due to the lower expansion ranges observed in CPT tests. Moreover, in the CPT model, the ALK coefficients are the second most influential parameters, remaining positive and ranging from 0.736 to 1.033. Although the impact of ALK diminishes slightly over longer durations, it remains a key factor in predicting AAR. This slight reduction in ALK influence could be attributed to the longer exposure period in CPT, which allows for more significant alkali leaching. As alkalis are gradually leached out of the concrete, their availability for driving AAR decreases, thereby slightly reducing the overall impact of alkali content on AAR probability over time.

The TEMP coefficients in CPT are slightly negative across various ages, ranging from -0.017 to 0.000. These negative coefficients are relatively close to zero, indicating a minimal effect of temperature on AAR probability in this context. However, the CPT model consistently shows negative RH coefficients across all test ages, with slightly more significant values ranging from -0.089 to -0.053, including the environmental factor in the model.

Finally, the analysis of the logistic regression model for both AMBT and CPT data reveals that laboratory expansion and alkali content are the most influential variables in predicting AAR probability, with consistently strong positive correlations. However, although

temperature and relative humidity show more complex and minimal effects, their effects can be captured when comparing different climates. Also, it is worth mentioning that the temperature and RH coefficients are associated with the data available mostly from two environments (i.e., Ottawa and Austin).

7.8 Flexible coupled threshold-time (FCTT) approach

Based on these findings, a parametric analysis is further conducted to define a flexible coupled threshold-time approach that accounts for variations in alkali content and environmental conditions, such as temperature and relative humidity (RH). Two distinct environments are considered based on data availability: a colder environment (i.e., TEMP = 7°C and RH = 75%) representing Ottawa, ON, Canada, and a warmer environment (i.e., TEMP = 21°C and RH = 63%) representing Austin, Texas, US. Subsequently, a time evolution analysis is performed to assess the progression of the reaction over defined time intervals (i.e., up to 5, 10, 15 and 22 years), evaluating its performance across different durations.

7.8.1 Impact of alkali content

In the context of a colder environment, the probabilistic estimation for AAR occurrence in the AMBT model reveals distinct behaviors based on varying alkali content. As observed in Figure 7. 15, the probability of AAR occurrence is strongly influenced by the combination of test duration and alkali content, with the expansion threshold playing a critical role in defining the risk levels.

At lower alkali content (i.e., 1.8 kg/m^3), the expansion threshold required to reach a high probability of AAR occurrence is relatively high, indicating that significant expansion is necessary before AAR becomes a major concern. However, as the alkali content increases, the expansion threshold decreases noticeably for the same risk levels. For instance, with an alkali content of 4.2 kg/m^3 or higher, even modest expansions within the 14-day test duration can result in a high probability of AAR, suggesting that concrete mixtures with higher alkali

levels are particularly susceptible to AAR, even when early expansion signs are not severe. This emphasizes the need to control alkali content in concrete as a mitigation action.

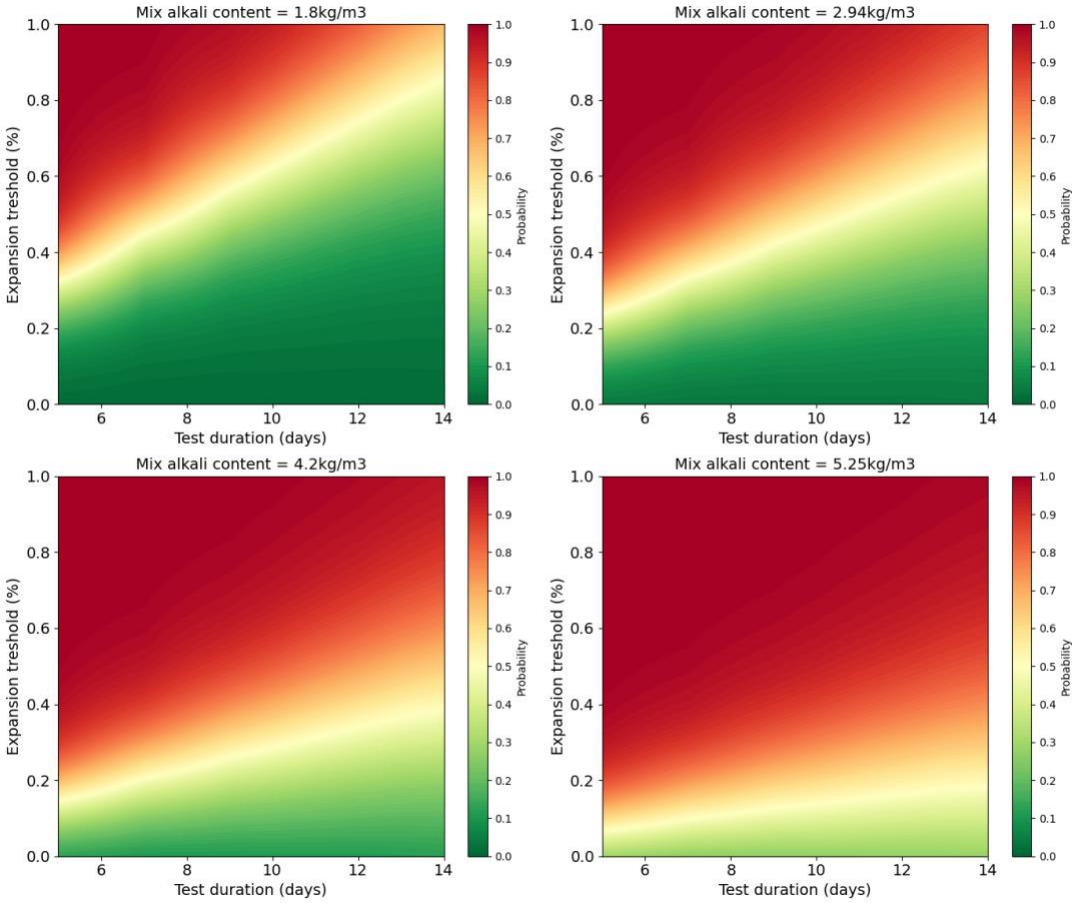


Figure 7. 15 – Probability of AAR occurrence based on AMBT results, under cold environment (i.e., TEMP=7C and RH=75%), for different alkali content: (a) 1.8 kg/m^3, (b) 2.94 kg/m^3, (c) 4.2 kg/m^3, and (d) 5.25 kg/m^3.

In the CPT model, also considering the colder environment, the influence of alkali content on AAR probability presents a slightly different pattern due to the longer test durations and lower expansion rates typically observed in CPT. As illustrated in Figure 7. 16, the expansion thresholds for AAR probability are generally lower than in AMBT, reflecting the more realistic field conditions simulated by the CPT.

At an alkali content of 1.8 kg/m^3, the probability of AAR remains low across most test durations, even with higher expansion thresholds. However, as the alkali content increases to

4.2 kg/m^3 and beyond, the probability of AAR rises significantly, even for relatively small expansions over extended durations, such as 365 days. Again, this indicates that higher alkali content can lead to significant AAR risk, even if the concrete exhibits only moderate expansion in the early stages. The CPT results reinforce the importance of controlling alkali levels in concrete, as even slight increases in alkali content can substantially raise the risk of AAR over time.

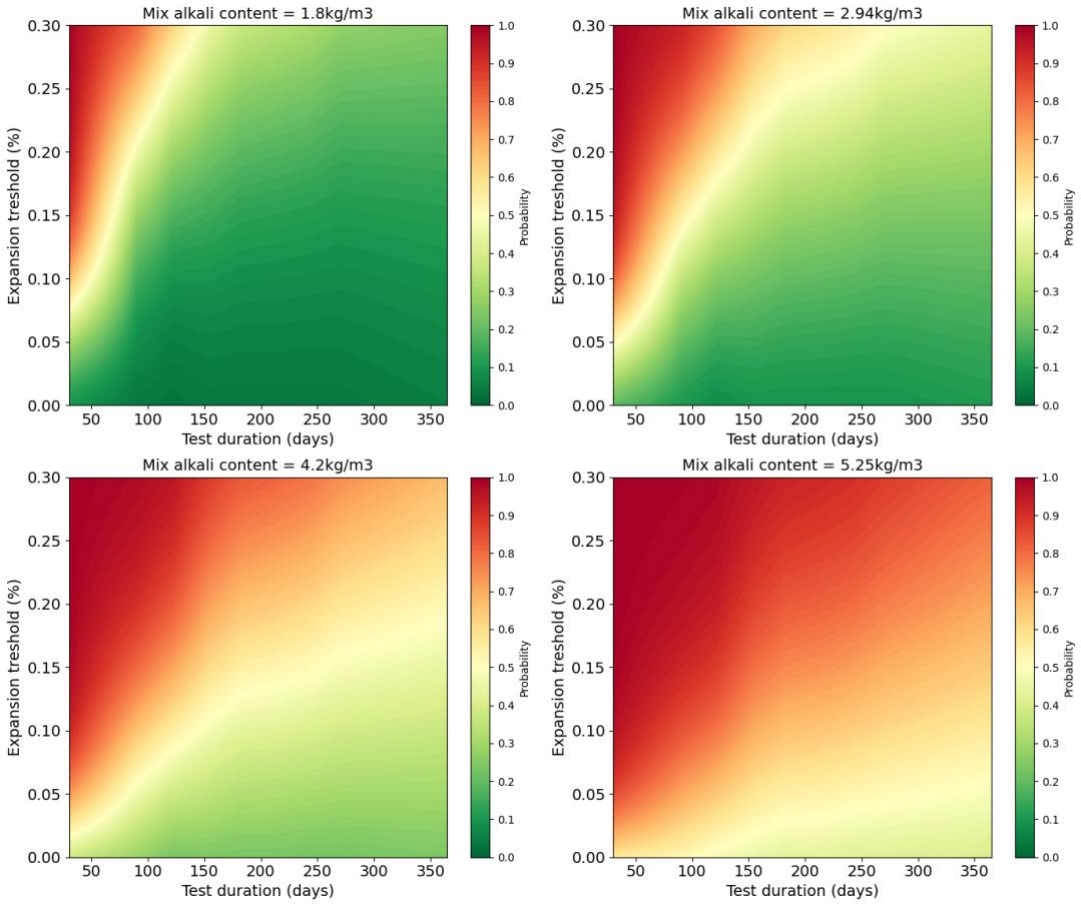


Figure 7. 16 – Probability of AAR occurrence based on CPT results, under cold environment (i.e., TEMP=7C and RH=75%), for different alkali content: (a) 1.8 kg/m^3, (b) 2.94 kg/m^3, (c) 4.2 kg/m^3, and (d) 5.25 kg/m^3.

7.8.2 Impact of environment

In examining the impact of environmental conditions on the probability of AAR occurrence, the AMBT model shows distinct differences when comparing cold (i.e., temperature=6.9^{o}C

and RH=75%) and warm (i.e., temperature=21°C and RH=63%) environments, with a constant alkali content of 2.94 kg/m^3, which is representative of most cements. As shown in Figure 7. 17a, under cold conditions, the expansion threshold required to reach a high probability of AAR is generally higher over the test duration. This suggests that in colder environments, the AAR process may be slower to initiate, requiring more significant expansion in the accelerated laboratory test before AAR becomes a critical concern.

However, when moving to a warmer environment as observed in Figure 7. 17b, the expansion threshold decreases, and the probability of AAR increases more rapidly over the same test duration. This indicates that warmer conditions accelerate the AAR process, making concrete more vulnerable to AAR at lower laboratory expansion levels.

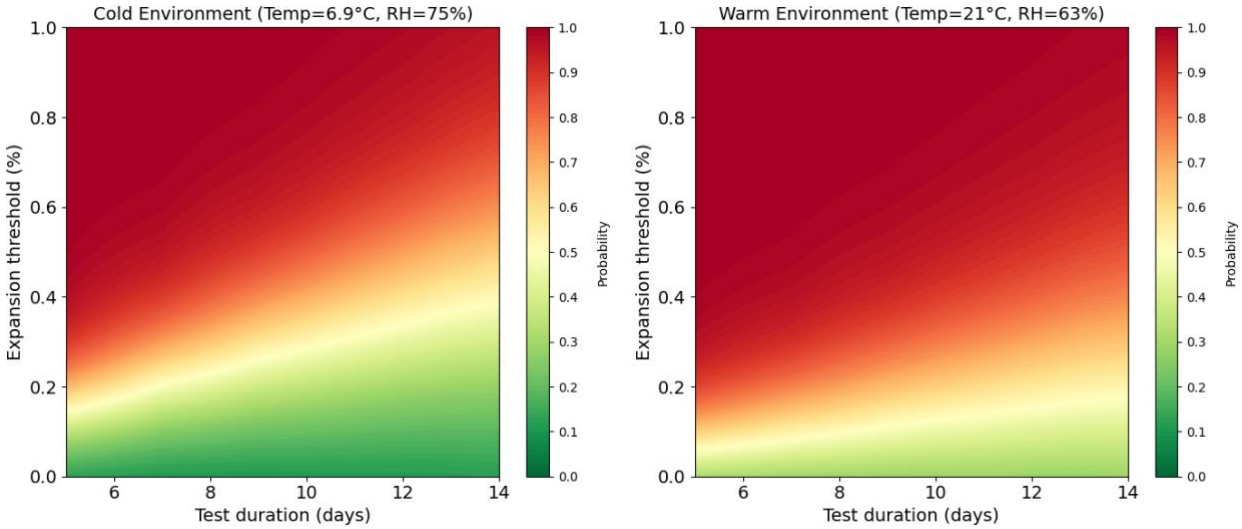


Figure 7. 17 – Probability of AAR occurrence based on AMBT results, under (a) cold environment (i.e., TEMP=7C and RH=75%), and (b) warm environment (i.e., TEMP=21C and RH=63%).

Similarly, the CPT model further illustrates the environmental impact on AAR probability, as shown in Figure 7. 18. In the cold environment scenario, the expansion thresholds are higher, particularly over the longer test durations. This reflects a slower AAR progression in colder conditions, where significant expansion over extended periods is necessary to reach a high probability of AAR.

Conversely, in a warm environment, the expansion thresholds are lower, and the AAR probability increases more quickly, even with smaller expansions. This suggests that in

warmer climates, AAR risk is heightened, and concrete structures may be more susceptible to damage due to faster reaction kinetics. Although the model coefficients for temperature and RH appear negligible (close to zero), these factors have a significant impact on the overall probability of AAR. In fact, the comparison between these environments emphasizes the critical role of temperature and humidity in influencing AAR and highlights the need for tailored mitigation strategies depending on the local climate conditions.

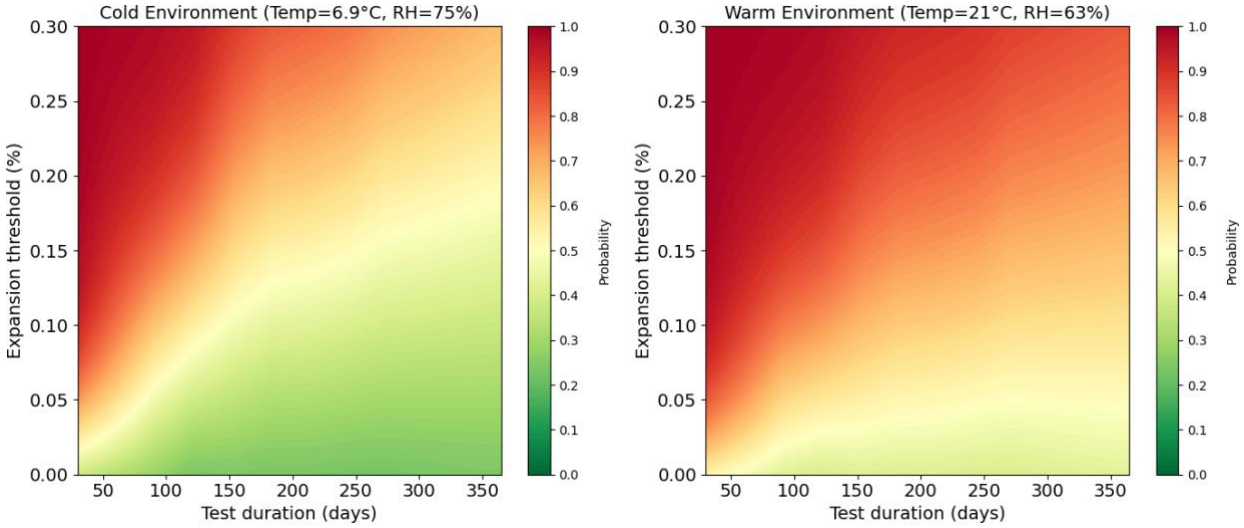


Figure 7. 18 – Probability of AAR occurrence based on CPT results, under (a) cold environment (i.e., TEMP=7C and RH=75%), and (b) warm environment (i.e., TEMP=21C and RH=63%).

7.8.3 Time analysis

Since the evaluations above are based on the last recorded expansion data for AMBT and CPT (average of 8.3 and 6.2 years of exposure, respectively), Figure 7. 19 and Figure 7. 20 provide a detailed visual representation of the model's predictions within specific timeframes of field exposure (i.e., up to 5, 10, 15, and 22 years). These figures help to illustrate how the expansion thresholds evolve over time and under different environmental conditions, enabling a clearer understanding of reactivity progression. In Appendix A-F, the logistic regression parameters are presented as the base of the current analysis.

For AMBT, the increasing probability of reactivity as older exposed blocks are incorporated into the analysis, as shown in Figure 7. 19, is closely linked to the reaction kinetics of AAR.

In the early years (i.e., up to 5 years), the reaction progresses more slowly, with reactivity primarily identified in highly reactive cases that exhibit large expansions. However, as data from blocks with longer exposure durations (i.e., 10 to 22 years) are added, the reaction has had more time to advance, making even cases with initially low expansion rates more evident. This cumulative effect allows the model to capture the entire spectrum of reactivity, including slower or delayed reactions that may not have manifested during shorter exposure periods. As a result, the inclusion of older blocks provides a more comprehensive view of the long-term progression of AAR, enabling the model to predict reactivity more reliably. Consequently, the model adjusts by lowering the thresholds for acceptable probabilities of reactivity. This enhanced sensitivity over time ensures that slow-reacting or delayed cases are detected, refining the model's predictive accuracy for field performance.

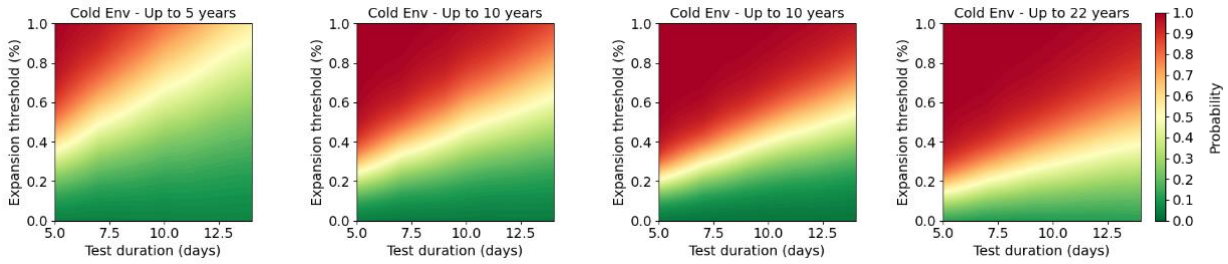


Figure 7. 19 – Probability of AAR occurrence based on AMBT results, under cold environment (i.e., TEMP=7C and RH=75%) and alkali content of 4.2 kg/m3, for (a) up to 5 years, (b) up to 10 years, (c) up to 15 years, and (d) up to 22 years.

For CPT, as seen in Figure 7. 20, the increasing probability of reactivity as older exposed blocks are incorporated into the analysis is similarly linked to the kinetics of the reaction. However, the trends are less distinct between 5 and 10 years. At the 10-year mark, there is a notable increase in the probability of reactivity at shorter test durations, suggesting that some reactivity were detected earlier or data from highly reactive aggregates were mostly incorporated in this stage.

Despite this variability between 5 and 10 years, the overall trend still refers to increasing probability of reactivity over time. As longer exposure durations (i.e., up to 22 years) are integrated, the model becomes more adept at capturing slower or delayed reactions. This results in a cumulative effect where the probability of detecting reactivity improves as the

reaction has more time to progress in the field, even for cases that initially exhibit lower expansion rates.

Similarly to AMBT, the inclusion of long-term data allows the model to provide a more reliable prediction of AAR, adjusting to capture a broader range of reactive cases over time. Consequently, the model increases its sensitivity, enhancing its ability to detect reactivity across varying test durations and field exposure periods. This ensures that the model continues to improve its predictive accuracy for field performance, even in cases where reactivity may develop more slowly.

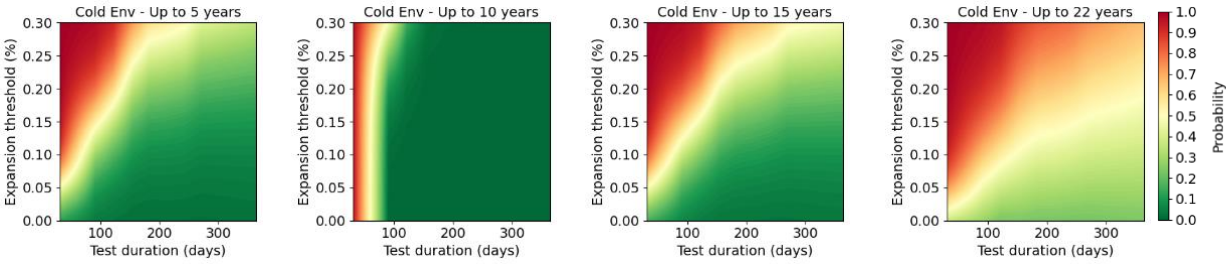


Figure 7. 20 – Probability of AAR occurrence based on CPT results, under cold environment (i.e., TEMP=7C and RH=75%) and alkali content of 4.2 kg/m3, for (a) up to 5 years, (b) up to 10 years, (c) up to 15 years, and (d) up to 22 years.

Given the critical findings from this analysis, it is evident that incorporating the probabilistic estimation of AAR occurrence into current structural evaluations could significantly enhance the accuracy and reliability of predicting long-term durability. By linking the probabilities of AAR occurrence to the structural safety levels of concrete elements, engineers can better assess the risk of potential damage and failure over time. This approach allows for more targeted mitigation strategies, optimizing the design and maintenance of concrete structures based on the specific environmental and material conditions they will face. Such integration could prove particularly valuable in high-risk areas or structures where the consequences of AAR-related degradation could compromise the integrity of critical infrastructure.

While this study provides valuable insights on AAR prediction, several limitations should be noted. First, the availability of data constrained the analysis, particularly regarding the environmental conditions, which were based mostly on two specific climates (i.e., Ottawa and Austin). This dataset may not fully capture the variability of global environmental

conditions and their influence on AAR. Second, the logistic regression model used in this analysis, while effective for identifying key trends, may oversimplify the complex interactions between variables. Future studies could benefit from incorporating more parameters into the modelling along with broader range of environmental and material data to enhance the predictive accuracy. Additionally, the model's reliance on specific test durations and conditions may limit its generalizability to other scenarios, necessitating further validation across different contexts. Despite these limitations, the findings offer a robust framework for understanding and mitigating AAR risks in concrete structures.

7.9 Conclusion

This study explored the application of machine learning techniques to propose a flexible coupled threshold-time (FCTT) approach, enabling a risk-based framework for mitigating AAR-induced damage under diverse exposure conditions in concrete structures. By integrating laboratory data with environmental variables, the analysis enhanced the predictability of field reactivity. The investigation led to the following conclusions:

- While the application of decision tree models offered valuable insights into how data can be classified based on test duration and expansion values, it also highlighted the complexities of establishing a single period for threshold determination. The results suggest that a more nuanced approach, considering the expansion overtime, environmental conditions and alkali content, may be necessary to accurately predict long-term AAR risk.
- In warmer environments, AAR was found to accelerate, which creates the need to lower expansion thresholds for the reaction to occur given test durations. Conversely, in colder environments, the reaction progresses more slowly, requiring higher expansion thresholds to reach the same probability of AAR. This indicates that environmental context must be factored into the establishment of accurate and reliable AAR thresholds, as these factors directly influence the reaction kinetics and, consequently, the long-term durability of concrete structures.

- The research confirmed that higher alkali content in concrete significantly elevates the risk of AAR, even at lower expansion thresholds. The study established that as alkali content increases, the expansion thresholds decrease, meaning that even small expansions can result in a high probability of AAR when alkali levels are elevated. This finding underscores the importance of strict control over alkali levels in concrete mixes, as proper management can help in setting more accurate thresholds that better predict AAR risk, ultimately improving the long-term performance and safety of concrete structures.

7.10 References

[1] M.G. Alexander, Alkali–aggregate reaction, in: Developments in the Formulation and Reinforcement of Concrete, Elsevier, 2019: pp. 87–113. https://doi.org/10.1016/B978-0-08-102616-8.00004-6.

[2] B. Fournier, M.-A. Bérubé, Alkali–aggregate reaction in concrete: a review of basic concepts and engineering implications, 27 (2000).

[3] P.J. Nixon, I. Sims, eds., RILEM Recommendations for the Prevention of Damage by Alkali-Aggregate Reactions in New Concrete Structures, Springer Netherlands, Dordrecht, 2016. https://doi.org/10.1007/978-94-017-7252-5.

[4] B. Fournier, V. Malhotra, Evaluation of Laboratory Test Methods for Alkali-Silica Reactivity, Cement, Concrete, Aggr. 21 (1999) 173. https://doi.org/10.1520/CCA10431J.

[5] J. Lindgård, Ö. Andiç-Çakır, I. Fernandes, T.F. Rønning, M.D.A. Thomas, Alkali–silica reactions (ASR): Literature review on parameters influencing laboratory performance testing, Cement and Concrete Research 42 (2012) 223–243. https://doi.org/10.1016/j.cemconres.2011.10.004.

[6] I. Sims, P. Nixon, RILEM Recommended Test Method AAR-0: Detection of Alkali-Reactivity Potential in Concrete—Outline guide to the use of RILEM methods in assessments of aggregates for potential alkali-reactivity, Mat. Struct. 36 (2003) 472–479. https://doi.org/10.1007/BF02481527.

[7] A. Leemann, M. Góra, B. Lothenbach, M. Heuberger, Alkali silica reaction in concrete - Revealing the expansion mechanism by surface force measurements, Cement and Concrete Research 176 (2024) 107392. https://doi.org/10.1016/j.cemconres.2023.107392.

[8] B. Fournier, J. Lindgård, B.J. Wigum, I. Borchers, Outdoor exposure site testing for preventing Alkali-Aggregate Reactivity in concrete – a review., MATEC Web Conf. 199 (2018) 03002. https://doi.org/10.1051/matecconf/201819903002.

[9] J. Lindgård, T. Østnor, B. Fournier, Ø. Lindgård, T. Danner, G. Plusquellec, K. De Weerdt, Determining alkali leaching during accelerated ASR performance testing and in field exposed cubes using cold water extraction (CWE) and μXRF, MATEC Web Conf. 199 (2018) 03004. https://doi.org/10.1051/matecconf/201819903004.

[10] J. Custodio, D. Costa, A.B. Ribeiro, A.S. Silva, Assessment of potential alkali-silica reactivity of aggregates for concrete, 4TH INTERNATIONAL CONFERENCE ON STRUCTURAL INTEGRITY (ICSI 2021) 37 (2022) 590–597. https://doi.org/10.1016/j.prostr.2022.01.127.

[11] ASTM C1260, Test Method for Potential Alkali Reactivity of Aggregates (Mortar-Bar Method), ASTM International, 2022. https://doi.org/10.1520/C1260-22.

[12] R.E. Oberholster, G. Davies, An accelerated method for testing the potential alkali reactivity of siliceous aggregates, Cement and Concrete Research 16 (1986) 181–189. https://doi.org/10.1016/0008-8846(86)90134-1.

[13] ASTM C1778, Guide for Reducing the Risk of Deleterious Alkali-Aggregate Reaction in Concrete, ASTM International, 2022. https://doi.org/10.1520/C1778-22.

[14] ASTM C1293, Test Method for Determination of Length Change of Concrete Due to Alkali-Silica Reaction, (2024). https://doi.org/10.1520/C1293_C1293M-23A.

[15] B. Fournier, J.H. Ideker, K.J. Folliard, M.D.A. Thomas, P.-C. Nkinamubanzi, R. Chevrier, Effect of environmental conditions on expansion in concrete due to alkali–silica reaction (ASR), Materials Characterization 60 (2009) 669–679. https://doi.org/10.1016/j.matchar.2008.12.018.

[16] M. Thomas, B. Fournier, K. Folliard, J. Ideker, M. Shehata, Test methods for evaluating preventive measures for controlling expansion due to alkali–silica reaction in concrete,

Cement and Concrete Research 36 (2006) 1842–1856. https://doi.org/10.1016/j.cemconres.2006.01.014.

[17] M.-A. Bérubé, B. Fournier, Canadian experience with testing for alkali-aggregate reactivity in concrete, Cement and Concrete Composites 15 (1993) 27–47. https://doi.org/10.1016/0958-9465(93)90037-A.

[18] S. Multon, M. Cyr, A. Sellier, P. Diederich, L. Petit, Effects of aggregate size and alkali content on ASR expansion, Cement and Concrete Research 40 (2010) 508–516. https://doi.org/10.1016/j.cemconres.2009.08.002.

[19] M. Berube, J. Frenette, A. Pedneault, M. Rivest, Laboratory assessment of the potential rate of ASR expansion of field concrete, CEMENT CONCRETE AND AGGREGATES 24 (2002) 13–19.

[20] C.A. MacDonald, The Kingston Exposure Site for ASR at 32 Years, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 540–549. https://doi.org/10.1007/978-3-031-59349-9_62.

[21] B.J. Wigum, G.J. Einarsson, ALKALI AGGREGATE REACTION IN ICELAND RESULTS FROM LABORATORY TESTING COMPARED TO FIELD EXPOSURE SITE, (n.d.).

[22] M. Thomas, A. Dunster, P. Nixon, B. Blackwell, Effect of fly ash on the expansion of concrete due to alkali-silica reaction – Exposure site studies, Cement and Concrete Composites 33 (2011) 359–367. https://doi.org/10.1016/j.cemconcomp.2010.11.006.

[23] B. Fournier, A. Bilodeau, N. Bouzoubaa, P.-C. Nkinamubanzi, Field and Laboratory Investigations on the Use of Fly Ash and Li-Based Admixtures to Prevent ASR in Concrete, in: United Kingdom, 2018.

[24] J. Lindgård, P.J. Nixon, I. Borchers, B. Schouenborg, B.J. Wigum, M. Haugen, U. Åkesson, The EU "PARTNER" Project — European standard tests to prevent alkali reactions in aggregates: Final results and recommendations, Cement and Concrete Research 40 (2010) 611–635. https://doi.org/10.1016/j.cemconres.2009.09.004.

[25] T. Drimalas, K.J. Folliard, J.H. Ideker, Findings from the University of Texas at Austin ASR Exposure Site after 20 Years, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings

of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 516–523. https://doi.org/10.1007/978-3-031-59349-9_59.

[26] A.H.-S. Ang, W.H. Tang, A.H.-S. Ang, Probability concepts in engineering: emphasis on applications in civil & environmental engineering, 2nd ed, Wiley, New York, 2007.

[27] K.M. Ramachandran, C.P. Tsokos, Descriptive Statistics, in: Mathematical Statistics with Applications in R, Elsevier, 2015: pp. 1–52. https://doi.org/10.1016/B978-0-12-417113-8.00001-1.

[28] O.C. Ibe, Introduction to Descriptive Statistics, in: Fundamentals of Applied Probability and Random Processes, Elsevier, 2014: pp. 253–274. https://doi.org/10.1016/B978-0-12-800852-2.00008-0.

[29] B.R. Martin, Statistics, Experiments, and Data, in: Statistics for Physical Science, Elsevier, 2012: pp. 1–20. https://doi.org/10.1016/B978-0-12-387760-4.00001-9.

[30] T.M. Mitchell, Machine Learning, McGraw-Hill, New York, 1997.

[31] K. Yeturu, Machine learning algorithms, applications, and practices in data science, in: Handbook of Statistics, Elsevier, 2020: pp. 81–206. https://doi.org/10.1016/bs.host.2020.01.002.

[32] H. Belyadi, A. Haghighat, Supervised learning, in: Machine Learning Guide for Oil and Gas Using Python, Elsevier, 2021: pp. 169–295. https://doi.org/10.1016/B978-0-12-821929-4.00004-4.

[33] A. Bergmann, L.F.M. Sanchez, Bibliometric analysis and database compilation of laboratory test procedures for assessing concrete field performance against alkali-aggregate reaction (AAR), Unpublished Manuscript (n.d.).

[34] A. Bergmann, L.F.M. Sanchez, Multifactorial analysis of AAR development: Integrating laboratory and field data with statistical and probabilistic modelling, Unpublished Manuscript (n.d.).

[35] T. Drimalas, K.J. Folliard, A. Parashar, M.D.A. Thomas, A. Ghanizadeh, A.M. Hossack, B. Fournier, National Cooperative Highway Research Program, Transportation Research Board, National Academies of Sciences, Engineering, and Medicine, Alkali-Silica Reactivity Potential and Mitigation: Test Methods and State of

Practice, Transportation Research Board, Washington, D.C., 2023. https://doi.org/10.17226/27435.

[36] J. Lindgård, T.F. Rønning, M.D.A. Thomas, B. Fournier, A.S. Silva, ASR - PERFORMANCE TESTING: MAIN FINDINGS IN THE NORWEGIAN COIN PROJECT, in: Sao Paulo – Brazil, 2016.

[37] B. Fournier, A. Ferro, V. Sivasundaram, CANMET/Industry Research Consortium on Alkali-Aggregate Reactivity in Concrete, EPRI and CANMET, Ontario, Canada, 2001. https://www.epri.com/research/products/1004031.

[38] J. Custódio, J. Lindgård, B. Fournier, A. Santos Silva, M.D.A. Thomas, T. Drimalas, J.H. Ideker, R.-P. Martin, I. Borchers, B. Johannes Wigum, T.F. Rønning, Correlating field and laboratory investigations for preventing ASR in concrete – The LNEC cube study (Part I – Project plan and laboratory results), Construction and Building Materials 343 (2022) 128131. https://doi.org/10.1016/j.conbuildmat.2022.128131.

[39] C.A. MacDonald, Kingston Outdoor Exposure Site for ASR –29 Year Update, (2020).

[40] I. Fernandes, A. Leemann, B. Fournier, E. Menendez, J. Lindgård, I. Borchers, J. Custódio, PARTNER project post-documentation study. Condition assessment of field exposure site cubes. Results of microstructural analyses, Cement and Concrete Research 162 (2022) 107006. https://doi.org/10.1016/j.cemconres.2022.107006.

[41] K.J. Folliard, R. Barborak, T. Drimalas, L. Du, S. Garber, J. Ideker, T. Ley, S. Williams, M. Juenger, B. Fournier, M.D.A. Thomas, Preventing ASR/DEF in New Concrete: Final Report (0-4085-5), Texas Department of Transportation and the Federal Highway Administration, 2006.

[42] B.J. Wigum, L.T. Pedersen, B. Grelk, J. Lindgård, State-of-the art report: Key parameters influencing the alkali aggregate reaction, SINTEF Building and Infrastructure, 2006. https://www.sintef.no/globalassets/upload/byggforsk/partner/report-2.1-final-a06018.pdf.

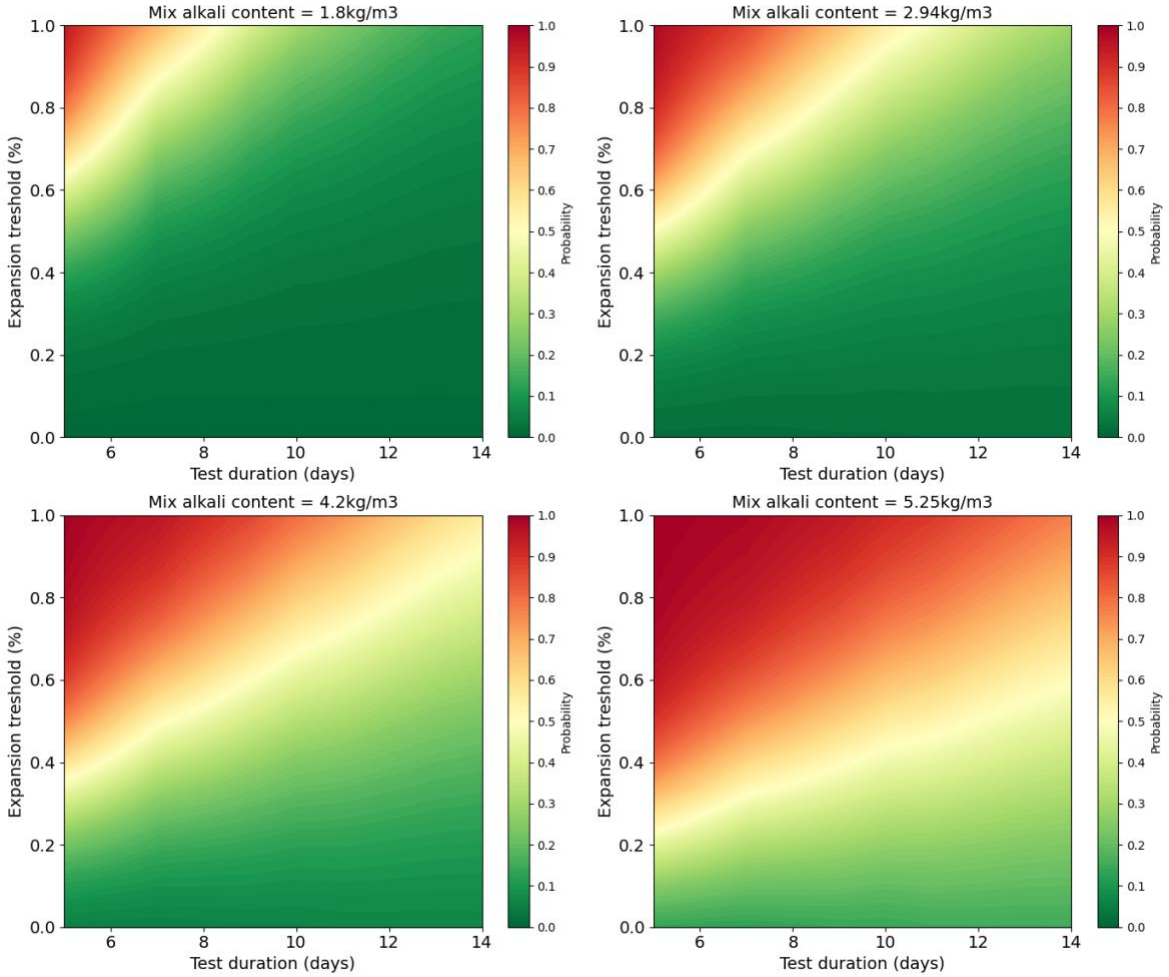
[43] A. Bergmann, L. Sanchez, Assessing the Variability of Laboratory Test Procedures for Predicting Concrete Field Performance Against Alkali-Aggregate Reaction (AAR), in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference

on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 507–515. https://doi.org/10.1007/978-3-031-59349-9_58.

[44] A. Bergmann, L.F.M. Sanchez, Assessing the variability of laboratory test procedures for predicting concrete field performance against alkali aggregate reaction (AAR), Manuscript Submitted for Publication (n.d.).

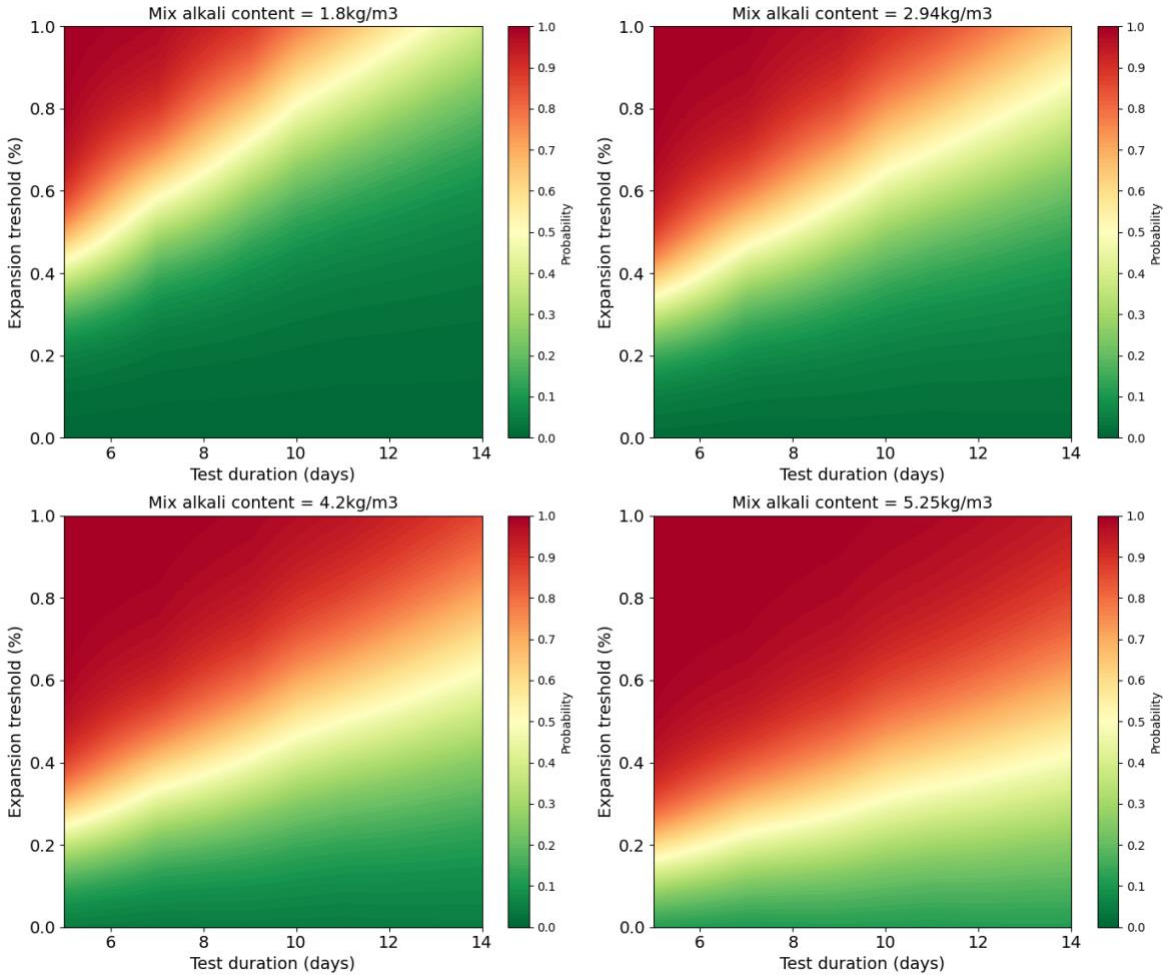
7.11 Appendix A - Logistic regression coefficients and threshold analysis for AMBT including mixes without SCMs, based on field results up to 5 years (Data availability: 141 rows).

Age (days)	β_0	β_{TEMP}	β_{RH}	β_{EXP}	β_{ALK}	Accuracy
5	0.040	0.059	-0.096	7.948	0.932	0.72 (29)
7	-0.257	0.061	-0.093	5.889	0.933	0.72 (29)
9	-0.321	0.058	-0.090	4.617	0.925	0.76 (29)
10	0.008	0.054	-0.094	4.155	0.923	0.76 (29)
11	-0.057	0.053	-0.092	3.778	0.921	0.76 (29)
12	-0.116	0.053	-0.091	3.438	0.920	0.76 (29)
13	-0.020	0.051	-0.092	3.161	0.918	0.76 (29)
14	0.140	0.049	-0.093	2.927	0.906	0.76 (29)



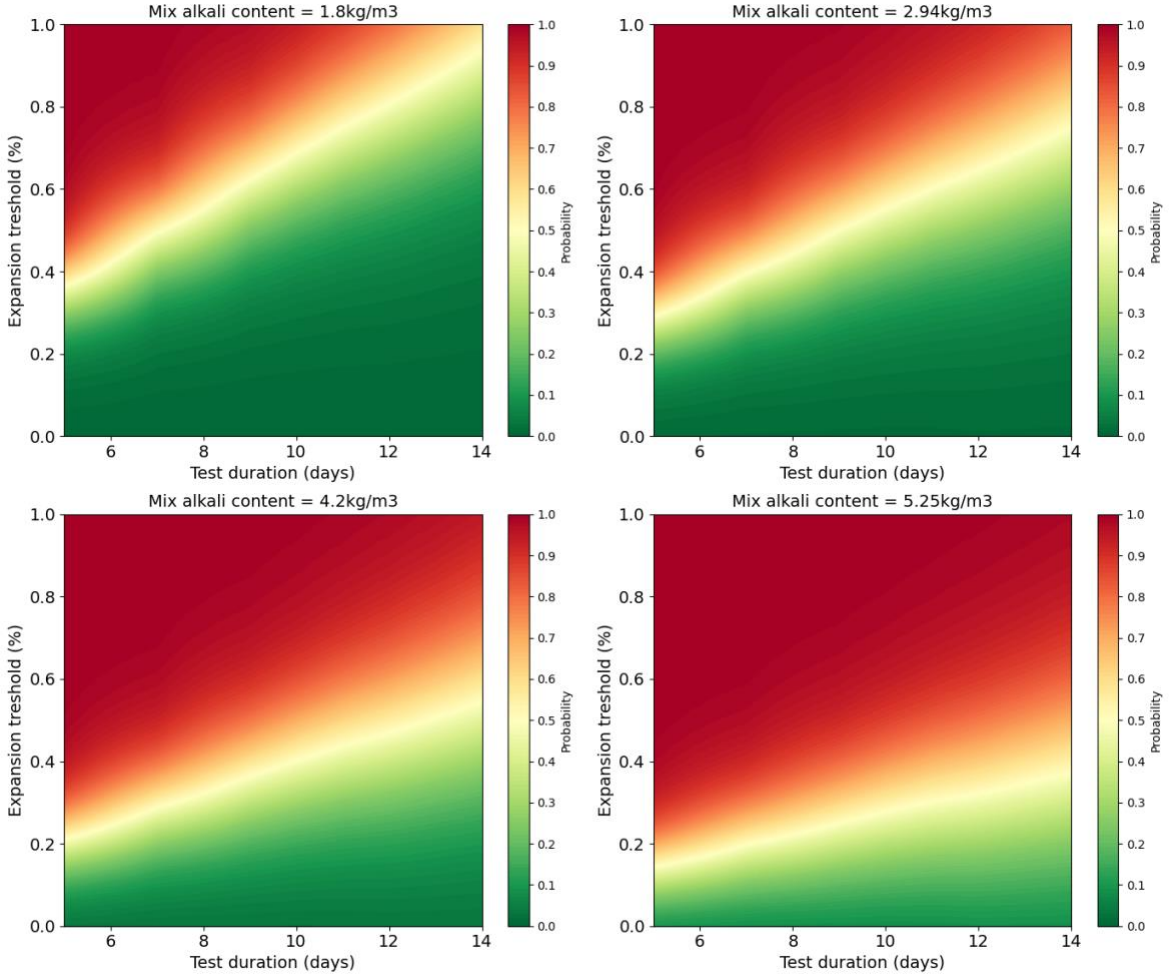
7.12 Appendix B - Logistic regression coefficients and threshold analysis for AMBT including mixes without SCMs, based on field results up to 10 years (Data availability: 155 rows).

Age (days)	β_0	β_{TEMP}	β_{RH}	β_{EXP}	β_{ALK}	Accuracy
5	0.153	0.045	-0.098	12.113	0.929	0.84 (31)
7	0.092	0.046	-0.100	9.233	0.945	0.81 (31)
9	0.262	0.041	-0.101	7.355	0.947	0.81 (31)
10	-0.811	0.053	-0.088	6.563	0.946	0.81 (31)
11	-0.048	0.043	-0.098	6.141	0.952	0.81 (31)
12	0.041	0.041	-0.098	5.620	0.947	0.81 (31)
13	-0.149	0.042	-0.095	5.186	0.941	0.81 (31)
14	-0.337	0.042	-0.092	4.800	0.936	0.81 (31)



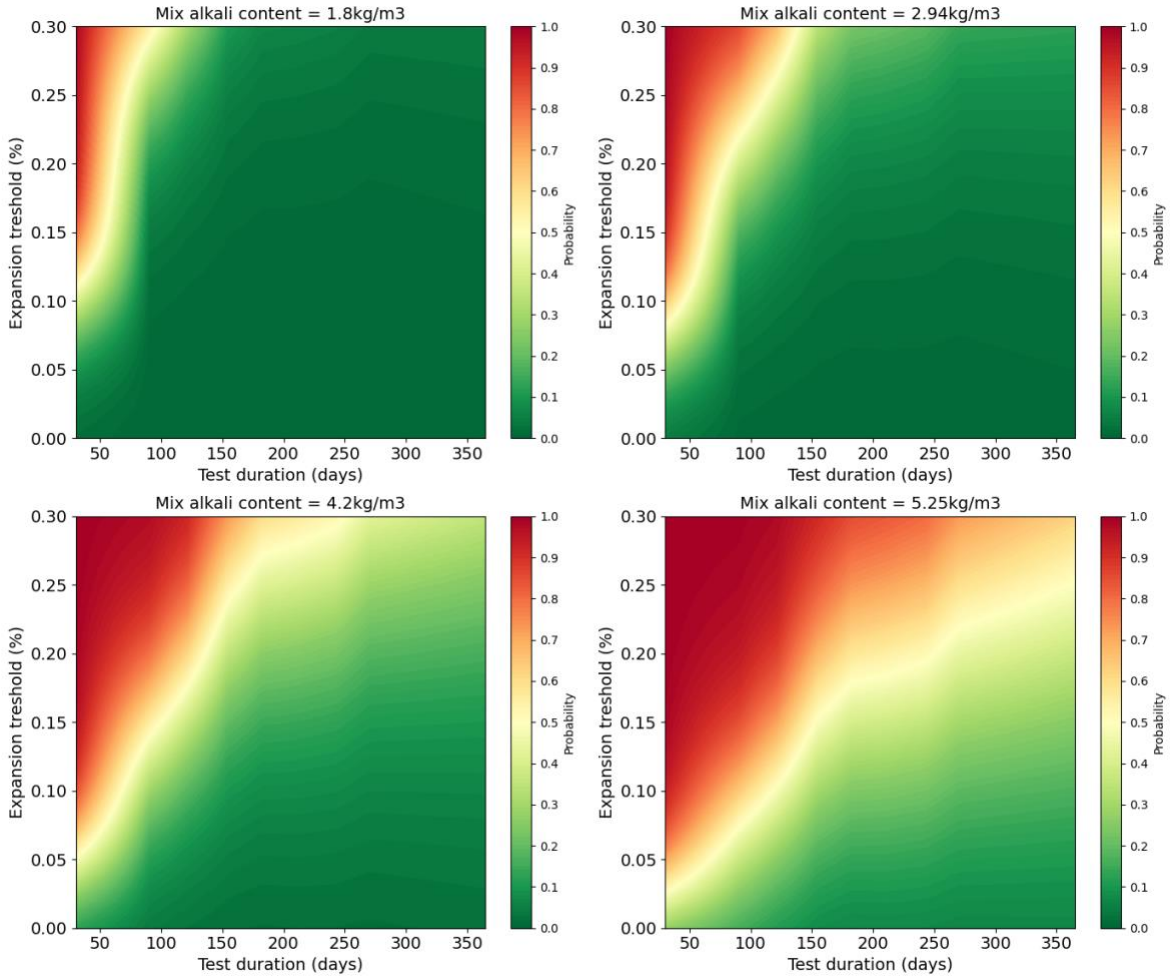
7.13 Appendix – Logistic regression coefficients and threshold analysis for AMBT including mixes without SCMs, based on field results up to 15 years (Data availability: 157 rows).

Age (days)	β_0	β_{TEMP}	β_{RH}	β_{EXP}	β_{ALK}	Accuracy
5	0.171	0.044	-0.104	15.269	0.975	0.72 (32)
7	-0.054	0.047	-0.106	11.813	1.014	0.72 (32)
9	-0.066	0.045	-0.107	9.473	1.027	0.72 (32)
10	0.003	0.044	-0.108	8.618	1.025	0.72 (32)
11	0.198	0.041	-0.110	7.879	1.023	0.72 (32)
12	0.611	0.034	-0.114	7.242	1.022	0.72 (32)
13	0.000	0.042	-0.107	6.748	1.026	0.72 (32)
14	-0.046	0.041	-0.107	6.305	1.033	0.72 (32)



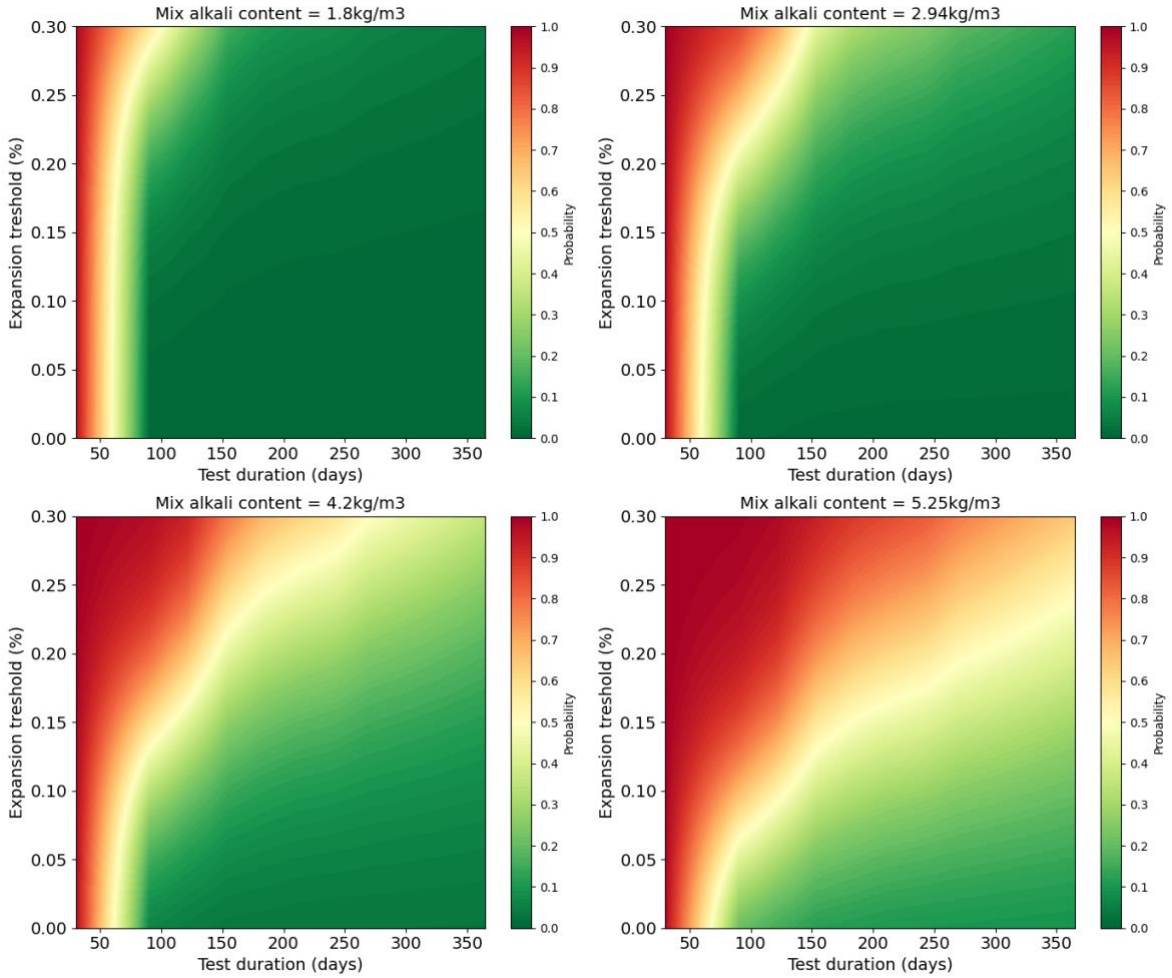
7.14 Appendix D – Logistic regression coefficients and threshold analysis for CPT including mixes without SCMs, based on field results up to 5 years (Data availability: 232 entries).

Age (days)	β_0	β_{TEMP}	β_{RH}	β_{EXP}	β_{ALK}	Accuracy
28	-5.154	0.095	-0.015	37.066	0.891	0.66 (47)
90	-3.378	0.071	-0.082	22.295	1.406	0.77 (47)
121	-2.960	0.080	-0.091	19.702	1.368	0.74 (47)
154	-3.134	0.097	-0.091	15.998	1.302	0.72 (47)
182	-3.214	0.103	-0.091	14.432	1.282	0.72 (47)
210	-3.467	0.104	-0.086	13.835	1.263	0.72 (47)
243	-3.570	0.104	-0.084	13.291	1.249	0.70 (47)
270	-4.318	0.109	-0.074	12.225	1.223	0.70 (47)
365	-4.201	0.112	-0.064	10.545	1.058	0.74 (47)



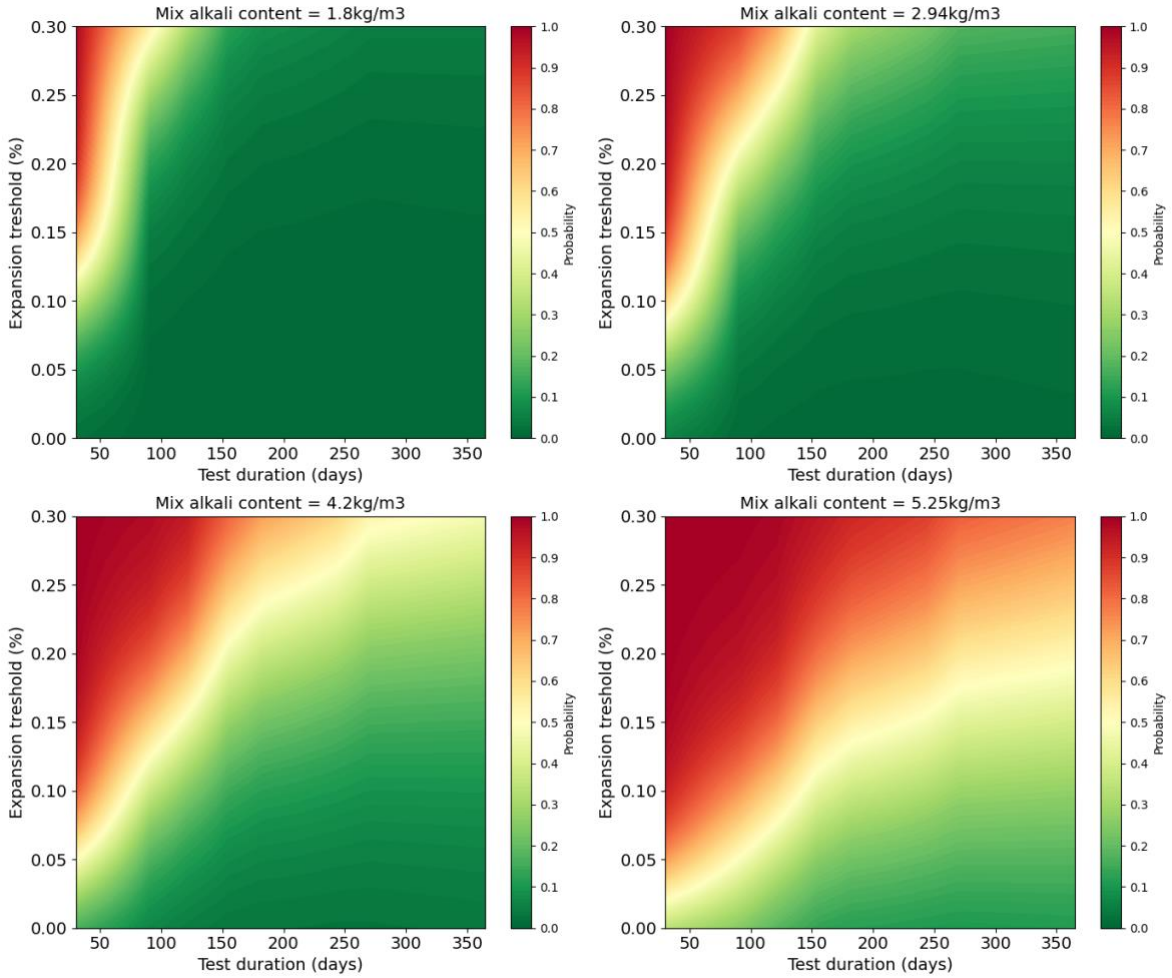
7.15 Appendix E – Logistic regression coefficients and threshold analysis for CPT including mixes without SCMs, based on field results up to 10 years (Data availability: 246 entries).

Age (days)	β_0	β_{TEMP}	β_{RH}	β_{EXP}	β_{ALK}	Accuracy
28	-2.463	0.081	0.081	33.764	0.843	0.72 (50)
90	0.163	0.040	-0.116	20.638	1.317	0.74 (50)
121	1.268	0.038	-0.133	18.455	1.286	0.70 (50)
154	1.861	0.045	-0.142	15.214	1.234	0.70 (50)
182	1.952	0.048	-0.143	13.707	1.213	0.68 (50)
210	1.683	0.050	-0.140	13.152	1.202	0.68 (50)
243	1.263	0.054	-0.134	12.560	1.182	0.68 (50)
270	0.784	0.058	-0.127	11.535	1.163	0.68 (50)
365	0.177	0.073	-0.119	9.689	1.128	0.71 (49)



7.16 Appendix F – Logistic regression coefficients and threshold analysis for CPT including mixes without SCMs, based on field results up to 15 years (Data availability: 246 entries).

Age (days)	β_0	β_{TEMP}	β_{RH}	β_{EXP}	β_{ALK}	Accuracy
28	-1.568	0.064	-0.057	32.545	0.907	0.70 (50)
90	1.477	0.012	-0.138	21.136	1.466	0.68 (50)
121	2.593	0.011	-0.158	19.259	1.465	0.64 (50)
154	3.057	0.022	-0.167	16.067	1.420	0.66 (50)
182	3.170	0.026	-0.169	14.633	1.397	0.66 (50)
210	2.663	0.032	-0.162	13.995	1.379	0.64 (50)
243	2.401	0.034	-0.159	13.397	1.370	0.64 (50)
270	1.948	0.039	-0.151	12.168	1.327	0.62 (50)
365	2.509	0.024	-0.150	10.882	1.237	0.66 (50)



Chapter 8: Framework for risk assessment of mixture designs incorporating alkali-silica reactive aggregates

Ana Bergmann and Leandro Sanchez

Abstract: Alkali-aggregate reaction (AAR), a harmful durability mechanism affects concrete structures worldwide. Over the years, it has become clear that prevention is the best strategy against AAR. In this sense, guidelines have been developed relying on accelerated laboratory tests, including the Accelerated Mortar Bar Test (AMBT) and Concrete Prism Test (CPT), to define mitigation strategies and plans. However, current thresholds frequently fail to predict the actual field behavior of concrete over extended periods. In this sense, this study introduces a probabilistic framework integrating laboratory and field data with logistic regression to predict ASR risk, a flexible coupled threshold-time (FCTT) approach. The framework incorporates environmental conditions, alkali content, and structural classification to define the risk of AAR field occurrence. It allows for dynamic risk assessments and informed decisions on mitigation strategies based on a structure's specific information, improving alignment between laboratory outcomes and real-world durability.

Keywords: Alkali-aggregate reaction (AAR); Concrete durability; Probabilistic risk assessment

8.1 Introduction

Alkali-aggregate reaction (AAR), a harmful durability mechanism affecting concrete structures worldwide, impacts the integrity and serviceability of structures in over 50 countries [1–7]. This reaction, predominantly alkali-silica reaction (ASR), involves the interaction between siliceous phases in the aggregates and the alkali hydroxides present in the concrete pore solution; and originates a secondary product (ASR-gel) that swells upon water uptake, leading to internal stresses and cracking, and reducing the mechanical properties of affected concrete [8,9].

Over the years, it became quite evident that the best approach against AAR is its prevention; in this context, a wide number of laboratory test procedures were developed to appraise the

potential reactivity of aggregates and the efficiency of preventive measures prior to concrete pouring. The accelerated mortar bar test (AMBT) and concrete prism test (CPT) are widely employed and essential for assessing aggregate's reactivity and the effectiveness of preventative measures [8,10]. These methods classify aggregates based on their reactivity levels, ranging from non-reactive to very highly reactive [10–12]. However, a critical challenge remains since the current established thresholds often fail to accurately predict long-term field performance [8,10]. This has led to significant discrepancies between laboratory test outcomes and the actual performance of concrete in field conditions [13–19]. Given the difficulty of remediating structures once AAR is triggered, prevention has emerged as the most effective strategy. Over the years, various laboratory test methods have been developed to evaluate both the reactivity of aggregates and the efficacy of preventive measures prior to concrete pouring. The Accelerated Mortar Bar Test (AMBT) and the Concrete Prism Test (CPT) are two of the most widely adopted methods for assessing the potential for ASR [8,10]. The AMBT is performed by immersing mortar bars (i.e., 25x25x285 mm^3) in a high-alkali solution (1M NaOH) at 80°C, with expansions measured over 14 to 28 days for mixtures without and with SCMs, respectively [20,21]. This accelerated method yields rapid results and is often recommended for preliminary screenings of aggregate reactivity. In contrast, the CPT employs concrete prisms (i.e., 75x75x250 mm^3) and monitors expansion over a longer duration, typically 1 to 2 years for mixtures without and with SCMs, under conditions of moderate temperature (i.e., 38°C) and high relative humidity (i.e., 95%) [22]. These tests classify aggregates according to their reactivity levels, ranging from non-reactive to very highly reactive [10–12]. However, despite their widespread use, a critical challenge has been the discrepancy between laboratory test results and long-term field performance [8,10]. For instance, studies have shown that current thresholds used in AMBT and CPT tests frequently fail to predict the actual field behavior of concrete over extended periods [2,6,12–18].

In response to this issue, recent research has focused on integrating field data with laboratory results to better assess the risk of ASR in real-world conditions [23–26]. Discrepancies remain particularly significant when samples are submitted to varying environments (i.e., cold vs warm), alkali content (i.e., low, moderate, high) and the presence of supplementary

cementitious materials (SCMs). Although AMBT outperforms in identifying non-reactive cases for both mixes with and without SCMs, both AMBT and CPT demonstrate similar reliability for detecting reactive cases, with CPT slightly better at higher alkali levels [25]. However, the reliability of both methods diminishes in warm climates with high alkali levels and no SCMs, while performance improves in colder climates with low alkali levels and the presence of SCMs. Therefore, it underscores the need for a more dynamic assessment approach, considering laboratory expansion levels, test duration, environmental factors, and alkali content to better predict AAR occurrence [26].

In this sense, this study aims to propose a comprehensive framework for assessing the risk of ASR in concrete mixtures by integrating laboratory and field data with probabilistic methods, such as logistic regression. The framework also aims to include the structure class to limit potential risks associated with the reaction, ensuring that critical infrastructure is given greater consideration. This approach will allow for more reliable and assertive AAR risk assessments, helping to align laboratory predictions with field outcomes and improve the long-term durability of concrete structures exposed to ASR.

8.2 ASR risk assessment in concrete mixtures

Guiding documents have been proposed on how to interpret the results of ASR test methods or suggest steps to reduce the risk of ASR [27]. In North America, the current strategies for mitigating ASR in concrete structures are outlined in the ASTM C1778, AASHTO R 80, and CSA A23.2-27A standards [12,28–30]. In general, the standards offer two pathways: performance-based and prescriptive approaches [27].

In the performance-based approach, laboratory tests, such as the CPT, are employed to assess the effectiveness of SCMs in mitigating ASR [31]. Therefore, The SCM replacement level is tested and must successfully reduce expansion below the established threshold to be considered effective. Conversely, in the prescriptive approach, aggregates are first classified through AMBT or CPT tests. Then, the exposure conditions of the concrete structure and its structural classification are evaluated to determine the necessary preventive measures, such as the required SCM replacement level or alkali content limits [28].

CSA A23.2-27A provides a detailed flowchart for determining aggregate reactivity and selecting appropriate preventive measures [29]. It evaluates the petrographic composition of the aggregates and includes a brief geological field examination, considering factors such as rock type, age, formation, and quarry location, alongside durability tests. When reviewing past performance history in Portland cement (PC) concrete, CSA recommends assessing structures that are at least 10 years old and have similar mixture proportions and exposure conditions as the intended project. The standard sets a fixed 0.04% expansion limit for the CPT at 1 year for mixes without SCMs and at 2 years for mixes with SCMs, while the AMBT expansion limit is 0.15% at 14 days for mixes with SCMs and 28 days for mixes without SCMs. If preventive measures are required, the size of the structure and its environmental conditions are considered to assess the ASR risk level, determining the necessary prevention strategy.

Similarly, ASTM C1778 provides guidelines for interpreting test results and selecting mitigation strategies to prevent ASR [12]. It incorporates a flowchart and considers field history for aggregate approval, alongside the concept of maximum alkali loading for determining the level of prevention. The expansion thresholds are set at 0.04% for CPT over 1-2 years and 0.10% for AMBT over 14-28 days, depending on whether the mix includes SCMs or not. Lastly, AASHTO R80-17 is closely aligned with ASTM C1778 [28,30], but a key distinction is its guidance on testing the efficacy of lithium nitrate to prevent ASR, specifically within the AMBT, which is not covered in ASTM C1778.

Regarding the level of ASR risk, the standards take into account three key factors: the size of the structure, its exposure conditions, and the reactivity of the aggregate. These factors help categorize the risk of ASR into six levels, as shown in Table 8. 1. Based on this risk level, the structure's classification, outlined in Table 8. 2, is used to determine the required level of prevention, which is further defined by the maximum alkali loading in Table 8. 3. The consequences of ASR damage are typically evaluated in terms of safety, economic costs, and environmental impact. For instance, if a massive concrete element is exposed to humid air, buried, or immersed and uses R1 (i.e., moderately reactive aggregate), Table 8. 1 categorizes this scenario as Risk Level 3. Next, if the structure is a Class SC4 type, such as a nuclear facility or dam, where even minor ASR damage cannot be tolerated, Table 8.

3 indicates that the appropriate level of prevention is Y, which would require a reduction in alkali loading to 1.8 kg/m³.

Table 8. 1 – Level of ASR Risk [12].

Size and Exposure conditions	Aggregate-Reactivity Class			
	R0	**R1**	**R2**	**R3**
Non-massive[A] concrete in a dry[B] environment	Level 1	Level 1	Level 2	Level 3
Massive[A] concrete in a dry[B] environment	Level 1	Level 2	Level 3	Level 4
All concrete exposed to humid air, buried or immersed	Level 1	Level 3	Level 4	Level 5
All concrete exposed to alkalis in service[C]	Level 1	Level 4	Level 5	Level 6

[A]A massive element has at least dimensions greater than 0.9m [3ft].
[B]A dry environment corresponds to an average ambient relative humidity lower than 60%, normally only found in the interior of buildings.
[C]Examples of structures exposed to alkalis in service include marine structures exposed to seawater and highway structures exposed to de-icing salts (for example, NaCl) or anti-icing salts (for example, potassium acetate, sodium formate, and so forth).

Table 8. 2 – Structures classified based on the severity of consequences should ASR[A] occur [12].

Class	Consequence of ASR	Acceptability of ASR	Examples[B]
Class SC1	Safety, economic, or environmental consequences small or negligible	Some deterioration from ASR may be tolerated	Non-load-bearing elements inside buildings Concrete elements no exposed to moisture Temporary structures (service life < 5 years)
Class SC2	Safety, economic, or environmental consequences if major deterioration	Moderate risk of ASR is acceptable	Sidewalks, curbs, and gutters Elements with service life <40 years
Class SC3	Safety, economic, or environmental consequences if minor deterioration	Minor risk of ASR may be acceptable	Pavements Foundation elements Retaining walls Culverts Highway barriers Rural, low-volume roads Precast elements in which economic costs of replacement are severe Service life normally 40 to 74 years
Class SC4	Serious safety, economic, or environmental consequences if minor damage	ASR cannot be tolerated	Major bridges Power plants Dams Nuclear facilities Water treatment facilities Waste water treatment facilities Tunnels Critical elements that are very difficult to inspect of repair Service life normally ≥ 75 years

[A]This table does not consider the consequences of damage as a result of ACR. This protocol does not permit the use of alkali-carbonate aggregates.
[B]The types of structures listed under each class are meant to serve as examples. Some owners may decide to use their own classification system. For example, sidewalks, curbs and gutters may be placed in the SC3 class in some jurisdiction.

Table 8. 3 – Level of prevention and respective maximum alkali loading. Adapted from [12].

Level of ASR Risk	Classification of Structure			
	Class SC1	**Class SC2**	**Class SC3**	**Class SC4**
Risk Level 1	V (No limit)	V (No limit)	V (No limit)	V (No limit)
Risk Level 2	V (No limit)	V (No limit)	W (3 kg/m^3)	X (2.4 kg/m^3)
Risk Level 3	V (No limit)	W (3 kg/m^3)	X (2.4 kg/m^3)	Y (1.8 kg/m^3)
Risk Level 4	W (3 kg/m^3)	X (2.4 kg/m^3)	Y (1.8 kg/m^3)	Z[B]
Risk Level 5	X (2.4 kg/m^3)	Y (1.8 kg/m^3)	Z[B]	ZZ[B]
Risk Level 6	Y (1.8 kg/m^3)	Z[B]	ZZ[B]	[A]

[A]It may not be permitted to construct a Class SC4 structure if the risk of ASR is Level 6. Measures should be taken to reduce the level of risk in these circumstances.
[B]SCMs may be used in prevention levels Z and ZZ

While this approach offers a detailed framework for developing mitigation strategies, it overlooks the influence of varying environmental conditions (i.e., warm vs cold), which are known to affect the reaction kinetics of ASR. Factors such as temperature and relative humidity, which can significantly accelerate or slow down the reaction, are not fully considered. Therefore, there is an opportunity to refine this approach by incorporating these environmental variables alongside laboratory test results, alkali content, and specific structural classification. This could lead to the development of a risk-based system, where acceptable risk levels vary for different structural classifications, guiding the appropriate mitigation strategies and preventive measures for each class.

8.3 Model development

Predicting the occurrence of AAR in the field based on laboratory test outcomes is a complex task due to the numerous interacting variables and the variability in test durations. As previously defined [26], to address these challenges and better understand the impact of each variable on AAR, a logistic regression model has been employed. This model allows for probabilistic estimation of AAR occurrence in field conditions, taking into account environmental factors, mix design parameters, and the expansion over time of accelerated laboratory tests. The ultimate goal is to establish a fexible threshold-time approach that can be systematically applied to reduce the risk of AAR across diverse concrete structure scenarios. This model has been specifically designed to include variables that are most relevant to AAR development, such as environmental conditions (i.e., temperature (TEMP),

relative humidity (RH)), laboratory expansion (EXP), and field alkali content (ALK). The model is defined as per Equation 8.1 and the probability of reactivity can, therefore, be calculate as per Equation 8.2.

$$x=\beta_0+\beta_{TEMP}\cdot TEMP+\beta_{RH}\cdot RH+\beta_{EXP}\cdot EXP+\beta_{ALK}\cdot ALK \quad \text{(Equation 8.1)}$$

$$P(x)=\frac{1}{1+e^{-(\beta_0+\beta_{TEMP}\cdot TEMP+\beta_{RH}\cdot RH+\beta_{EXP}\cdot EXP+\beta_{ALK}\cdot ALK)}} \quad \text{(Equation 8.2)}$$

Based on a robust database correlating laboratory and field information [32], the binary outcome of field reactivity (i.e., YES or NO) was based on the last recorded expansion of field blocks, meaning an average of 8.3 and 6.2 years for AMBT and CPT, respectively. The outcomes of the logistic regression model are provided for both the AMBT and the CPT without the use of SCMs, as observed in Table 8. 4 and Table 8. 5.

Table 8. 4 – Overall coefficients for the AMBT regression model.

Age (days)	β_0	β_{TEMP}	β_{RH}	β_{EXP}	β_{ALK}	Accuracy
5	0.276	0.008	-0.085	13.156	0.983	0.77 (44)
7	0.012	0.014	-0.087	10.151	1.029	0.77 (44)
9	0.047	0.011	-0.087	8.013	1.024	0.75 (44)
10	0.127	0.008	-0.087	7.204	1.018	0.75 (44)
11	0.024	0.009	-0.085	6.520	1.009	0.75 (44)
12	-0.200	0.010	-0.082	5.992	1.008	0.75 (44)
13	-0.029	0.007	-0.083	5.478	0.997	0.75 (44)
14	0.065	0.006	-0.083	5.068	0.987	0.75 (44)

As observed, EXP emerges as the most significant predictor of AAR probability in AMBT, with consistently positive coefficients across all ages. In fact, EXP has a substantial impact, with coefficients starting at 13.156 at 5 days and decreasing to 5.068 by 14 days. This strong relationship indicates that the degree of expansion observed in laboratory conditions is a critical indicator of AAR potential in the field, with early higher expansions strongly correlating with higher AAR risk.

ALK consistently exhibits a positive correlation with AAR probability in the AMBT, underscoring the critical role of alkalis in promoting AAR. The ALK coefficient is relatively stable, ranging from 0.983 to 1.029. This stability suggests that alkali content has a sustained influence on AAR risk, regardless of the test duration. The consistent positive impact of ALK across all ages highlights the need for stringent control of alkali levels in concrete to mitigate AAR.

In the AMBT model, temperature exhibits a generally positive correlation with the likelihood of AAR occurrence. This aligns with the understanding that higher temperatures accelerate the chemical reactions associated with AAR, leading to increased expansion and cracking in concrete [17]. Finally, RH generally has a negative correlation also around zero with AAR probability. This trend suggests that lower humidity may have a minimal effect or even reduce the likelihood of AAR, given the combination of the other parameters.

Table 8. 5 – Overall Coefficients for the CPT regression model.

Age (days)	β_0	β_{TEMP}	β_{RH}	β_{EXP}	β_{ALK}	Accuracy
28	0.413	0.000	-0.053	30.014	0.736	0.78 (67)
90	0.420	-0.017	-0.071	15.654	0.975	0.78 (67)
121	-0.717	0.000	-0.064	13.921	1.033	0.79 (67)
154	1.857	-0.018	-0.089	10.408	0.904	0.79 (67)
182	1.860	-0.016	-0.089	9.308	0.887	0.79 (67)
210	1.480	-0.012	-0.084	8.887	0.883	0.79 (67)
243	1.259	-0.009	-0.081	8.466	0.870	0.78 (67)
270	1.064	-0.007	-0.078	7.713	0.859	0.78 (67)
365	1.009	0.002	-0.071	6.191	0.749	0.82 (67)

In the CPT model, laboratory expansion (EXP) plays a dominant role, with coefficients ranging from 30.014 at 28 days to 6.191 at 365 days. The strong positive correlation between EXP and AAR probability underscores the importance of laboratory expansion as a primary predictor of AAR, particularly due to the lower expansion ranges observed in CPT tests. Moreover, in the CPT model, the ALK coefficients are the second most influential parameters, remaining positive and ranging from 0.736 to 1.033. Although the impact of ALK diminishes slightly over longer durations, it remains a key factor in predicting AAR.

This slight reduction in ALK influence could be attributed to the longer exposure period in CPT, which allows for more significant alkali leaching. As alkalis are gradually leached out of the concrete, their availability for driving AAR decreases, thereby slightly reducing the overall impact of alkali content on AAR probability over time.

The TEMP coefficients in CPT are slightly negative across various ages, ranging from -0.017 to 0.000. These negative coefficients are relatively close to zero, indicating a minimal effect of temperature on AAR probability in this context. However, the CPT model consistently shows negative RH coefficients across all test ages, with slightly more significant values ranging from -0.089 to -0.053, including the environmental factor in the model.

Finally, the analysis of the logistic regression model for both AMBT and CPT data reveals that laboratory expansion and alkali content are the most influential variables in predicting AAR probability, with consistently strong positive correlations. However, although temperature and relative humidity show more complex and minimal effects, their effects can be captured when comparing different climates. Also, it is worth mentioning that the temperature and RH coefficients are associated with the data available mostly from two environments (i.e., Ottawa and Austin).

Therefore, a parametric analysis is further conducted to define a flexible coupled threshold-time approach that accounts for variations in alkali content and environmental conditions (i.e., temperature and RH). Two distinct environments are considered based on data availability: a colder environment (i.e., TEMP = 7°C and RH = 75%) representing Ottawa, ON, Canada, and a warmer environment (i.e., TEMP = 21°C and RH = 63%) representing Austin, Texas, US. As observed in Figure 8. 1, the probability of AAR occurrence is strongly influenced by the combination of test duration and alkali content, with the expansion threshold playing a critical role in defining the risk levels.

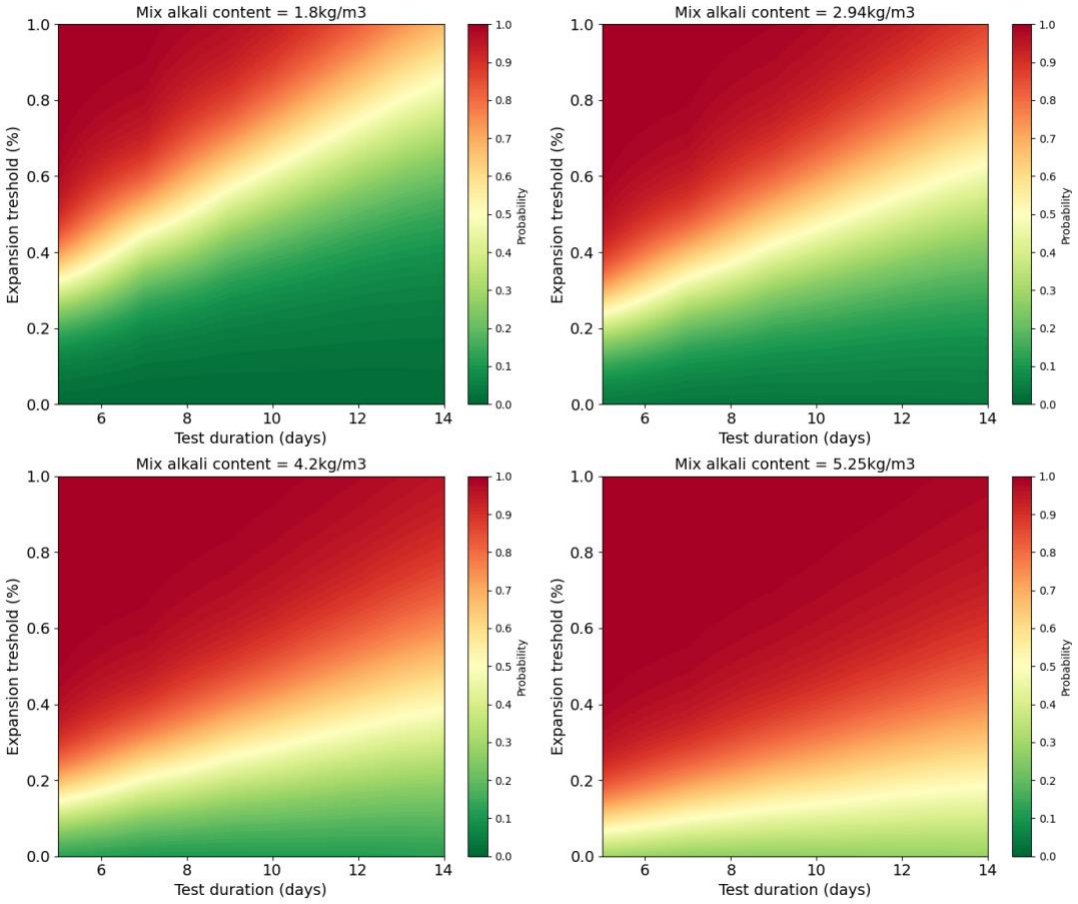


Figure 8. 1 – Probability of AAR occurrence based on AMBT results, under cold environment (i.e., TEMP=7C and RH=75%), for different alkali content: (a) 1.8 kg/m^3, (b) 2.94 kg/m^3, (c) 4.2 kg/m^3, and (d) 5.25 kg/m^3.

As shown in Figure 8. 2a, under cold conditions, the expansion threshold required to reach a high probability of AAR is generally higher over the test duration. This suggests that in colder environments, the AAR process may be slower to initiate, requiring more significant expansion in the accelerated laboratory test before AAR becomes a critical concern. However, when moving to a warmer environment as observed in Figure 8. 2b, the expansion threshold decreases, and the probability of AAR increases more rapidly over the same test duration. This indicates that warmer conditions accelerate the AAR process, making concrete more vulnerable to AAR at lower laboratory expansion levels.

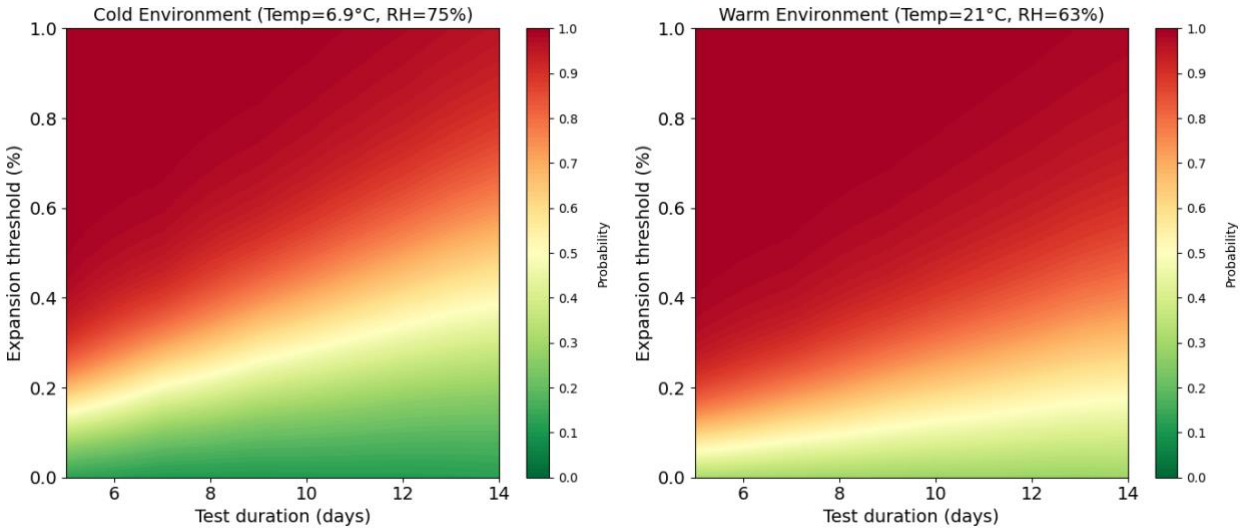


Figure 8. 2 – Probability of AAR occurrence based on AMBT results, under (a) cold environment (i.e., TEMP=7C and RH=75%), and (b) warm environment (i.e., TEMP=21C and RH=63%).

8.4 Proposed framework and risk limits

Figure 8. 3 presents the proposed framework, which consists of six essential steps. The process begins by establishing risk acceptance thresholds, where risk limits are defined according to the structural classification. Next, gather critical input data, including environmental factors such as temperature and relative humidity, as well as mix properties such as alkali content. In the following step, the probability of ASR occurrence is calculated using logistic regression to predict the likelihood of reactivity. Once the risk is computed, it is compared against the established acceptance limits to determine whether it falls within the acceptable risk range. If necessary, parameters are adjusted, and risk recalculated, with a focus on reducing alkali content or selecting aggregates with lower expansion potential (EXP). The final step is to make an informed decision on the appropriate mitigation strategies, which could involve modifying the mix design, incorporating SCMs, or implementing other preventive measures to effectively manage ASR risk.

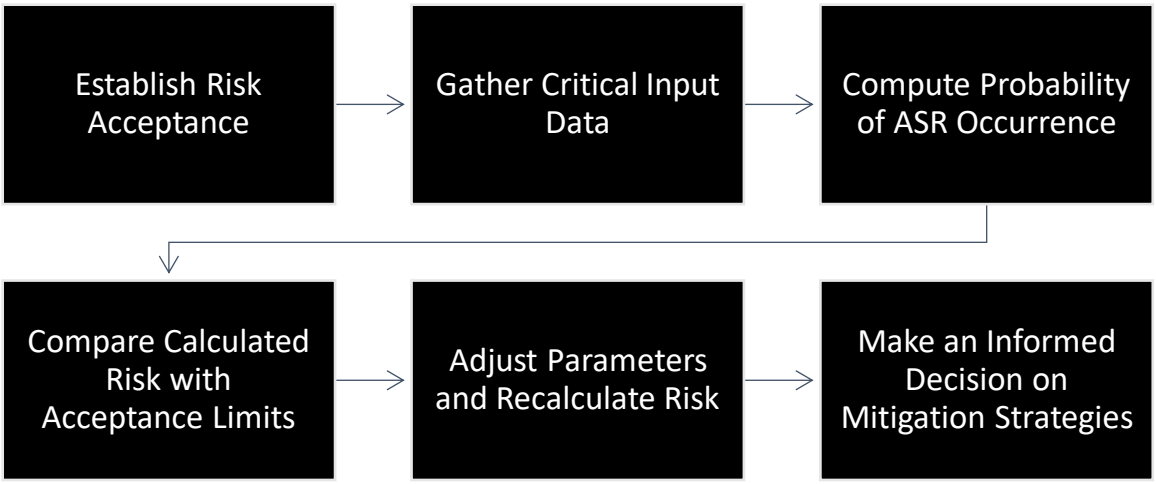


Figure 8. 3 – Proposed framework for ASR risk assessment.

The classification of ASR risk should be determined based on the structural classification and the probability of correctly identifying reactive outcomes, as outlined in Table 8.6. In this framework, less tolerance is allowed for structures with higher consequences of ASR damage, and stricter thresholds are applied to those classified as critical. For lower-risk structures, more tolerance for ASR may be acceptable, allowing for a higher probability of risk classification.

Table 8. 6 – Structural classification and risk levels based on the probability outcomes.

Structural Class	Consequence of ASR	Risk Level Acceptability	Very Low risk	Low risk	Moderate risk	High risk
Class SC1	Safety, economic, or environmental consequences small or negligible	Some deterioration from ASR may be tolerated	Probability < 40%	Probability 40-60%	Probability 60-80%	Probability >80%
Class SC2	Safety, economic, or environmental consequences if major deterioration	Moderate risk of ASR is acceptable	Probability < 30%	Probability 30-50%	Probability 50-70%	Probability >70%
Class SC3	Safety, economic, or environmental consequences if minor deterioration	Minor risk of ASR may be acceptable	Probability < 20%	Probability 20-40%	Probability 40-60%	Probability >60%
Class SC4	Serious safety, economic, or environmental consequences if minor damage	ASR cannot be tolerated	Probability < 10%	Probability 10-20%	Probability 20-40%	Probability >40%

In the following steps of calculating and comparing the probability of ASR occurrence, data gathering becomes a crucial aspect. For instance, Figure 8. 4 and Figure 8. 5 illustrate the assessment of ASR risk based on laboratory test results conducted under cold environmental conditions (i.e., 7°C and 75% relative humidity) with a fixed alkali content of 4.2 kg/m³. This highlights how the proposed framework integrates critical input data from environmental conditions and mix properties to predict the likelihood of ASR across different structural classifications. Such data is essential for accurately calculating the probability of occurrence and enables a detailed comparison against the suggested risk acceptance thresholds, ensuring that mitigation strategies are tailored to the specific structural and environmental context.

Figure 8. 4, which is based on AMBT results, shows how different structural classes respond to potential ASR development over a 14-day period. For Class SC1 structures (i.e., non-load-bearing elements), the risk remains low until the probability of occurrence exceeds 40%. In contrast, for Class SC4 (i.e., critical structures such as dams and nuclear facilities), even minor expansion presents a significant risk, emphasizing the need for stricter preventive

measures. This pattern illustrates how the framework dynamically adjusts the level of risk based on both the structural classification and the conditional evaluated parameters.

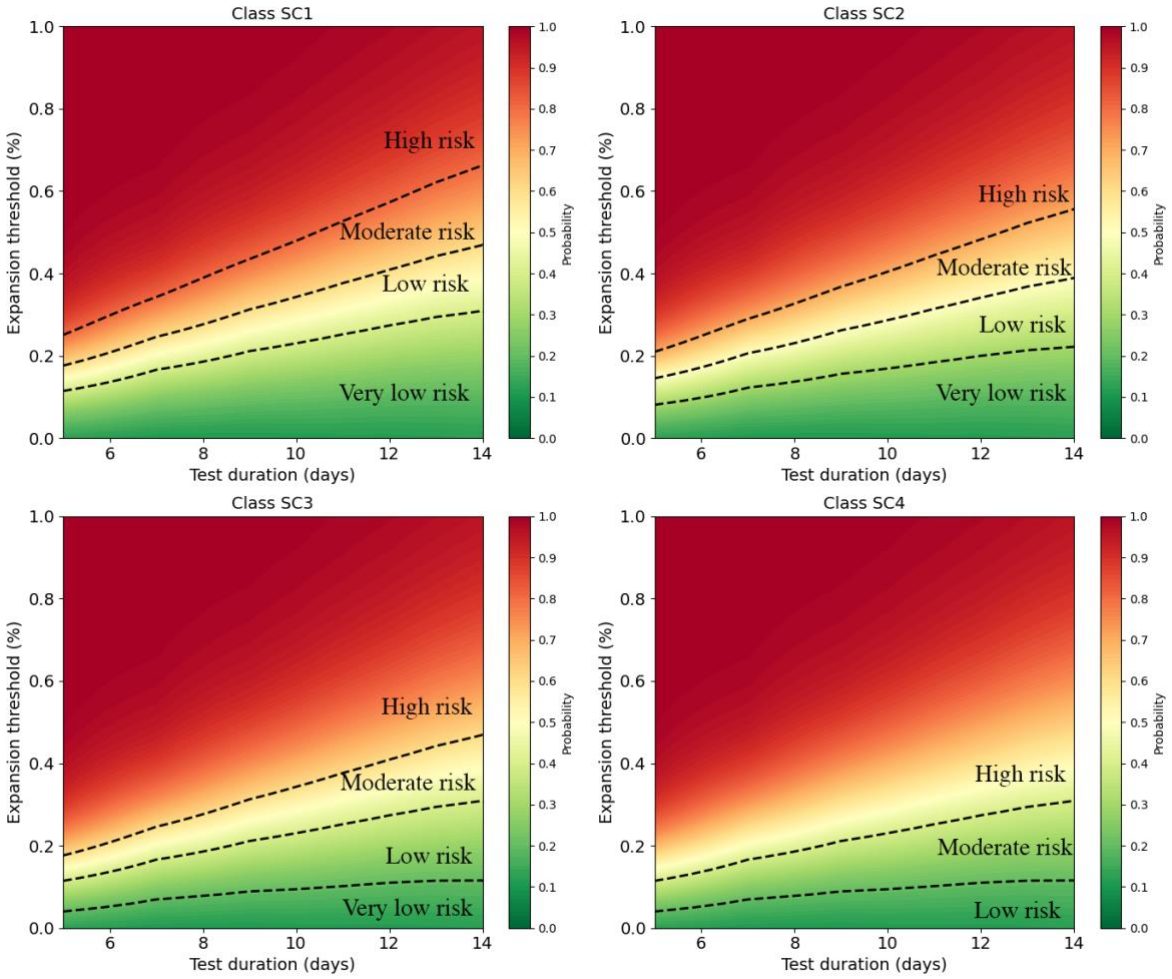


Figure 8. 4 – Risk of AAR occurrence based on AMBT results, under cold environment (i.e., TEMP=7C and RH=75%), 4.2 kg/m^3 alkali content for different structural classes: (a) SC1, (b) SC2, (c) SC3, and (d) SC4.

Similarly, Figure 8. 5, which is based on CPT results, illustrates how different structural classes respond to potential ASR development over a longer period. For Class SC1 structures (i.e., non-load-bearing elements), the risk of ASR remains low until the probability of occurrence exceeds 40%. In contrast, for Class SC4 structures (i.e., critical infrastructure like dams and nuclear facilities), even small expansion levels pose a significant risk much earlier in the test. This demonstrates how critical structures have much stricter tolerances,

underscoring the framework's ability to adjust risk levels dynamically based on both the test duration and the structural importance of the element being evaluated.

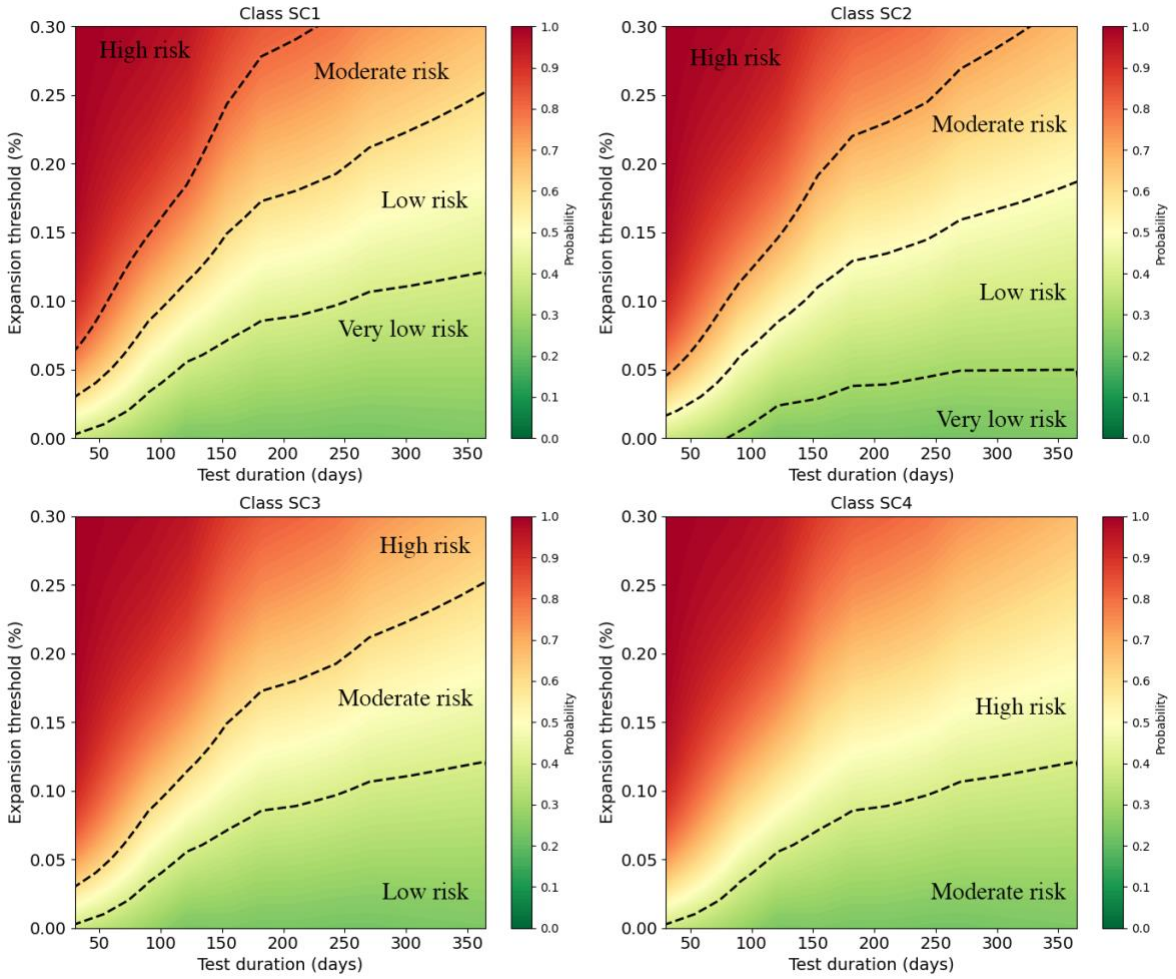


Figure 8. 5 – Risk of AAR occurrence based on CPT results, under cold environment (i.e., TEMP=7C and RH=75%), 4.2 kg/m^3 alkali content for different structural classes: (a) SC1, (b) SC2, (c) SC3, and (d) SC4.

This process of calculating and comparing risk levels against acceptable thresholds is key to the decision-making phase. For instance, if the calculated risk exceeds the acceptable level for a given structural class, parameters like alkali content or the choice of aggregates with lower expansion potential can be adjusted to mitigate the risk. Ultimately, the framework allows for informed decisions on mix design adjustments, the incorporation of SCMs, or other preventive measures to effectively manage ASR risk across various structural applications.

8.5 Conclusion

In summary, this study presents a comprehensive approach to assessing ASR risk in concrete structures by integrating laboratory and field data with probabilistic methods. The main findings are:

- A probabilistic framework is proposed, combining logistic regression with laboratory and field data to assess ASR risk in concrete mixtures.
- Environmental factors such as temperature, relative humidity, and alkali content are integrated with structural classification to provide dynamic risk assessments.
- Mitigation strategies are dependent on structural classification, with stricter requirements applied to critical infrastructure to ensure long-term durability.

8.6 References

[1] M.D.A. Thomas, K.J. Folliard, J.H. Ideker, North America (USA and Canada), in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[2] J. Lindgård, B. Grelk, B.J. Wigum, J. Trägårdh, K. Appelqvist, E. Hol, M. Ferreira, M. Leivo, Nordic Europe, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[3] E.M.R. Fairbairn, South and Central America, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[4] K. Yamada, T. Miyagawa, Japan, China and South-East Asia, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[5] A. Shayan, S. Freitag, Australia and New Zealand, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[6] A.K. Mullick, Indian Sub-Continent, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[7] T. Kay, A.B. Poole, I. Sims, Middle East & North Africa, in: Alkali-Aggregate Reaction in Concrete: A World Review, CRC Press, 2017. https://doi.org/10.1201/9781315708959.

[8] B. Fournier, M.-A. Bérubé, Alkali–aggregate reaction in concrete: a review of basic concepts and engineering implications, 27 (2000).
[9] P.J. Nixon, I. Sims, eds., RILEM Recommendations for the Prevention of Damage by Alkali-Aggregate Reactions in New Concrete Structures, Springer Netherlands, Dordrecht, 2016. https://doi.org/10.1007/978-94-017-7252-5.
[10] I. Sims, P. Nixon, RILEM Recommended Test Method AAR-0: Detection of Alkali-Reactivity Potential in Concrete—Outline guide to the use of RILEM methods in assessments of aggregates for potential alkali-reactivity, Mat. Struct. 36 (2003) 472–479. https://doi.org/10.1007/BF02481527.
[11] B. Fournier, M. Berube, Alkali-aggregate reaction in concrete: a review of basic concepts and engineering implications, CANADIAN JOURNAL OF CIVIL ENGINEERING 27 (2000) 167–191. https://doi.org/10.1139/l99-072.
[12] ASTM C1778, Guide for Reducing the Risk of Deleterious Alkali-Aggregate Reaction in Concrete, ASTM International, 2022. https://doi.org/10.1520/C1778-22.
[13] M. Thomas, A. Dunster, P. Nixon, B. Blackwell, Effect of fly ash on the expansion of concrete due to alkali-silica reaction – Exposure site studies, Cement and Concrete Composites 33 (2011) 359–367. https://doi.org/10.1016/j.cemconcomp.2010.11.006.
[14] B. Fournier, A. Bilodeau, N. Bouzoubaa, P.-C. Nkinamubanzi, Field and Laboratory Investigations on the Use of Fly Ash and Li-Based Admixtures to Prevent ASR in Concrete, in: United Kingdom, 2018.
[15] I. Borchers, J. Lindgård, C. Müller, Evaluation of laboratory test methods for assessing the alkali-reactivity potential of aggregates by field site tests, Materconstrucc 72 (2022) e286. https://doi.org/10.3989/mc.2022.17221.
[16] J. Custodio, D. Costa, A.B. Ribeiro, A.S. Silva, Assessment of potential alkali-silica reactivity of aggregates for concrete, 4TH INTERNATIONAL CONFERENCE ON STRUCTURAL INTEGRITY (ICSI 2021) 37 (2022) 590–597. https://doi.org/10.1016/j.prostr.2022.01.127.
[17] J. Lindgård, P.J. Nixon, I. Borchers, B. Schouenborg, B.J. Wigum, M. Haugen, U. Åkesson, The EU "PARTNER" Project — European standard tests to prevent alkali reactions

in aggregates: Final results and recommendations, Cement and Concrete Research 40 (2010) 611–635. https://doi.org/10.1016/j.cemconres.2009.09.004.
[18] T. Drimalas, K.J. Folliard, J.H. Ideker, Findings from the University of Texas at Austin ASR Exposure Site after 20 Years, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 516–523. https://doi.org/10.1007/978-3-031-59349-9_59.
[19] S. Hayman, M. Thomas, N. Beaman, P. Gilks, Selection of an effective ASR-prevention strategy for use with a highly reactive aggregate for the reconstruction of concrete structures at Mactaquac generating station, Cement and Concrete Research 40 (2010) 605–610. https://doi.org/10.1016/j.cemconres.2009.08.015.
[20] ASTM C1260, Test Method for Potential Alkali Reactivity of Aggregates (Mortar-Bar Method), ASTM International, 2022. https://doi.org/10.1520/C1260-22.
[21] R.E. Oberholster, G. Davies, An accelerated method for testing the potential alkali reactivity of siliceous aggregates, Cement and Concrete Research 16 (1986) 181–189. https://doi.org/10.1016/0008-8846(86)90134-1.
[22] ASTM C1293, Test Method for Determination of Length Change of Concrete Due to Alkali-Silica Reaction, (2024). https://doi.org/10.1520/C1293_C1293M-23A.
[23] A. Bergmann, L. Sanchez, Assessing the Variability of Laboratory Test Procedures for Predicting Concrete Field Performance Against Alkali-Aggregate Reaction (AAR), in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 507–515. https://doi.org/10.1007/978-3-031-59349-9_58.
[24] A. Bergmann, L.F.M. Sanchez, Assessing the variability of laboratory test procedures for predicting concrete field performance against alkali aggregate reaction (AAR), Manuscript Submitted for Publication (n.d.).
[25] A. Bergmann, L.F.M. Sanchez, Multifactorial analysis of AAR development: Integrating laboratory and field data with statistical and probabilistic modelling, Unpublished Manuscript (n.d.).

[26] A. Bergmann, L.F.M. Sanchez, Enhancing alkali-aggregate reaction (AAR) predictive models using machine learning to integrate laboratory and field data, Unpublished Manuscript (n.d.).
[27] J.H. Ideker, T. Drimalas, K.J. Folliard, A. Ghanizadeh, A. Parashar, K.S.T. Chopperla, A. Snyder, M.D.A. Thomas, Preventing Alkali-Silica Reaction in Concrete, Ce Papers 6 (2023) 1101–1109. https://doi.org/10.1002/cepa.2935.
[28] J.H. Ideker, A. Parashar, L. Fenstemacher, A. Ghanizadeh, T. Drimalas, K.J. Folliard, A. Hossack, M.D.A. Thomas, A. Snyder, A New Approach to Preventing ASR, in: L.F.M. Sanchez, C. Trottier (Eds.), Proceedings of the 17th International Conference on Alkali-Aggregate Reaction in Concrete, Springer Nature Switzerland, Cham, 2024: pp. 476–484. https://doi.org/10.1007/978-3-031-59419-9_56.
[29] CSA A23.2-27A, Standard Practices to identify degree of alkali-reactivity of aggregates and to identify measures to avoid deleterious expansion in concrete., in: CSA Standards A23.1–09/A23.2–09 Concrete Materials and Methods of Concrete Construction/Test Methods and Standard Practices for Concrete, 13th edition, CSA Group, Ontario, Canada, 2019: pp. 594–610.
[30] AASHTO R80-17. (2021) Standard Practice for Determining the Reactivity of Concrete Aggregates and Selecting Appropriate Measures for Preventing Deleterious Expansion in New Concrete Construction, (2021).
[31] C09 Committee, ASTM C1293-23 Test Method for Determination of Length Change of Concrete Due to Alkali-Silica Reaction, (2023). https://doi.org/10.1520/C1293_C1293M-23A.
[32] A. Bergmann, L.F.M. Sanchez, Alkali-Aggregate Reaction (AAR) Comparative Data: Accelerated Laboratory Outcomes vs. Field Performance, (2024). https://doi.org/10.17632/5cscpggnhg.1.

Chapter 9: Recommendations and Future Work

After conducting this comprehensive research, further recommendations and investigations are suggested:

- Expanding the dataset by incorporating a wider range of environmental conditions, aggregate types, and geographical contexts will improve the model's applicability across different concrete exposure conditions. Updating the established database, through collaboration with international research institutions and technical committees, will capture diverse environmental and material influences on AAR, creating a more universally relevant and adaptable framework.
- Incorporating alternative methodologies beyond AMBT and CPT can broaden the understanding of variations in testing setups, including local standards. By integrating these diverse methods, the model will better account for additional scenarios, improving its ability to predict AAR risks in different operational settings and environmental exposures.
- Broadening the investigation to include different SCMs and dosages would provide a more nuanced understanding of mitigation options for AAR. Due to data availability, the current study was unable to develop a model for each SCM type, which could potentially enhance AAR mitigation strategies. Expanding datasets to include diverse SCMs will support optimized material selection, enabling concrete mixes with targeted, limited expansions across various exposure conditions.
- Similarly, the analysis could be expanded to include chemical admixtures, specifically to evaluate their potential in mitigating AAR. Assessing admixture effectiveness would broaden the range of preventative measures, providing additional strategies to manage AAR risk across diverse environmental conditions and better quantify the probabilities of AAR occurrence in the field.
- A user-oriented software tool based on the flexible coupled threshold-time (FCTT) approach would enable the easy application of a risk-based AAR framework for real-world projects. By translating the model's predictive capacity into an accessible format,

users can more easily make informed decisions regarding materials, test conditions, and risk thresholds, thereby improving structural durability.

- Conducting field validation studies across different environments is essential for testing the flexible coupled threshold-time (FCTT) approach robustness in practical applications. These studies will help validate its predictive capacity across variable field conditions, leading to more confident implementation of AAR mitigation strategies. Extending this validation to full-scale structures would address the next critical gap, bridging the current work's focus on laboratory-to-block correlations to a broader block-to-structure scale.
- Developing dynamic threshold models that adjust to environmental changes, such as temperature and humidity fluctuations, would make AAR assessments more responsive to climate variability. The flexible coupled threshold-time (FCTT) could capture real-time environmental influences, enabling a more responsive and accurate prediction of AAR risk under changing exposure conditions. Considering the impacts of climate change, future analyses could also incorporate more detailed environmental variables such as annual rainy days, extreme temperature variations and freeze-tawing cycles, to better anticipate the evolving risks associated with AAR.