Assessing Broadband and Spectral Irradiance Variability for Solar Nowcasting Using Statistical Analysis and Machine Learning

Nick Anderson

Thesis submitted to the University of Ottawa in partial fulfillment of the requirements for the MASc degree in Electrical & Computer Engineering with Concentration in Applied Artificial Intelligence

School of Electrical Engineering and Computer Science
Faculty of Engineering
Supervisor: Professor Henry Schriemer
University of Ottawa
Ottawa, Canada

Table of Contents

List of Figures

List of Tables

Abstract

Solar photovoltaic (PV) resources are a key enabling technology in the global energy transition towards a more sustainable future. However, PV generation is highly variable due to the dynamic shading caused by clouds. To mitigate the effects of PV variability on electrical grid stability, grid operators rely on solar forecasts to proactively dispatch grid assets and balance supply and demand. To gain insights into the nature of solar variability, which is key for effective solar forecasting, this thesis presents a statistical assessment of high resolution spectral and broadband solar irradiance in Ottawa, Canada. The statistical assessment investigates the first- and second-order spectral and temporal dependencies of irradiance time series within the context of stationarity. The temporal structures indicate that solar irradiance processes are at best weakly stationary, and the implications for forecasting are discussed. The results of the statistical assessment are leveraged to develop several deterministic machine learning solar forecasting models (LSTM, XGBoost, and 1D-CNN). These models are implemented and compared in terms of computational complexity and prediction accuracy. It was found that under all sky conditions, the inclusion of spectral irradiance data improved forecasting performance compared to only using broadband irradiance. A ramp regime classification algorithm is then described, which enables the training and testing specialized ramp regime forecasting sub-models. These specialized sub-models were found to yield even greater forecasting accuracy within their respective ramp regimes, compared with the all-sky models. Further optimization and ensembling of the presented solar forecasting models is recommended for future work.

Acknowledgements

First and foremost, I would like to express my immense gratitude to Professor Henry Schriemer for all of his guidance and support throughout the past years. Every opportunity I have had during my Master's, and every opportunity in front of me today, has been made possible by him. I am also deeply thankful to Professor Karin Hinzer for all of her support, kindness, and invaluable wisdom. To Dr. Viktar Tatsiankou, your expertise and insights have made this entire project possible, for which I am sincerely grateful. To everyone at SUNLAB, thank you for making my Master's program an amazing and unforgettable experience.

The work in this thesis would not have been possible without funding from the National Science and Engineering Research Council Canada Graduate Scholarships-Master's (NSERC CGS-M) program. Thank you to the NSERC CREATE Training in Optoelectronics for Power: From Science and Engineering to Technology (TOP-SET) program for providing funding and educational resources. Thank you to our collaborators at Spectrafy for their knowledge, insight, and expertise.

Finally, thank you to all of my family and friends, who's unwavering support has kept me grounded and motivated over the years.

List of publications and contributions

This thesis begins with an introduction to several fundamental solar resource and machine learning principles, which are foundational to the analyses in later chapters. Then, the journal article manuscript listed below, which is intended for submission in 2023, is presented. This manuscript contains an extensive statistical analysis of solar irradiance variability within the context of data stationarity and its implications for forecasting approaches. Findings from this article are leveraged in the design and implementation of the machine learning forecasting models presented in Chapter 5.

1. **Nick Anderson**, Viktar Tatsiankou, Karin Hinzer, Richard Beal, and Henry Schriemer, "Statistical Assessment of Local Spectral and Broadband Solar Irradiance Variability Across Sub-Hourly to Seasonal Time Scales", to be submitted in 2023.

A second article, published through the SPIE Photonics West 2022 Conference Proceedings, is included in the appendix of this thesis. Though it is largely superseded by the above manuscript, some results exclusive to this conference proceeding are leveraged in an irradiance ramp regime classification algorithm, as outlined in Chapter 5 of this thesis.

1. **Nick Anderson**, Viktar Tatsiankou, Karin Hinzer, Richard Beal, and Henry Schriemer, "Probabilistic description of short-term cloud dynamics from rapid sampling of the solar spectral irradiance", Proc. SPIE 11996, Physics, Simulation, and Photonic Engineering of Photovoltaic Devices XI, 119960B (4 March 2022); doi: 10.1117/12.2616231.

Additional contributions from the author during their graduate studies, which are not featured in this thesis, are listed below.

1. **Nick Anderson**, Viktar Tatsiankou, Karin Hinzer, Richard Beal, and Henry Schriemer, “Cloud Dynamics Probed by Narrow and Broadband Solar Irradiance Probability Densities”, 8th World Conference on Photovoltaic Energy Conversion, Abstract and Poster Presentation, September 2022.
2. **Nick Anderson**, Viktar Tatsiankou, Karin Hinzer, Richard Beal, and Henry Schriemer, “Probabilistic Assessment of Narrowband vs Broadband Solar Irradiance Temporal Variability in Ottawa”, 49th IEEE Photovoltaic Specialists Conference, Abstract and Poster Presentation, June 2022.

3. **Nick Anderson**, Viktar Tatsiankou, Karin Hinzer, Richard Beal, and Henry Schriemer, "Rapid Spectral Sampling of Global Spectral Irradiance Variability Enables Probabilistic Description of Diverse Sky Dynamics", Photonics West 2022 Online, Abstract and Oral Presentation (On-Demand), Feb 2022.

Copyright Permissions

The journal article manuscript in Chapter 4 is reprinted with permission from co-authors V. Tatsiankou, K. Hinzer, R. Beal, and H. Schriemer. The conference proceeding in the appendix is reprinted with permissions from SPIE and co-authors V. Tatsiankou, K. Hinzer, R. Beal, and H. Schriemer.

Declaration of Originality

Unless otherwise stated, the work contained in this thesis was performed and written by Nick Anderson under the supervision of Henry Schriemer, with contributions from co-authors, as specified. To the best of the author's knowledge, the results contained in this thesis are original.

Chapter 1: Introduction

Solar photovoltaic (PV) energy resources are a key technology for enabling the global energy transition. The world's current dependence on fossil fuels for energy is considered unsustainable in terms of the environmental, societal, and economic impacts [1]. Moreover, as the global population and demand for energy continuously grow, limited fossil fuel reserves are rapidly depleting and thus cannot guarantee future energy security [1], [2]. However, the many environmental, societal, and economic consequences of relying on fossil fuels can, in part, be mitigated by shifting the world's energy mix to one that is dominated by renewable energy sources such as solar PV [1], [3], [4].

The installed capacity of PV has seen tremendous growth in recent years. Though it remains a small fraction of the required global energy supply, the high growth rate of PV makes it a promising resource for meeting future electricity needs [4]. However, while the potential for PV is great, its rapid growth presents many challenges – among which is its inherent variability [5]. During daylight periods, short term solar energy production is highly variable due to the dynamic shading caused by clouds moving across the sky. As clouds move in front of the sun, the amount of sunlight, or *irradiance*, that is incident upon the Earth's surface can drop substantially and rapidly, while the opposite occurs when clouds leave the sun's position in the sky. These fluctuations in incident irradiance result in similar fluctuations in the output power of PV modules, which hinders power quality and stability. At present, electricity grids can absorb most of these PV power fluctuations with their *grid inertia*, since the majority of grid electricity comes from large, stable, centralized power plants [6]. However, the future energy mix is expected to shift away from these central power plants and instead consist of mostly decentralized renewable sources, meaning much of the inherent grid inertia will be lost. Additionally, the expected growth of grid-connected PV will significantly increase the amount of variability being fed to the grid. Future electricity grids must therefore be able to accommodate higher variability with lower grid inertia in order to maintain stability and energy security. Thus, for PV to successfully drive the global energy transition, its variability must be understood and controlled.

1.1 Thesis Objectives

The challenge of achieving a stable electricity grid with PV as a primary contributor is a multidisciplinary one. Understanding the effects of Earth's dynamic atmostphere on the incoming

solar irradiance – and managing the resulting impacts on PV power – requires a mix of engineering, atmospheric sciences, and data analytics [7]. Once these atmospheric effects are understood, the power quality control mechanisms required for ensuring a stable grid with high PV penetration requires can be deployed more effectively.

To stabilize PV power generation, slower fluctuations caused by slow moving clouds can be balanced by dispatchable grid assets like electricity storage systems or small fossil fuel generators. Faster PV power fluctuations caused by fast moving clouds must be offset in a much shorter time frame, so the power electronics in PV inverters become responsible for handling these events. In both cases, however, a reactive approach to managing PV fluctuations after they have already occurred is insufficient. Instead, a proactive approach to managing PV fluctuations before they occur must be adopted.

To enable control measures to be taken in advance of fluctuation events, accurate forecasting of incident solar irradiance and PV generation is vital. The prediction of both irradiance and PV power is combined under the term *solar forecasting*. Although electricity grid operators are directly concerned with mitigating the impacts of PV power fluctuations, these fluctuations are merely the result of irradiance variability. Therefore, to gain a better understanding of the problem as a whole, this thesis addresses the root of the problem by assessing the underlying characteristics of solar irradiance variability and cloud dynamics.

The first major objective of this thesis is to investigate the statistical properties of irradiance variability. As earlier described, clouds move at different ground speeds which results in a range of irradiance ramp rates at the ground level. It therefore becomes necessary to assess the irradiance variability across multiple time scales in order to capture and quantify the fluctuations over very short (sub-second) to longer (hourly) periods. Furthermore, as clouds have different spectral impacts on the incoming irradiance, the spectral irradiance variability must be investigated as well. Thus, this thesis provides probabilistic descriptions of the spectral and broadband irradiance variability across time scales relevant to PV generation.

The second major objective builds upon the first by leveraging the spectral and broadband irradiance variability statistical behaviours in the development of multiple sub-hourly solar forecasting models. Three machine learning algorithms are used to develop these forecasting models, which are trained on one full year of high resolution data to capture all seasonal effects. The machine learning models are compared against the well-known smart persistence model [8], which is considered to have no forecasting skill as its predictions are made by simply shifting the time series forward. Given the differences in irradiance variability statistics at different sub-hourly time scales, individual forecasting models are trained to make predictions at forecast

horizons ranging from a few seconds up to one hour ahead. Although the aim of these models is to forecast the solar irradiance, these predictions can later be mapped to PV generation. This, however, is beyond the scope of this thesis.

In the irradiance variability statistical assessment and solar forecasting sections of this thesis – namely, Chapters 4 and 5 – the data involved is from a new spectral irradiance database created in Ottawa, Canada. The data is generated by a custom SolarSIM-G from Spectrafy, which measures the spectral irradiance at nine wavelengths across the solar spectrum and derives the broadband irradiance from these measurements. The database contains measurements from June 2021 until present, with 250 ms resolution.

1.2 Thesis Overview

Chapter 2 of this thesis introduces the theory and contextual information necessary to motivate and understand the research presented in later chapters. This chapter begins with a review of the solar resource, its broadband and spectral components, and its sources of variability. These concepts are presented within the context of PV generation variability and the resulting impacts on electrical grid stability. The role of and best practices for operational solar forecasting are then discussed.

Chapter 3 provides the fundamental theory underlying the machine learning algorithms used to create the solar forecasting models in Chapter 5. The operating principles behind the long short-term memory (LSTM), extreme gradient boosting (XGBoost), and 1-dimensional convolutional neural network (1D-CNN) models, which make up the forecasting models in this thesis, are presented. Several foundational models from which the LSTM, XGBoost, and 1D-CNN models are derived are also described in order to provide a complete description of the algorithm architectures.

Chapter 4 presents a journal article manuscript intended for submission in 2023. In this manuscript, an extensive statistical assessment of spectral and broadband irradiance variability is provided within the context of solar forecasting and time series stationarity. This paper uses one full year of data from 2022 to investigate irradiance variability, from both spectral and broadband perspectives, on time scales ranging from seconds to seasons. The assessment focuses on the quantification of the temporal and spectral dependencies which exist on these time scales, and aims to explain their origins.

Chapter 5 describes the short-term solar forecasting (*nowcasting*) models implemented in this

thesis. Several machine learning models are trained, tested, and compared in accordance with proper solar forecasting research practices [9], using ground-based broadband and spectral irradiance measurement data from all sky conditions. An approach to irradiance ramp regime classification, which leverages the results from the manuscript in Chapter 4 and the conference proceeding provided in the appendix, is then presented. Different ramp-regime-specific nowcasting models are then trained using the proposed classification method, and their performances are compared with each other and the models trained on all sky conditions.

Chapter 6 concludes this thesis with a summary of its main contributions to the fields of solar resource assessment and solar forecasting. Several next steps are also suggested to help identify and overcome the limitations of the approaches taken in earlier chapters, such that the nowcasting models can be made operationally applicable.

Chapter 2: Solar Resource and Forecasting Background

2.1 The Solar Resource

The energy delivered by the sun to the Earth's surface can be converted to electricity by photovoltaic generators (solar panels). Before entering the Earth's atmosphere, the sun provides a total extraterrestrial solar irradiance of approximately 1361 W/m^2, which is known as air mass zero (AM0) [10]. As it travels through Earth's atmosphere towards the surface, the irradiance interacts with aerosols, clouds, and other atmospheric constituents which cause different scattering and absorption processes to occur. These processes disperse the light in different directions and, as a result, the total solar irradiance measured at the Earth's surface is decomposed into two components: the direct normal irradiance (DNI) and the diffuse horizontal irradiance (DHI). The DNI is the unobstructed irradiance traveling directly from the sun to the Earth's surface, while the DHI is all of the incident light that has been scattered or reflected by clouds, the ground, or other surfaces as illustrated in Figure 2.1.

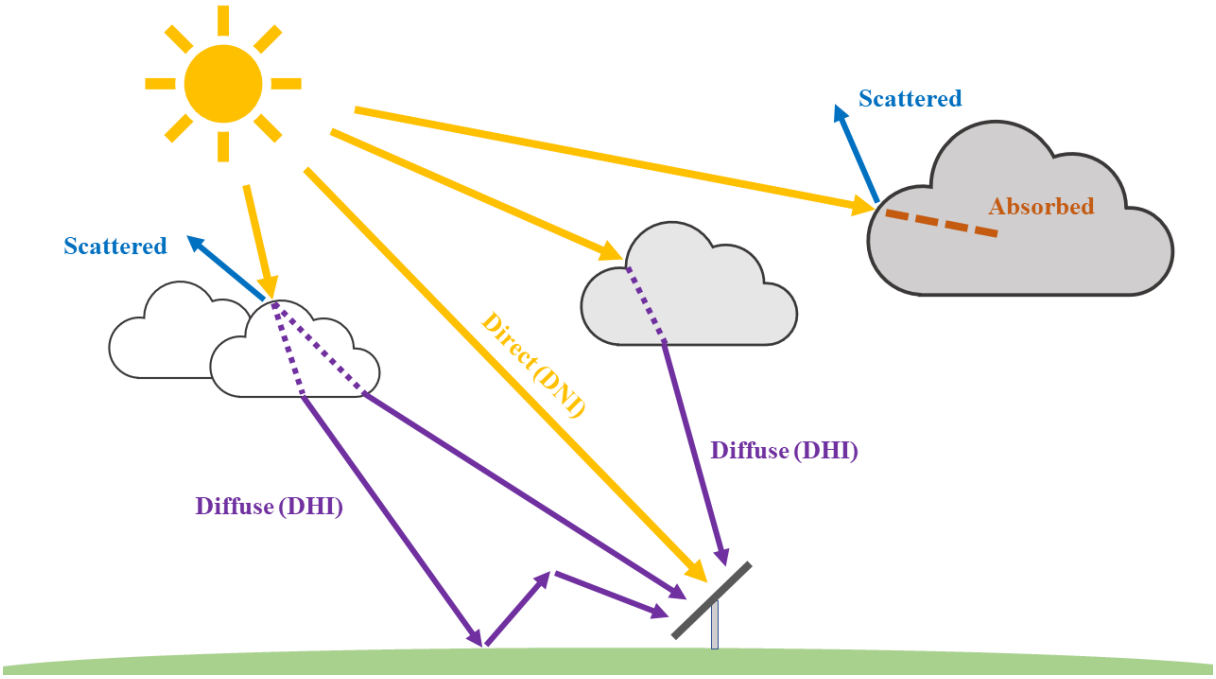

Figure 2.1: Illustration of irradiance components

The DNI and DHI can be combined in a total irradiance metric, known as the global horizontal irradiance (GHI):

$$GHI = DNI \cdot sin\theta + DHI \tag{1}$$

where θ is the solar elevation angle. Under clear-sky conditions (*i.e.*, when there are no clouds present), the DNI, DHI, and GHI can all be estimated using a theoretical clear-sky model. Clear-sky models take into account the Earth's atmospheric conditions in the absence of clouds, as well as the solar coordinates. The solar coordinates are used to identify the location of the sun relative to a specific site on the Earth's surface using a spherical coordinate system. Many clear-sky models require only two solar coordinates: the elevation angle (θ) and the azimuth angle (ϕ). The elevation angle corresponds to the sun's angular height relative to the horizontal plane (i.e., the ground). The elevation angle is 90° when the sun is directly overhead, while at sunrise and sunset the elevation angle is 0°. In contrast, the azimuth angle corresponds to the sun's lateral position relative to north. For instance, when the sun is directly south of an observation site the azimuth angle will be 180°.

2.1.1 Broadband and Spectral Irradiance

Fundamentally, solar irradiance is electromagnetic radiation emitted by the sun. It is made of a spectrum of wavelengths, ranging from the ultraviolet (UV) to infrared (IR), with varying intensities as illustrated in Figure 2.2. The broadband irradiance is the integrated solar spectrum, from UV to IR and all of the visible light wavelengths in between. In contrast, the spectral irradiance refers to the sunlight at a specific wavelength. In practice, the spectral irradiance is typically measured as a very narrow band of the solar spectrum surrounding the center wavelength.

The distinction between broadband and spectral irradiance is significant in the context of PV generation. Broadband irradiance data is typically considered to be sufficient for PV generation modelling and forecasting, meaning the information that is potentially embedded in the spectral components gets overlooked. The neglection of the spectral components is largely due to a lack of available spectral irradiance data. There are several broadband irradiance databases for sites around the world, such as NSRDB [12] and PVGIS [13], but very few which include spectral measurements. However, it has been shown that aerosols, cloud types, and other atmospheric constituents affect different spectral regions in different ways [14]. As the irradiance variability induced by clouds is therefore likely to be spectrally dependent, it is also likely that different PV materials will experience different variability under the same sky conditions. Therefore,

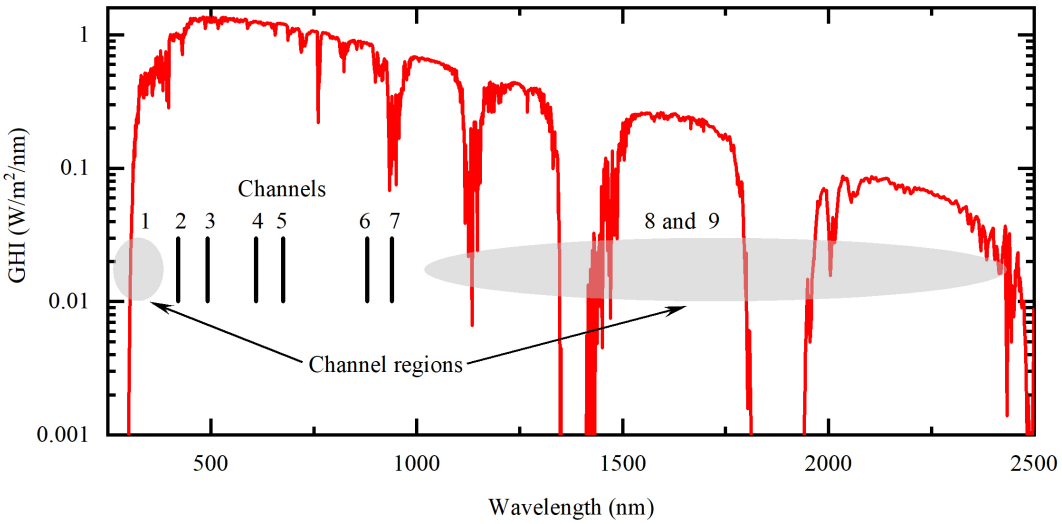

Figure 2.2: Snapshot of the measured solar spectrum capture by the nine spectral channels of a Spectrafy SolarSIM-G [11]

in the context of solar modelling and forecasting, both broadband and spectral irradiance data should be considered.

2.1.2 Solar Irradiance Variability

In cloudless sky conditions, the solar irradiance can be accurately derived using theoretical clear-sky models, in which there are only two sources of irradiance variability. First, there is the diurnal (daily) variability caused by the sun's path across the sky – rising in the east, reaching its maximum elevation angle at solar noon, and setting in the west. The received irradiance at the ground level naturally follows a similar diurnal pattern with the least irradiance in the early morning and late evening, the most around midday, and none at night. This diurnal behaviour is illustrated for three days using a clear-sky model in Figure 2.3.

The second source of clear-sky irradiance variability is the orbital cycle, which is caused by the Earth's elliptical orbit around the sun. During the summer solstice on June 21, the period of daylight and the maximum daily solar elevation angle are the largest in the year. In contrast, during the winter solstice on December 21, the period of daylight and the maximum daily solar elevation angle are the smallest. For days between the two solstices, the daylight periods and maximum solar elevation angles progressively increase and decrease depending on the time of year. These variations in day length and maximum daily solar irradiance are also taken into

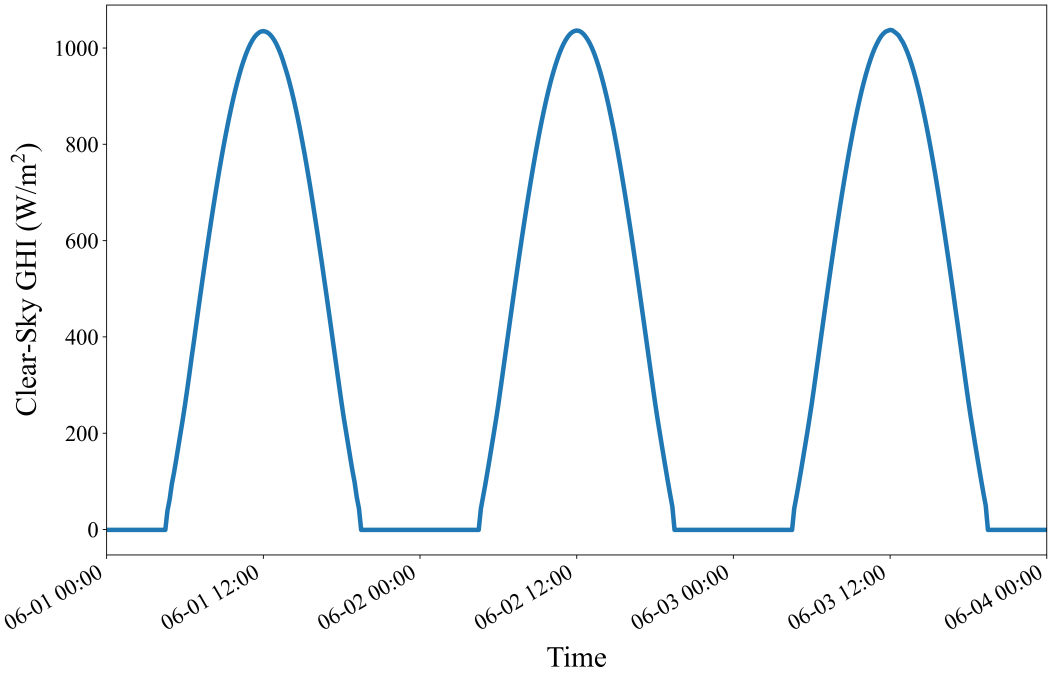

Figure 2.3: Modelled clear-sky GHI for the first three days in June 2022

account by clear-sky models.

Although climates vary widely between locations around the world, even very sunny places do not experience completely clear skies year-round. Hence, while clear-sky models are useful for capturing the diurnal and orbital irradiance trends, their applicability in the real world is limited because they neglect the third major source of variability: clouds. The impact of clouds on the incident solar irradiance is illustrated by comparing the modelled clear-sky GHI in Figure 2.3 with the actual GHI measured on the same days in Figure 2.4. Unfortunately, there is no theoretical model that can perfectly capture irradiance behaviour in the presence of clouds. The impact of any given cloud on ground-level irradiance is determined by many factors, such as the cloud's size, density, aerosol optical depth, altitude, ground speed, and position in the sky relative to the sun. This is further complicated by the presence of many clouds in the sky with different attributes, which is extremely common.

Formally, the total irradiance variability across all time scales and all sky conditions can be expressed by decomposing it into the three main contributors described above:

$$\Delta E = \Delta E_{orbital} + \Delta E_{diurnal} + \Delta E_{clouds} \tag{2}$$

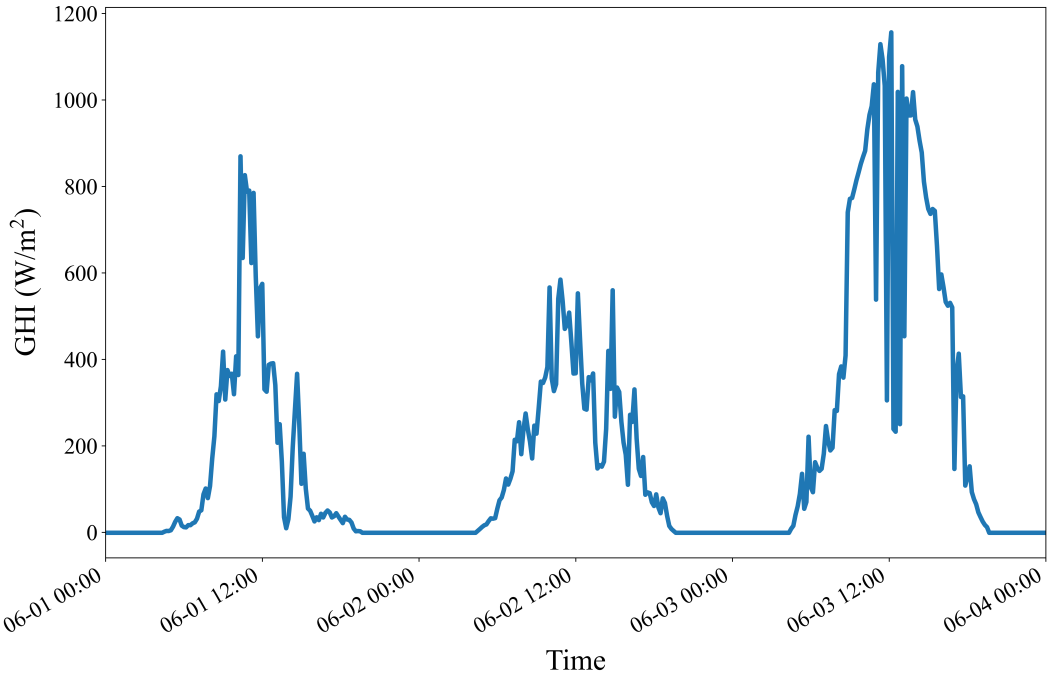

Figure 2.4: Measured GHI for the first three days in June 2022

where ΔE is the total irradiance variability, $\Delta E_{orbital}$ and $\Delta E_{diurnal}$ are the known (deterministic) variability components, and ΔE_{clouds} is the unknown (non-deterministic) variability induced by clouds. The orbital and diurnal components are well understood and can be easily predicted by clear-sky models. On the contrary, cloud-induced irradiance variability is the most challenging to capture and of major significance in the context of PV generation modelling and forecasting.

Given the complexity of cloud-induced irradiance variability, empirical (i.e., data-driven) approaches to modelling and characterizing it are more realizable than theoretical ones. In this thesis, a database containing ground-based broadband and spectral irradiance measurement data is used to quantify the cloud-induced irradiance variability on a range of time scales. Before the cloud-induced variability can be probed directly, the irradiance measurements must first be detrended of their deterministic diurnal and orbital patterns. This is commonly achieved using the clear-sky index, κ^*, which is the clear-sky normalization of the measured irradiance [15]:

$$\kappa^*(t) = \frac{E(t)}{E_{clr}(t)} \tag{3}$$

where $E(t)$ and $E_{clr}(t)$ are the measured irradiance and theoretical clear-sky irradiance values at a particular time, t, respectively. The asterisk included in the $\kappa^*(t)$ notation is used to differ-

entiate this clear-sky index, which uses a terrestrial clear-sky model that includes atmospheric irradiance effects, from the *clearness index*, $\kappa(t)$, which uses an extraterrestrial clear-sky model and neglects the Earth's atmosphere. Dividing the measured irradiance by the theoretical clear-sky irradiance, the diurnal and orbital patterns are effectively removed from the measured data. As such, the remaining variability in the clear-sky index is dominated by cloud dynamics. It should be noted that the value of the clear-sky index can exceed 1 – this situation occurs in lensing conditions, when the sun disk is unobstructed but surrounded by clouds which reflect additional light towards the ground site. While Eq. (3) represents the broadband clear-sky index, the spectral clear-sky index can be similarly defined as:

$$\kappa^*(\lambda;t) = \frac{E(\lambda;t)}{E_{clr}(\lambda;t)} \tag{4}$$

where $E(\lambda;t)$ and $E_{clr}(\lambda;t)$ are the measured and theoretical clear-sky spectral irradiance, respectively, located at wavelength, λ.

After isolating the cloud-induced variability (ΔE_{clouds}) using the clear-sky index, it can then be quantified using the clear-sky index increment [15], $\Delta\kappa^*_\tau$, which is defined as:

$$\Delta\kappa^*_\tau(t) = \kappa^*(t+\tau) - \kappa^*(t) \tag{5}$$

where $\kappa^*(t)$ and $\kappa^*(t+\tau)$ are two clear-sky index values separated by a time step, τ. In other words, the clear-sky index increment is the forward difference of the clear-sky index across a time horizon defined by τ. Again, while Eq. (5) expresses the broadband clear-sky index increment, the spectral equivalent is defined as:

$$\Delta\kappa^*_\tau(\lambda;t) = \kappa^*(\lambda;t+\tau) - \kappa^*(\lambda;t) \tag{6}$$

Intuitively, the value of the clear-sky index increment corresponds to the change in normalized irradiance caused by clouds. The clear-sky index increment is relatively large (closer to ± 1) when the clouds cause a significant irradiance change, and relatively small (closer to 0) when the clouds introduce little variability. The former situation is common in mixed sky conditions, when there are many clouds moving across the sun's position in the sky causing dynamic shading on the ground. The latter is common when the sky is more homogeneous around the sun's position, such as in stable clear or overcast conditions. Since the clear-sky index can have values

greater than 1, the clear-sky index increment likewise has the potential to exceed values of ± 1 in very extreme sky condition changes. Finally, the clear-sky index increment can have either positive or negative values, such as when a cloud leaves or blocks the sun's position in the sky, respectively. The mapping between the clear-sky index and the clear-sky index increment is illustrated in Figure 2.5.

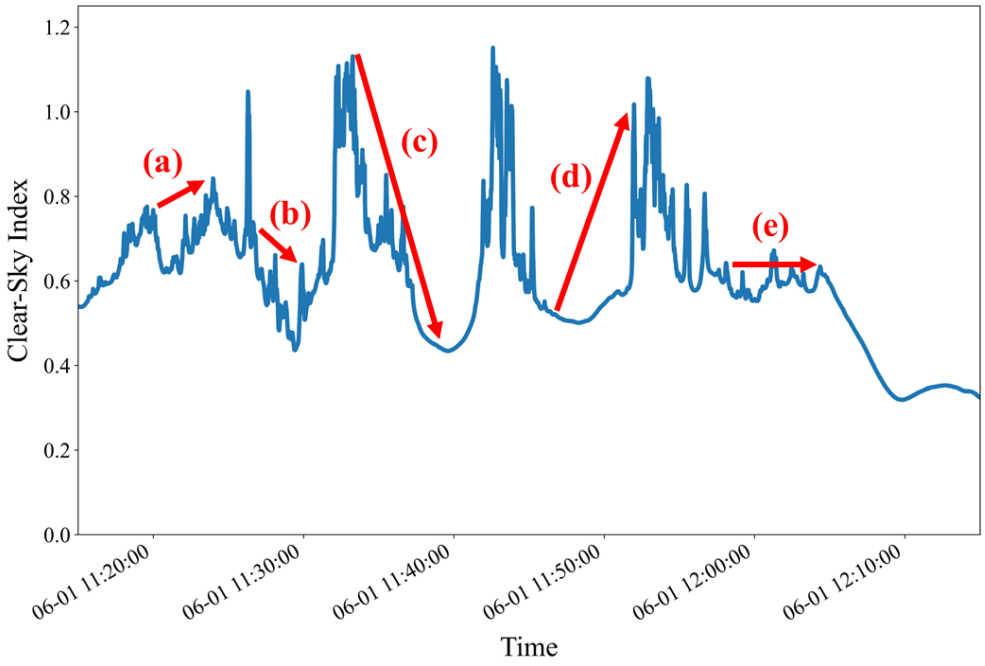

Figure 2.5: Snapshot of the GHI clear-sky index illustrating (a) small positive, (b) small negative, (c) large negative, (d) large positive, and (c) near-zero clear-sky index increments

As can be seen in Figure 2.5 above, irradiance fluctuations caused by moving clouds can occur over different time scales with different ramp rates. For this reason, the clear-sky index increment must be computed for a range of τ values in order to capture all variability, from rapid fluctuations taking only a few seconds to more gradual changes taking several minutes or even hours.

Chapter 4 of this thesis investigates the clear-sky index increment probability distributions for various sub-hourly time steps in order to statistically model the cloud-induced variability. These distributions are computed for the broadband and spectral irradiance time series using kernel density estimation [16], which empirically approximates the probability density function of a sample distribution. For a given sample distribution $(x_1, x_2, ..., x_n)$, the kernel density estimator is given by:

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-x_i}{h}) \tag{7}$$

where $\hat{f}_h(x)$ is the unknown probability density at any given point x, h is the bandwidth which acts as a smoothing parameter, and the kernel, K, is a non-negative window function. In this work, Gaussian kernel density estimation is used, meaning a normal (Gaussian) kernel is used. The bandwidth is chosen using Silverman's rule-of-thumb estimator [17], implemented as:

$$h = \left(\frac{n(d+2)}{4}\right)^{-\frac{1}{d+4}} \tag{8}$$

where n is the sample size and d is the number of dimensions of the data. Since the clear-sky index increment time series is one-dimensional, Eq. (8) simplifies to:

$$h = \left(\frac{3n}{4}\right)^{-\frac{1}{5}} \tag{9}$$

2.2 Impacts of High PV Penetration on Electricity Grids

Due to the cause-and-effect relationship, solar irradiance variability is directly coupled with PV generation variability; large fluctuations in incident solar irradiance will cause large fluctuations in PV power. In grids where PV is a relatively minor contributor, solar variability is not of great concern since the large, high-inertia baseload (e.g., coal, nuclear, or hydro) power plants can absorb the fluctuations introduced by PV. However, in grids where PV is a significant contributor, the impacts of its variability become a major concern. With enough grid-connected PV capacity, large fluctuations experienced in the overall PV supply can cause grid voltages to deviate beyond allowable limits. This can result in power quality issues for consumers, possible damage to grid-connected appliances and equipment, and even area-wide blackouts in extreme cases. These risks will also become greater over time as large baseload plants are removed from future grids, which are shifting towards a distributed resource architecture. Beyond the threat to grid stability, the balance between electricity supply and demand in grids that heavily rely on PV can be jeopardized during periods of low irradiance. While a crude solution to such supply-demand imbalances is load-shedding, this is not desirable for consumers who require uninterrupted energy access.

2.2.1 Mitigation of PV Variability

There are a number of actions which grid operators can take to mitigate the impacts of PV variability. To some degree, fast PV power fluctuations can be handled by inverters before feeding into the grid. Inverters play several key roles in PV systems, but their primary responsibilities include power conditioning and converting the direct current (DC) PV output to alternating current (AC) to synchronize with the grid. While some inverters are capable of preventing fast voltage transients from being fed into the grid, inverters cannot supply additional energy to offset the lacking PV generation during periods of low irradiance. For this, dispatchable grid assets are needed.

Dispatchable grid assets include generators and energy storage systems which can be controlled by grid operators and "dispatched" as needed [18]. Grid batteries are typically used to help absorb and offset the fast ramps in PV power, as they can be charged and discharged rapidly. However, utility-scale batteries are very expensive, so grid operators typically have a limited capacity of battery-stored energy at their disposal. To offset low PV power for many seconds or minutes, grid operators rely on other assets like small diesel or natural gas generators. These generators can be started up quickly to supplement the PV power, but their operation should only be as-needed since they can be expensive to run, environmentally harmful, and require frequent maintenance and refueling. This can be achieved with advanced warning of PV fluctuations, which allows operators to optimally schedule grid asset dispatching ahead of time. This way, PV variability can be effectively and proactively managed while allowing generators to be turned off whenever they are not needed.

Finally, grid operators attempt to reduce the amount of variability introduced by PV to the grid by penalizing prosumers and PV plant owners who cannot accurately predict how much power they will sell to the grid in the near future. In real-time (spot) markets, prosumers and PV plant owners bid on how much energy they expect to be able to sell a short time into the future. Bidding a high amount of electricity results in a higher payout from the grid, while bidding a low amount will decrease the payout – provided the bid matches how much electricity is actually supplied [19], [20]. However, when a high amount of electricity is bid but a low amount of electricity is delivered to the grid, a penalty is given to dissuade greedy bidding which prevents grid operators from scheduling their assets accurately. Hence, it is in the best economic interests of prosumers and PV plant owners to be able to accurately forecast how much electricity they will generate in the near future. Solar forecasting therefore not only improves grid stability and asset scheduling, but it can also improve PV economics and make the renewable energy technology more affordable in transactive energy markets [21].

2.3 Solar Forecasting

Having established the need for solar forecasting and its enabling role in the global energy transition, the remainder of this chapter outlines the solar forecasting approaches, requirements, and guidelines that will be used in later chapters of this thesis.

2.3.1 The Persistence Model

The simplest approach to solar forecasting is to use the persistence model [8]. The persistence model relies on the assumption that the conditions some time step into the future will be the same as the current conditions (i.e., the current conditions will *persist* into the future). The persistence model neglects all sources of variability, including the diurnal and orbital trends. Formally, the persistence model is defined as:

$$\hat{E}(t+\tau) = E(t) \tag{10}$$

where $E(t)$ is the current irradiance and $\hat{E}(t+\tau)$ is the predicted irradiance after some time step, τ. In the context of forecasting, the time step is referred to as the *forecast horizon*. The persistence model is considered to be a naïve model with no *forecasting skill* because it does not intelligently account for variability. In practice, implementing the persistence model is equivalent to shifting the measured irradiance time series data by the desired forecast horizon.

2.3.2 The Smart Persistence Model

The smart persistence model, which is a slightly more sophisticated version of the persistence model, takes into account the diurnal and orbital irradiance trends [8]. With the persistence model, there will always be some amount of forecasting error due to the change solar position over the forecast horizon – even if the sky condition is perfectly persistent. With smart persistence, forecasts are instead made using the clear-sky index:

$$\hat{\kappa}^*(t+\tau) = \kappa^*(t) \tag{11}$$

where $\kappa^*(t)$ is the current clear-sky index and $\hat{\kappa}^*(t+\tau)$ is the predicted future clear-sky index. As described in Section 2.1.2, the clear-sky index detrends the measured irradiance time series

of the diurnal and orbital trends. Therefore, the only forecasting errors for the smart persistence model will be strictly due to cloud-induced variability. The predicted clear-sky index value can then be mapped back to an irradiance value using the known clear-sky irradiance value at the target time stamp:

$$\hat{E}(t+\tau) = \kappa^*(t) \cdot E_{clr}(t+\tau) \tag{12}$$

While simple, smart persistence is foundational in the field of solar forecasting. The smart persistence model is typically used as a benchmark since it performs rather well at short forecast horizons. For instance, if the forecast horizon is 1 second into the future, it is likely that the measured irradiance will not change significantly – since cloud motion is limited within that 1 second – which would naturally favour the smart persistence model's predictions. In fact, the smart persistence model typically outperforms even the most complex modern forecasting models at sufficiently short forecast horizons. However, smart persistence becomes increasingly unreliable as the forecast horizon gets longer since sky conditions have more time to change drastically. Nonetheless, smart persistence serves as a universal benchmark for forecasting and models that cannot outperform it at a particular forecast horizon are considered to have no forecasting skill and are insufficient.

2.3.3 Approaches to Solar Irradiance Forecasting

In recent years, many approaches to solar irradiance forecasting have been proposed. Several methods use numerical weather prediction (NWP) models, such as [22], [23], and [24] which use current weather conditions to predict the future irradiance. Other approaches like [25], [26], and [27] use satellite data to make irradiance forecasts, which typically involve mapping satellite images to the irradiance incident at ground-level. NWP and satellite irradiance forecasting approaches have the benefit of being widely applicable around the world; these models and data often cover entire countries or even continents. However, NWP- and satellite-based solar forecasting approaches are often limited to low spatial and temporal resolutions. That is, most data NWP and satellite data is between 15 minute to 1 hour resolution and covers large spatial regions (often several kilometers wide) [12], [28], [29]. Hence, very short term local forecasting is limited with these approaches.

Sky imagers, which capture images of the sky from the ground, have been used extensively in solar forecasting, as in [30], [31], and [32]. Sky imagers help overcome the limitations of

satellite data by taking the ground-based perspective; rather than a wide area image of the sky from above, sky imagers capture the local sky conditions. Additionally, sky imagers often have much higher temporal resolutions, which means shorter forecast horizons can be addressed. Sky imagers are also able to track local cloud motion [33], which results in the spatial element of cloud dynamics to be accounted for in forecasts. While sky imagers are useful for irradiance forecasting, processing sky images and mapping them to irradiance can be a complex and computationally intensive task [34].

Groud-based data has also been found to be useful for solar forecasting. This can included the use of pyranometers [35], which measure local (point) irradiance with typically high temporal resolution. While individual pyranometers do not capture spatial information, such as cloud or shadow motion, networks of pyranometers spread out over a small region can overcome this challenge [36]. Combining spatial information with high temporal data resolution, accurate short-term forecasts of local irradiance can be achieved. To overcome the limitations of the different types of data used for forecasting, multiple data sources can be combined (e.g., ground- and satellite-based) for improving irradiance forecasting [37]. This way, the higher temporal resolution and precision of ground-based sensors is combined with the wide-area spatial information provided with satellite or sky imaging data.

Besides the type of data used for generating solar irradiance forecasts, there have also been many forecasting algorithms proposed. Several comprehensive review papers have compiled some of the more promising forecasting models. For instance, [38] assesses linear (statistical) and non-linear (machine learning) models and their applicability to the different data sources described above. Other studies, such as [39] and [40], focus on the opportunities for better forecasting provided by highly non-linear deep learning models. In [41], an extensive evaluation of 68 machine learning models for generating hourly forecasts is presented. Conclusions from this study reveal that there is unlikely a universally best algorithm for forecasting; each algorithm has different strengths and weaknesses, and model selection is often application (i.e., climate and forecast horizon) dependent. Hence, solar forecasters should test and compare different algorithms, or combine multiple models in ensembles.

Finally, there are two main categories of forecasting approaches: deterministic and probabilistic [9]. Deterministic models provide a single output value determined by the forecasting model parameters. As such, deterministic models neglect sources of randomness that might cause their predictions to be inaccurate. Deterministic forecasting can be useful for processes without much or any randomness; for instance, a deterministic linear regression is useful for making predictions along a straight line. However, deterministic forecasts are limited when used to forecast a

physical process like solar irradiance, given the many sources of randomness involved. In contrast, probabilistic forecasting models acknowledge that it is not possible to predict the target variable's exact future value with absolute certainty. Instead, probabilistic forecasting models account for randomness by predicting a probability density rather than a single value. Thus, although deterministic forecasting approaches are more abundant and easier to implement, developing probabilistic models for solar forecasting should be the ultimate objective. Although probabilistic solar forecasting is beyond the scope of this thesis, converting deterministic forecasts to probabilistic is possible and is left as future work [42].

2.4 The ROPES Guideline

The field of solar forecasting has received a great deal of research attention in recent years, resulting in a plethora of proposed models in the literature. Ranging from traditional statistics-based approaches to sophisticated machine learning and artificial intelligence (AI) algorithms, there are seemingly endless solar forecasting model options to choose from. However, the rapid expansion of the solar forecasting research field in absence of standardized practices has resulted in a lack of consistency between works. This has introduced challenges for researchers and industry engineers when trying to make fair and complete comparisons between proposed forecasting approaches.

To help resolve the issues outlined above, Yang proposed the ROPES guideline for standardizing solar forecasting research practices [9]. With the aim of making solar forecasting research more useful to industry engineers and other researchers, the ROPES guideline suggests 5 key principles for researchers to follow: forecasting approaches should be **r**eproducible, **o**perational, **p**robabilistic and/or **p**hysically-based, an **e**nsemble, and evaluated using a **s**kill score. Each of these key principles are further described below.

- **Reproducible:** The code and data used in the work should be made publicly available (if possible) such that the results can be reproduced and verified. If proprietary concerns exist, partial and/or synthetic data and code should be made available.
- **Operational:** The forecasting method should account for the operational constraints and requirements of the application (e.g., computational resource demand and lead times). Reporting the approach's ability to meet real-world operational requirements enables grid operators to properly assess and select the most suitable forecasting methods for their needs.

- **Probabilistic and/or Physically-Based:** The forecasting approach should ideally be probabilistic rather than deterministic. Proposed approaches should also be based on models of the physical processes being forecasted, rather than purely data-driven; data-driven models should be based on data that incorporates physical models.
- **Ensemble:** The forecasting method should consist of an ensemble which diversifies the forecasts to reduce uncertainty (e.g., multiple parallel models should be strategically combined). Ensembles can also aid in data and dimensionality reduction, which results in more lightweight models. While hybrid models are considered ensembles, simply rearranging existing hybrid configurations lacks innovation and significance and should therefore be avoided.
- **Skill Score:** The forecasting approach should be evaluated using a skill score. For deterministic (single output value) approaches, the skill score should be computed with the smart persistence model as the reference. For probabilistic (probability density output) approaches, the skill score should use the persistence ensemble as the reference. Using the skill score as a universal evaluation metric helps increase the consistency and comparability between solar forecasting works. It also discourages researchers from cherry-picking evaluation metrics (e.g., root mean square error, mean squared error, mean bias error, etc.) which best suit their models and biases the apparent performance.

The aforementioned skill score (*SS*), which is typically expressed as a percentage, is becoming the commonly accepted metric for evaluating all deterministic solar forecasting approaches. The skill score is defined as:

$$SS = 1 - \frac{e_f}{e_{ref}} \tag{13}$$

where e_f is the error of the forecasting model being evaluated and e_{ref} is the error of the reference forecasting model. As outlined above, the smart persistence model should be used as the reference for evaluating deterministic models. This also implies that all forecasts should be made using the clear-sky index rather than the measured irradiance, since the former enables a direct evaluation on the model's ability to forecast cloud-induced variability. A forecasting model is said to be skillful if it achieves a positive skill score, with higher skill scores indicating better forecasting accuracy. However, Yang identifies that there are practical limits which bound skill scores, and that reasonable forecasting models will likely have skill scores below 70%.

To avoid the issue of error metric cherry-picking, Yang proposes that the root mean squared error (RMSE) should always be used when computing the skill score. Hence, Eq. (13) becomes:

$$SS = 1 - \frac{RMSE_f}{RMSE_{ref}} \tag{14}$$

where $RMSE_f$ is the RMSE of the forecasting model being evaluated and $RMSE_{ref}$ is the RMSE of the reference (i.e., smart persistence) model. The RMSE can be computed as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{15}$$

where $\hat{y}_i$ is the predicted value, y_i is the actual future observation, and N is the number of forecast samples being used to compute the RMSE.

Chapter 3: Machine Learning Forecasting Models

Forecasting models can largely be categorized as either conventional or machine learning (ML) based. In conventional models, programmers must explicitly outline all rules, logic, and decision-making steps involved in making a forecast. This requires the programmer to have extensive knowledge of the process being forecast, and is limited by the parametric models that have been developed to describe that process. In contrast, ML algorithms do not require a fully-defined model to be explicitly programmed. Instead, ML algorithms are data-driven – they are capable of deriving (or "learning") their own internal parameters to model the process. When forecasting complex physical processes like cloud-induced irradiance variability, it is extremely challenging to develop the theoretical models required for conventional approaches. Therefore, data-driven ML models are more appropriate for solar forecasting.

While more flexible and less reliant on intricate theoretical models, there are drawbacks to ML algorithms which must be considered. First, these models require a large volume of historical samples (or *training data*) to learn from. With too little training data, ML models will be unable to capture the relationships between the input and the output variables, and will therefore *underfit* the process. Conversely, when ML models see the same data too many times during the training process (i.e., too many *epochs*, or passes through the training data), they become susceptible to *overfitting*. An overfit model memorizes the training data rather than learning the relationships between the input and output variables, which means the model will have poor *generalization* to new, unseen data. Hence, finding a balance between underfitting and overfitting is key to training effective ML models.

Another drawback of machine learning models are their lack of explainability. Unlike conventional forecasting models which rely on explicitly defined parameters and logic, ML models derive their own internal parameters to capture relationships and patterns in the training data. The types of patterns learned by ML models and the ways in which they are used for making forecasts are unknown to the programmer in most cases, making them "black box" models. In an operational context, grid operators and solar forecasters may not care about the way in which a model makes a forecast as long as the forecast is accurate. However, the lack of explainability of how ML models operate internally can limit their reproducibility and the amount of research insight that can be drawn.

In this thesis, three carefully selected machine learning forecasting models are developed and compared in terms of their forecasting performances, training times, and prediction times to

assess their efficacy in operational contexts. The selected models are the Long Short-Term Memory (LSTM) network, the Extreme Gradient Boosting (XGBoost) algorithm, and the 1-Dimensional Convolutional Neural Network (1D-CNN). The objective of these models is to make skillful (i.e., positive skill score) forecasts of the broadband solar irradiance on sub-hourly time scales. Using one full year of 1 s resolution spectral and broadband irradiance data, it is possible to train these models to make forecasts on horizons down to a few seconds ahead. The remainder of this chapter describes the operational theory underlying the LSTM, XGBoost, and 1D-CNN model architectures, as well as the reasoning behind the algorithm selections.

3.1 The Long Short-Term Memory Network Model

Before diving into the complex architecture of LSTMs, it is important to first understand the basic ML models from which LSTMs are derived. This section first provides descriptions of two elementary ML architectures, namely artificial neural networks and recurrent neural networks, before presenting the LSTM model.

3.1.1 Artificial Neural Networks

The feed forward artificial neural network (ANN) is one of the most fundamental neural network architectures [43], [44]. As illustrated in Figure 3.1, the basic ANN architecture consists of three stages: the input layer, the hidden layer, and the output layer. For time series prediction problems, the input layer accepts the input variables for a given timestamp, t, in the form of an input vector, $\vec{x}$. The output layer provides the predicted future value of the target variable, $\hat{y}_{t+\tau}$.

The *hidden layer* exists between the input and output layers of the ANN and consists of nodes, or *neurons*, which are represented by the orange circles in the above illustration. In basic ANNs, each neuron receives the full input vector from the input layer and combines each value in a weighted summation. The weights assigned to each input variable for each neuron ($w_{n,m}$) are illustrated in Figure 3.1 by the lines connecting the input and hidden layers. The weighted summation is then passed through an activation function which determines whether the neuron is activated or not. A common activation function is the Sigmoid function, which is a non-linear S-shaped curve that compresses its input between 0 and 1. The output of the hidden layer in an ANN with N input variables, M neurons in the hidden layer, and a Sigmoid activation function will be:

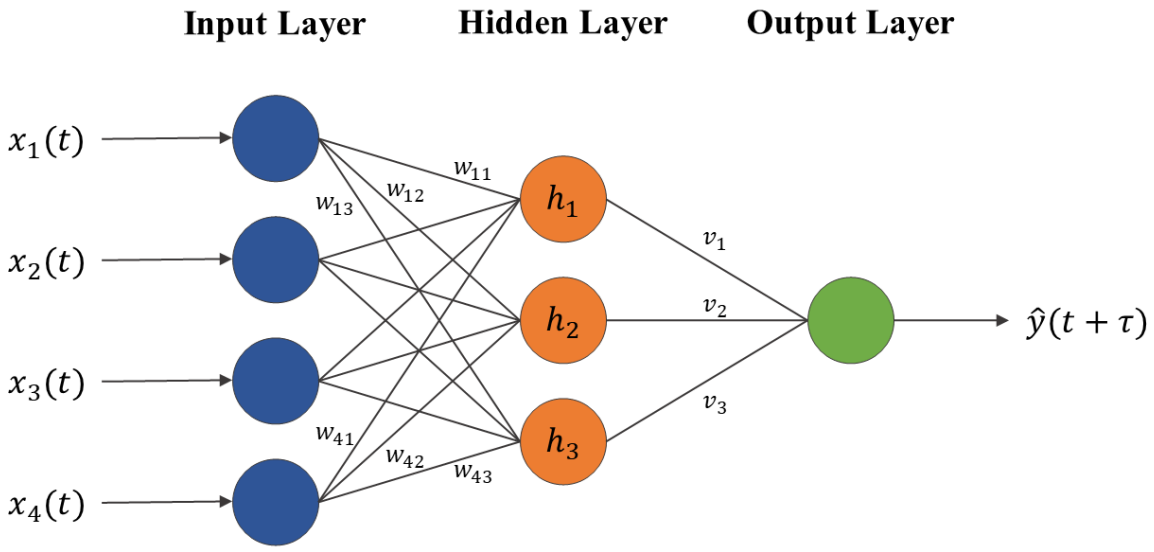

Figure 3.1: Illustration of a simple feed forward neural network architecture [44]

$$\vec{h} = \sigma(W\vec{x}) = \sigma\left(\begin{bmatrix} w_{11} & w_{21} & \cdots & w_{N1} \\ w_{12} & w_{22} & \ldots & w_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1M} & w_{2M} & \cdots & w_{NM} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}\right) = \begin{bmatrix} h_1 & h_2 & \cdots & h_M \end{bmatrix} \tag{16}$$

where $\vec{x}$ is the input vector, W is a matrix containing the input weights, and $\vec{h}$ is the vector containing the outputs of each neuron in the hidden layer. The input weight matrix is automatically tuned by the ANN during the model training process in order to optimize its prediction accuracy. It should be noted that an optional bias can be added to all neurons to further improve the ANN's performance. The Sigmoid function, which determines whether each neuron is activated or not, is defined as:

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{17}$$

where z in this case is the result of the weighted summation for a particular neuron. Finally, the output of an ANN can then be obtained through a weighted summation of the hidden neuron outputs:

$$\hat{y}_{t+\tau} = \vec{v} \cdot \vec{h} = \begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{bmatrix} \tag{18}$$

where $\vec{v}$ is an output weight vector connecting the hidden layer with the output layer. Similar to the input weight vector, the output weight vector is also tuned automatically during the model training process. It should also be noted that if the ANN has multiple output nodes, the output weight vector becomes a matrix.

In machine learning, the parameters which the programmer must specify are referred to as the model's *hyperparameters*. For an ANN, these include (but are not necessarily limited to) the:

- number of input nodes
- number of neurons in the hidden layer
- number of hidden layers (if more than 1, it becomes a *deep learning* model)
- type of activation function used
- learning rate (i.e., how fast the weights are tuned)

The *learned* parameters of an ANN are the input weight matrix (W) and the output weight vector ($\vec{v}$), which are tuned automatically during training.

3.2 Recurrent Neural Networks

LSTMs fall under the category of *recurent neural networks* (RNNs), which have a similar architecture to ANNs [44], [45]. However, as illustrated in Figure 3.2, the key difference between RNNs and ANNs is that the former includes a feedback element in the hidden layer. This feedback element enables RNNs to process sequences of time series data (i.e., current *and* past observations), unlike ANNs which can only process the current observation. As such, RNNs are able to account for historical information and sequential patterns in the data which contain useful information for making predictions.

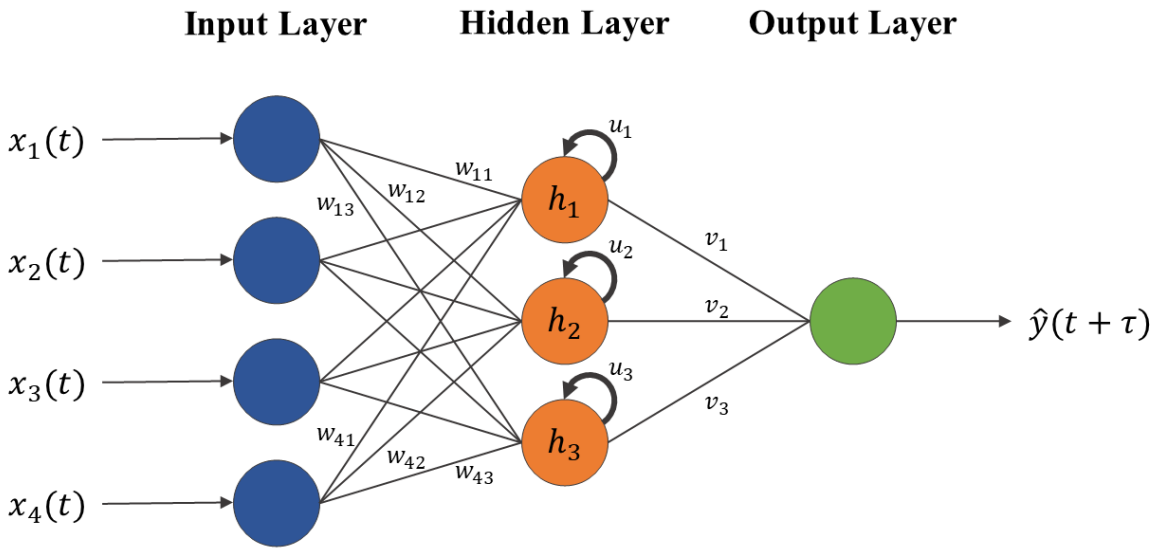

Figure 3.2: Comparison of recurrent vs feed forward neural network architectures [44]

For solar forecasting, it can be highly advantageous to use a sequence of input data rather than a single observation. For instance, if recent irradiance measurements are similar (i.e., persistent conditions) then it may be more likely that the future irradiance will also be similar. Alternatively, if recent irradiance measurements have been increasing or decreasing linearly, then the future irradiance may be more likely to follow this behaviour as well. Such trends would be largely unrecognizable with only a single input observation, giving RNNs an advantage over ANNs for solar forecasting. It should be noted, however, that there is a limit to the amount of historical data (i.e., the *lookback window* size) that will be useful for making predictions. For instance, the observed irradiance from 3 days earlier is unlikely to help an RNN predict the irradiance 5 minutes into the future. Providing too much historical data can degrade forecasting performance by overloading the model with extraneous information and increase computational resource demands needlessly.

Conventional RNN models have both a long-term memory aspect (i.e., the weight vectors and matrices) and a short-term memory aspect. The short-term memory is introduced by the feedback element, which provides information on each neuron's previous state. The output, h_t, of any given neuron, m, in the hidden layer is given by:

$$h_t = f(h_{t-1}, \vec{x}_t) \tag{19}$$

where h_{t-1} is the previous state (i.e., the embedded historical information) of the neuron and $\vec{x}_t$

is the current input vector. Similar to ANNs, the activation function, f, determines whether the neurons are activated or not; however, in an RNN the activation function takes both the current input vector and the previous state of that specific neuron as inputs. In an RNN, the most commonly used activation function is the hyperbolic tangent function which can be substituted into Eq. (19) to give:

$$h_t = \tanh\left(u_m\, h_{t-1} + \vec{w}_m \cdot \vec{x}_t\right) \tag{20}$$

where u_m is the weight applied to the feedback element of the neuron and $\vec{w}_m$ is the input weight vector connecting the inputs to the neuron. Finally, the output of an RNN can be obtained as:

$$\hat{y}_{t+\tau} = \vec{v} \cdot \vec{h}_t = \begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{bmatrix} \tag{21}$$

where $\vec{v}$ is the output weight vector and $\vec{h}_t$ is a vector containing the outputs of each neuron in the hidden layer. Similar to ANNs, the RNN model automatically learns the input and output weights; however, RNNs must also learn the optimal weights of the feedback elements for each neuron.

3.2.1 LSTM

Having outlined the relevant concepts of ANNs and RNNs, this section describes the theory behind the LSTM models [46] used for solar forecasting in Chapter 6 of this thesis. LSTM models have similar architectures to RNNs as they incorporate a feedback element which allows them to process sequences of past data. However, while RNNs use basic weighted-summation neurons in the hidden layer, LSTM networks instead use LSTM cells, which help avoid some of the limitations of traditional RNNs and improve the overall forecasting performance. The LSTM cell contains an input gate, an output gate, and a forget gate, as shown in Figure 3.3.

The forget gate enables the LSTM cell to decide whether to retain or forget past information, meaning finite sequences of historical observations can be retained until the forget gate discards them to start a new sequence. The forget gate of a single LSTM cell can be expressed as:

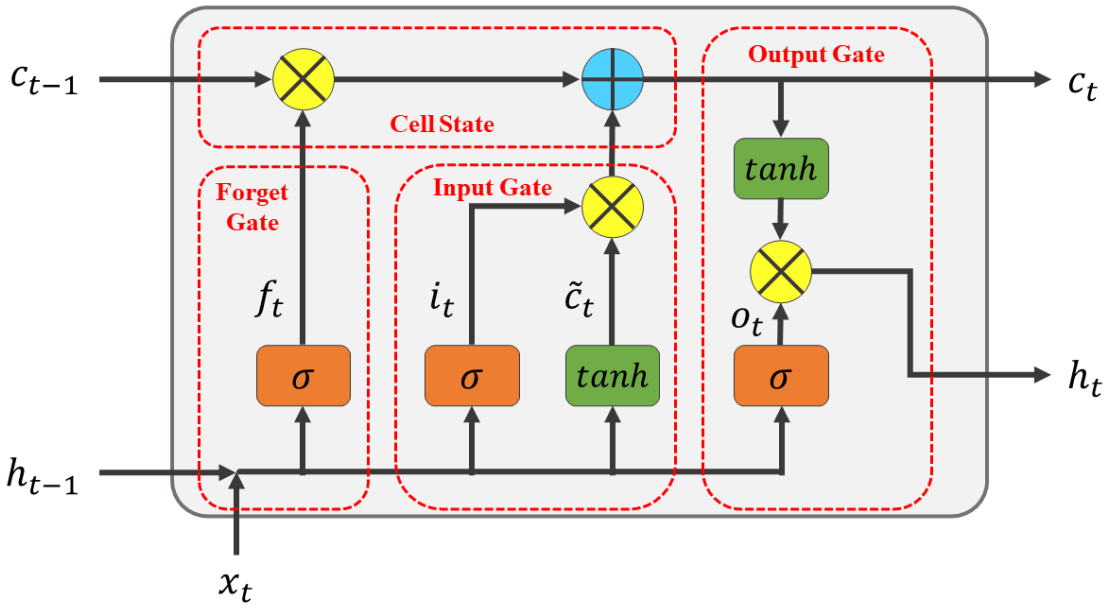

Figure 3.3: LSTM cell internal architecture [46]

$$f_t = \sigma(u_f\, h_{t-1} + \vec{w}_f \cdot \vec{x}_t) \tag{22}$$

where u_f is the weight connecting the previous state to the forget gate, $\vec{w}_f$ is the weight vector connecting the inputs to the forget gate, and (σ) is the Sigmoid function defined in Eq. (17). When all past information is to be forgotten (i.e., at the start of a new sequence), f_t will be close to 0. While f_t is close to 1, all past information since the last memory reset will be retained.

The input gate receives the current input vector and combines it with the previous hidden state in a weighted summation to update the cell state. It can be expressed as:

$$i_t = \sigma(u_i\, h_{t-1} + \vec{w}_i \cdot \vec{x}_t) \tag{23}$$

where u_i is the weight connecting the previous state to the input gate and $\vec{w}_i$ is the weight vector connecting the input variables to the input gate. The input gate is used to determine how much of the previous state and current inputs' values should be added to the new cell state using these weights. The same information is also often passed through a tanh operator to generate an intermediate cell state, $\tilde{c}$, which helps which helps overcome an issue encountered by RNNs known as the vanishing gradient problem [47]. The intermediate cell state is given by:

$$\tilde{c}_t = \tanh\left(u_c\, h_{t-1} + \vec{w}_c \cdot \vec{x}_t\right) \tag{24}$$

where u_c is the weight applied to the previous state and $\vec{w}_c$ is the weight vector applied to the inputs for generating the intermediate cell state.

The output gate combines the previous cell state with the current. As illustrated in Figure 3.3, the output gate state, o_t, can be expressed as:

$$o_t = \sigma(u_o\, h_{t-1} + \vec{w}_o \cdot \vec{x}_t) \tag{25}$$

where u_o is the weight connecting the previous state with the output gate and $\vec{w}_o$ is the weight vector connecting the inputs with the output gate.

Having defined the transfer functions for the forget gate, input gate, intermediate cell state, and output gate, we can obtain an updated cell state as:

$$c_t = (i_t\, \tilde{c}_t) + (f_t\, c_{t-1}) \tag{26}$$

and an updated hidden state (i.e., the LSTM cell output that will be passed to the network output) as:

$$h_t = o_t\, \tanh\left(c_t\right) \tag{27}$$

Finally, the output of an LSTM network can be obtained as:

$$\hat{y}_{t+\tau} = \vec{v} \cdot \vec{h}_t = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix} \tag{28}$$

where $\vec{v}$ is the output weight vector and $\vec{h}_t$ is a vector containing the outputs of each LSTM cell in the hidden layer.

During the training process, LSTM models must learn all of the weights for each cell in the network. As Eq. (22) to (27) are for a single LSTM cell in the hidden layer, in a full LSTM network – which has the same architecture an RNN – all of the weight vectors in these equations become weight matrices and the feedback weights become vectors. Therefore, LSTM models must derive nine optimal weight matrices and vectors, namely W_f, W_i, W_o, W_c, $\vec{u}_f$, $\vec{u}_i$, $\vec{u}_o$, $\vec{u}_c$, and $\vec{v}$.

3.3 The XGBoost Model

Extreme gradient boosting, or XGBoost, is a type of gradient-boosted decision tree algorithm which has several advantages over other types of machine learning models [48]. Unlike LSTMs, XGBoost does not rely on computationally expensive like Sigmoid or hyperbolic tangent functions to be sequentially executed, making it much more lightweight and significantly reducing training time. Since it is based on regression trees, XGBoost can even provide information about which inputs (features) it found to be most useful for making predictions, unlike neural networks which are black box models. In this section, the underlying principles of regression trees and tree-based ensembling techniques are described to provide necessary foundational knowledge. Then, the XGBoost architecture and operational theory is outlined and its essential parameters are described.

3.3.1 Decision and Regression Trees

Decision trees are an supervised machine learning algorithm which accept a set of input variables and make a series of decisions to classify, or categorize, an output [49], [50]. If a decision tree is predicting a numeric value rather than a class, it is instead referred to as a *regression tree*. Since solar forecasting involves predictions of numerical values, this section focuses on regression trees; however, many of the principles apply to both decision and regression trees. In either case, these models are structured in a top-down hierarchy of decision making stages, as shown in Figure 3.4. In this architecture, the decision nodes which split the data the most are near the top, while the decision nodes which make more acute splits on smaller data subsets are near the bottom.

The input variables are accepted by a *root node*, which makes the first decision based on the input variable that most significantly splits the data. The outputs of the root node are then passed along *branches* to *internal nodes* that further divide the data. The internal nodes, which may be

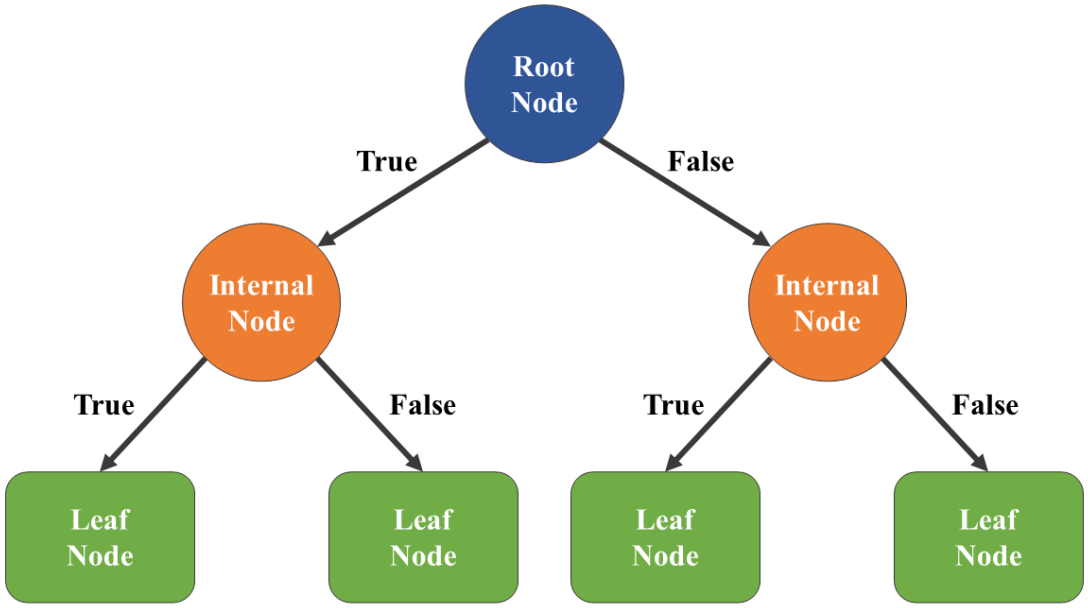

Figure 3.4: Illustration of a decision and/or regression tree structure [49]

several layers deep, divide the data as much as needed to make accurate predictions. The last internal nodes in the hierarchy pass their outputs to the *leaf nodes* (or *terminal nodes*), which contain all of the possible output options. While the leaf nodes in a decision tree represent the different classes into which the data is being divided, the leaf nodes of a regression tree represent different possible prediction values. The more leaf nodes there are in a regression tree, the higher the prediction resolution; however, regression trees which are too complex may be susceptible to overfitting the data.

The objective of a regression tree is to ensure that the most appropriate decision path is taken to select the most likely output, given a particular set of input values. This is illustrated in Figure 3.5, where a series of decisions are made based on the values of different input variables, leading to an ultimate prediction of the target variable. The decisions made by the root and internal nodes of a regression tree use numeric thresholds, which are learned by the model during the training process to optimize the prediction accuracy.

Regression trees split the data into groups that are most likely to lead to similar future values of the target variable. Each group corresponds to a different leaf node, which contains the prediction value. Naturally, if there are very few leaf nodes then there will be very few possible prediction values; hence, the model will likely underfit the data and perform poorly. However,

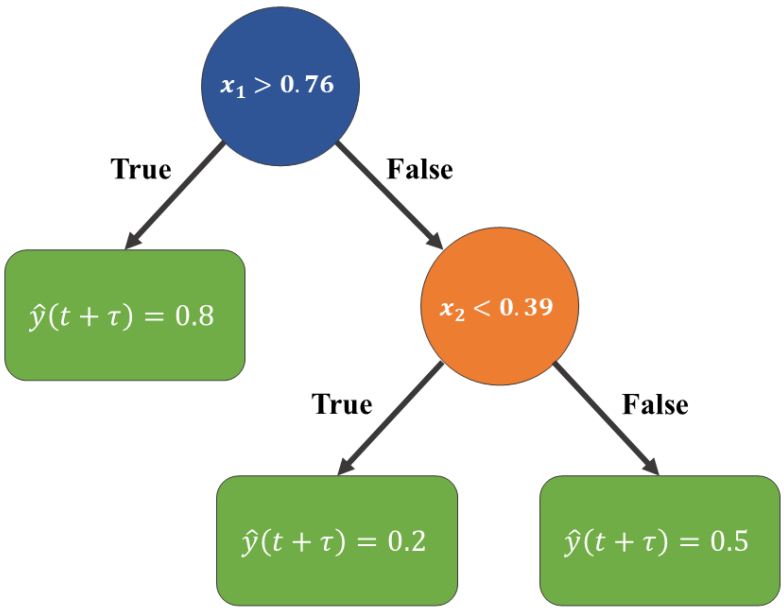

Figure 3.5: Example of a regression tree [49]

underfitting is generally avoidable if the model is provided with a sufficiently high volume of data during its training. In contrast, with too many leaf nodes the data groupings will become too acute and lead to overfitting and poor generalization to new data. While the model should include enough leaf nodes to achieve sufficient prediction accuracy, it is important to impose a maximum limit on the number of leaf nodes (which also limits the number of internal nodes). Overfitting can also be avoided by imposing a rule on internal nodes such that a minimum number of observations (datapoints) fit each output of that node. This eliminates any internal nodes which do not make significant splits of the data.

During the training process, regression trees use a greedy top-down approach known as binary recursive partitioning to select split thresholds at each decision node. This means that the tree successively optimizes the thresholds, beginning with the root node and working its way down the hierarchy. Thus, the splits at the top will affect the thresholds of nodes further down; however, the splits at nodes further down the tree will not affect the thresholds of nodes higher up.

To select a split threshold at the root node, the model takes the entire training set and tries every value of every input variable as the threshold. The algorithm computes the sum of squared errors (*SSE*) between all values in each resulting partition and the partition means:

$$SSE = \sum_{i \in R_1} (y_i - \mu_1)^2 + \sum_{i \in R_2} (y_i - \mu_2)^2 \tag{29}$$

where y_i is the value of each observation in the partition, μ_1 and μ_2 are the partition means, and R_1 and R_2 are the subsets of the input data resulting from the split. Note that each y_i will appear in either the first term or the second term, but not both. The split threshold which minimizes the *SSE* is taken as the optimal root node threshold. The model then moves to the next highest internal nodes and repeats this process. For the internal nodes, the input observations will only be those contained in the partition being passed to that node rather than the full input dataset. Hence, internal nodes that are further down the tree will make more acute splits on smaller data subsets.

As previously mentioned, a limit should be imposed on the model to prevent it from growing too deep and overfitting the data. This limit, referred to as the stopping criterion, allows the model to be *pruned* to remove excessive or redundant leaf nodes. This can be done using cost complexity pruning (CCP), in which a penalty is applied to the objective function defined in Eq. (29). A commonly used form of CCP is weakest link pruning (WLP), which calculates a *tree score* for the fully trained tree and all differently pruned sub-tree versions of it as:

$$Tree\ Score = SSE + \alpha|T| \tag{30}$$

where α is a cost complexity parameter and T is the number of leaf nodes in the tree. For a tree with n leaf nodes, the tree score will be calculated for the full tree with $T = n$, the sub-tree with $T = (n-1)$ leaf nodes, the sub-tree with $T = (n-2)$ leaf nodes, and so on until only a single leaf node is left. The sub-tree with the lowest tree score is selected as the optimal model, since it will have the best balance between prediction performance and complexity. The cost complexity parameter, α, is a tuning parameter that can be determined using cross-validation techniques.

3.3.2 Tree-Based Ensembles

The simplicity of regression trees offers the advantages of being fast, lightweight, and transparent. The transparency of regression trees is one of the primary reasons for including a tree-based model in this thesis. In contrast to the neural networks described in Section 3.2, which are

black box models with unknown internal logic, tree-based models are considered to be *white box* models which have highly explainable internal logic.

However, a single regression tree on its own will not perform well in complex prediction problems like solar forecasting. A regression tree cannot simply be grown larger and more complex to improve its performance on complicated data – this will only lead to overfitting. Instead, most practical tree-based models use several grown trees in ensemble architectures which can generally be categorized as either a bagging [50] or boosting [51] ensemble. A third ensemble method is stacking [52]; however, stacking is merely an ensemble of bagged or boosted models (i.e., an ensemble of ensembles) which is not foundational for XGBoost. The bagged and boosted ensemble architectures are illustrated in Figure 3.6.

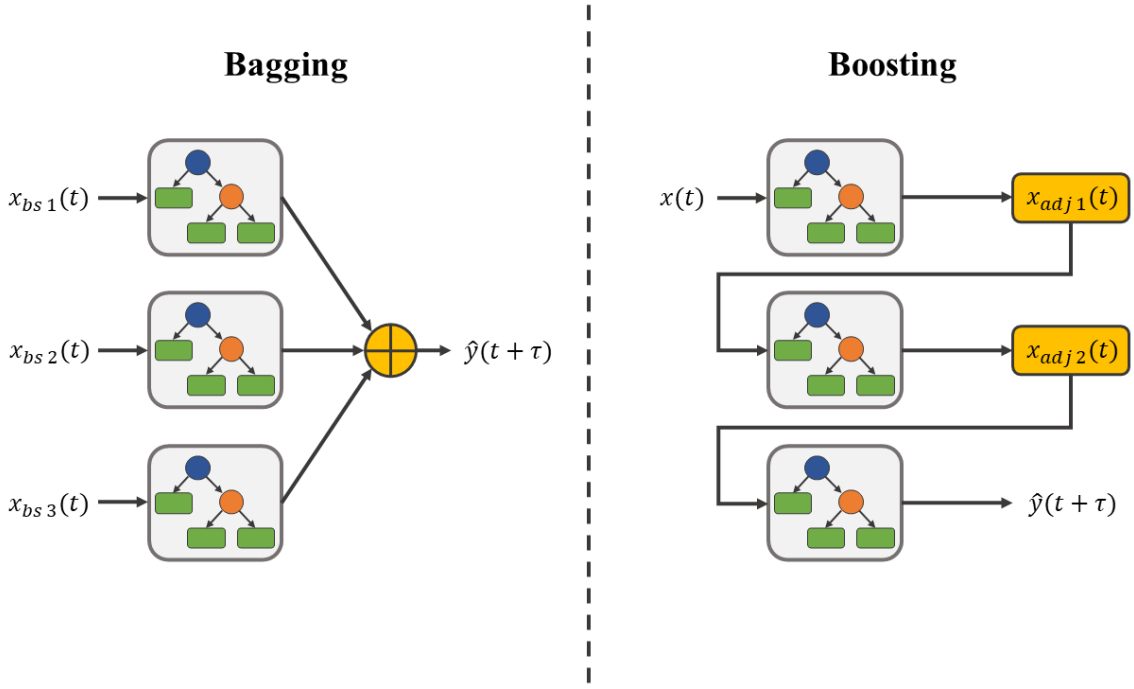

Figure 3.6: Illustration of bagging vs boosting architectures [49]

Bagging, or *bootstrap aggregation*, combines the outputs of several weak learners to form a strong learner ensemble model [50]. The weak learner models (i.e., individual regression trees) are trained separately on different subsets of the training data. These data subsets are generated using the bootstrapping sampling technique – or random sampling with replacement. After the individual trees are trained, new data is provided to their inputs and each tree makes its own prediction. The predictions are then aggregated such that the overall ensemble output is the average of all tree outputs.

Boosting differs from bagging in that it does not use bootstrap sampling, nor does it combine weak learners in parallel [51]. Instead, the trees are trained sequentially using information about the errors of the previous tree. This way, the each subsequent tree can take into account which training observations caused the previous tree to make larger prediction errors. Each observation in the training dataset is scaled, or weighted, according to how difficult it is to make predictions on. The concept behind the boosting ensemble method is that each subsequent weak learner will generally perform better than the previous ones. While the performance of some trees may be worse than their predecessors, the overall error trend across the sequence of trees should be decreasing.

3.3.3 Gradient Boosting

Gradient boosting, which underlies the XGBoost model, is a boosting method in which each tree directly predicts the errors of the previous tree in the sequence [53]. The first tree in a gradient boosting ensemble is often a single leaf node which uses the average value of the target variable as its prediction. The second tree in the sequence is trained on the prediction errors, or *residuals*, of the first tree. The predicted residuals that are output by the second tree are then scaled by a learning rate, which is a specified parameter that determines how much each sequential tree can adjust the ensemble output. The scaled error predictions from the second tree are then added to the first tree's prediction of the target variable, thus reducing the first tree's prediction errors. The third tree in the sequence is then trained on the residuals of these adjusted predictions, predicts the new residuals, and adjusts the target variable prediction again. This process is repeated either until the desired ensemble size is reached or until new trees do not improve the predictions.

The initial prediction of the target variable, which is output by the first tree in the ensemble, is taken as the single value that minimizes the prediction errors for all points in the training dataset. This initial value, denoted $\hat{y}_{initial}$, can be found as:

$$\hat{y}_{initial} = \underset{\hat{y}}{\operatorname{argmin}} \sum_{i=1}^{N} L(y_i, \hat{y}) \tag{31}$$

where L is a loss function which quantifies the prediction error, $\hat{y}$ is the single prediction used for the entire dataset, and y_i is the observed target value (label). While there are several to choose from, a commonly used loss functions is:

$$L(y_i, \hat{y}_i) = \frac{1}{2}(y_i - \hat{y}_i)^2 \tag{32}$$

where $\hat{y}_i$ is each predicted value, which in this initial step will be a constant value for all points in the training dataset. Multiplying by one half is commonly done to simplify the derivative that will later be computed, and it does not impact the final performance of the model. Substituting Eq. (32) into Eq. (31) yeilds:

$$\hat{y}_{initial} = \underset{\hat{y}}{\operatorname{argmin}} \sum_{i=1}^{N} \frac{1}{2}(y_i - \hat{y}_i)^2 \tag{33}$$

To find the value for $\hat{y}_i$ that minimizes the loss function, the chain rule can be used to find the partial derivative with respect to $\hat{y}_i$ as:

$$\frac{\partial L(y_i, \hat{y}_i)}{\partial \hat{y}_i} = -\sum_{i=1}^{N} (y_i - \hat{y}_i) \tag{34}$$

into which $\hat{y}_i = \hat{y}_{initial}$ can be substituted and solved as:

$$0 = -\sum_{i=1}^{N} (y_i - \hat{y}_{initial}) \tag{35}$$

$$N\hat{y}_{initial} = \sum_{i=1}^{N} (y_i) \tag{36}$$

$$\hat{y}_{initial} = \frac{1}{N} \sum_{i=1}^{N} (y_i) \tag{37}$$

Therefore, according to Eq. (37) the initial prediction output by the first tree should be the mean value of the target variable. As earlier indicated, this can be achieved by a single-leaf-node tree which outputs the mean value.

After initializing the ensemble output using a single tree, more trees can be sequentially added. In an ensemble with M trees (excluding the initial single-leaf-node tree), each subsequent tree takes the residuals of each previous tree's predictions, $r_{i,m}$, as its input:

$$r_{i,m} = -\left[\frac{\partial L(y_i, \hat{y}_i)}{\partial \hat{y}_i}\right]_{\hat{y}=\hat{y}^{m-1}} \quad \text{for } i = (1, ..., N) \tag{38}$$

where N is the number of samples in the training dataset. The partial derivative contained within the square brackets in Eq. (38) is referred to as the *gradient*, which is key to the *gradient boosting* method. Next, Eq. (34) can be substituted in for the gradient to yield:

$$r_{i,m} = \left[(y_i - \hat{y}_i)\right]_{\hat{y}=\hat{y}^{m-1}} \quad \text{for } i = (1, ..., N) \tag{39}$$

Using Eq. (39), the inputs to the second tree in the ensemble can be found using the residuals of the first tree. Likewise, the inputs to the third tree can be found using the residuals of the second tree, and so on. To further illustrate, since the first tree's prediction is $\hat{y}_{initial}$ (i.e., the mean) for all observations, the inputs to the second tree ($m = 1$) will be:

$$r_{i,1} = \left[(y_i - \hat{y}_{initial})\right] \quad \text{for } i = (1, ..., N) \tag{40}$$

The output leaf nodes of each sequentially added tree, which contain the possible predictions of the previous tree's residuals, are denoted as $R_{j,m}$, where j is the index for each leaf and m still denotes the index of the tree. For instance, the first leaf node of the second tree will be labelled $R_{1,1}$, the second leaf in this tree will be labelled $R_{2,1}$, and so on. The possible output values contained by each leaf node in a new tree, $\hat{y}_{j,m}$, can be found for a tree with J_m leaf nodes using:

$$\hat{y}_{j,m} = \underset{\hat{y}}{\operatorname{argmin}} \sum_{i \in R_{k,j}} L(y_i, \hat{y}_i^{m-1} + \hat{y}) \tag{41}$$

For instance, in the second tree ($m = 1$) with J_1 leaf nodes, the output values of each leaf node would be $\hat{y}_{j,1}$. The $(i \in R_{k,j})$ in the summation indicates that the output value for each leaf node is computed using only the inputs which reach that leaf node. That is, all samples which take paths through the tree that lead to other leaf nodes are not included in the calculation. The loss function from Eq. (32) can be substituted in to give:

$$\hat{y}_{j,m} = \underset{\hat{y}}{\operatorname{argmin}} \sum_{i \in R_{k,j}} \frac{1}{2}(y_i - (\hat{y}_i^{m-1} + \hat{y}))^2 \tag{42}$$

which can again be solved by setting the partial derivative to zero, as done for finding $\hat{y}_{initial}$. The result will be the optimal output value of leaf node j, which again happens to be the average value of the inputs that reached leaf node j. This computation must be done for each leaf node in each new tree.

Finally, the overall prediction made by the gradient boosting ensemble is given by:

$$(\hat{y}_{t+\tau})_i = \hat{y}_{initial} + \nu \sum_{m=1}^{M} (\hat{y}_i)_m \tag{43}$$

where ν is the learning rate which ranges from 0 to 1. As earlier mentioned, the learning rate defines how much of an impact each tree in the ensemble will have on the final prediction. As shown in Eq. (43), the initial prediction made by the first single-leaf-node tree is adjusted by each tree in the sequence to improve the ensemble's prediction accuracy. Naturally, the first few trees will result in larger adjustments since the early residuals predicted by these trees will be larger. Since the predictions get more accurate as the ensemble grows, the trees further down the sequence will make smaller and smaller adjustments.

3.3.4 XGBoost

XGBoost, or *extreme gradient boost*, is a tree-based ensemble model which is architecturally similar to the gradient boost method outlined in the previous section [48]. The main difference between these two ensemble models is that gradient boosting uses normal regression trees while XGBoost uses a different type of tree, which is referred to in this thesis as an *XGBoost tree*. XGBoost trees have better generalization to new data compared with gradient boosting due to their use of advanced regularization, which will be discussed further in this section.

Similar to gradient boosting, the first step for training an XGBoost model is to make an initial (constant) prediction that can be applied to all observations with a single-leaf-node tree. However, unlike gradient boosting which uses the mean of the target variable, the default initial prediction in an XGBoost model is 0.5. This value is arbitrary and can be defined otherwise; however, this will not impact the model performance.

The residuals of the initial XGBoost prediction are used to train a second sequential tree, as in gradient boosting. Each new XGBoost tree added to the sequence begins as a single leaf, and will grow if the algorithm deems it necessary. To decide if a tree should grow further, a *similarity score*, S, is computed using the residuals passed to the input of the tree:

$$S = \frac{1}{N+\lambda}\left(\sum_{i=1}^{N} r_i\right)^2 \tag{44}$$

where r_i is the residual of observation i, N is the number of observations in the training dataset, and λ is a regularization parameter. Since the residuals are added together before being squared, positive and negative values will partially cancel each other out. If the residuals in a leaf node are relatively different from each other (i.e., a balance of positive and negative residuals), the similarity score will be low since many will partially cancel each other out. In contrast, if the residuals are relatively similar to each other (i.e., most residuals are either positive or negative), the similarity will be higher. The regularization parameter, λ, is used to reduce the model's sensitivity to individual observations and avoid overfitting the training data. The regularization parameter decreases the similarity scores of all leaf nodes, but the amount of decrease is much greater in leaf nodes which contain a small number of residuals (N).

A tree can be grown to further divide the input residuals into two clusters of similar values using a greater-than-less-than threshold. The *gain* of adding two new leaf nodes with this threshold can be found by combining the similarity scores of each node as:

$$Gain = S_{(New\ Leaf\ 1)} + S_{(New\ Leaf\ 2)} - S_{(Root\ Leaf)} \tag{45}$$

This can be done for all possible threshold values, and the threshold which results in the highest gain will be selected since it splits the data into clusters with the greatest similarity. The tree can continue growing, further splitting the residuals into smaller and more similar groups. To prevent the tree from growing too large and overfitting the residuals, a maximum tree depth is imposed; by default this limit is set to 6 levels. Fully grown XGBoost trees can also be *pruned* to remove unnecessary leaf nodes which do not make impactful splits. A leaf node is pruned if its gain smaller than a fixed threshold value, γ:

$$\begin{cases} Gain < \gamma & \text{Prune leaf node} \\ \text{Else} & \text{Do not prune leaf node} \end{cases} \tag{46}$$

The output (residual prediction) values contained in a given leaf node, j, on the fully grown and pruned XGBoost tree can be computed as:

$$\hat{y}_j = \frac{1}{N_j + \lambda} \sum_{i \in R_j} r_i \tag{47}$$

where R_j is the subset of residuals which reach leaf node j, and N_j is the number of points in R_j. The inclusion of the regularization parameter, λ, helps decrease the impact of leaf nodes with few residuals (i.e., outliers) on the overall ensemble prediction. If λ is set to zero, the output value of each leaf node will simply be the average value of the residuals contained by that leaf node, as is the case in gradient boosting.

The overall output of the XGBoost ensemble is found in the same way as in gradient boosting:

$$(\hat{y}_{t+\tau})_i = \hat{y}_{initial} + \varepsilon \sum_{m=1}^{M} (\hat{y}_i)_m \tag{48}$$

where M XGBoost trees are sequentially added to the ensemble and ε denotes the learning rate, which by default is equal to 0.3. XGBoost and gradient boost are therefore the same architecturally; however, the use of regularization in XGBoost trees helps XGBoost avoid overfitting, thus achieving better performance and generalization. As a final note, given that each added tree in both XGBoost and gradient boost will contribute smaller and smaller adjustments to the overall prediction, a limit on the number of trees, M, should be imposed to prevent excessive model complexity and overfitting.

3.4 The 1-Dimensional Convolutional Neural Network Model

The third and final machine learning model implemented in this thesis is the 1-dimensional convolutional neural network (1D-CNN). The 1D-CNN is a type of feed-forward neural network architecture which performs 1-dimensional convolutions between a sequence of input data and a *kernel*, which will be discussed further in this section. A 1D-CNN uses scalar multiplication to compute the convolutions, which makes it significantly faster and more lightweight than the more commonly used – but computationally expensive – LSTM model.

3.4.1 2-Dimensional Convolutional Neural Networks

To understand the functionality of a 1D-CNN, it can be helpful to first consider the more well known 2-dimensional convolution neural network (2D-CNN) variant [54]. A 2D-CNN provides

an intuitive illustration of how convolutions in the spatial domain can be used in image classification problems. The architecture of a 2D-CNN is shown in Figure 3.7. In general, these models contain an input layer, a convolutional layer, a pooling layer, and an output neural network.

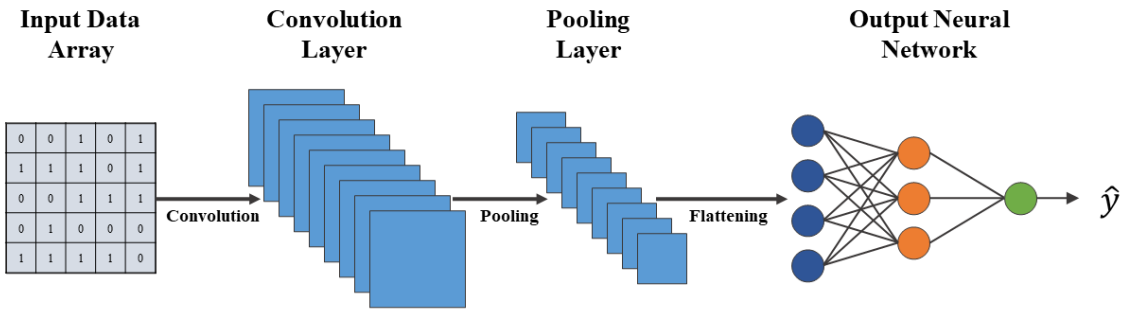

Figure 3.7: Architecture of a 2D-CNN model with one hidden layer [54]

A common application of the 2D-CNN is image classification. Images displayed on a computer screen are discretized (pixelated) into an array of pixels, where each pixel is assigned a value to represent its colour or brightness. For instance, in a black and white image, dark pixels may be assigned a value close to zero, light pixels may be assigned a value close to one, and pixels with varying shades of grey will be given a value between zero and one depending on their brightness. In image classification problems, 2D-CNNs attempt to learn the spatial patterns which correspond to different images (e.g., characteristic lines, curves, or shading patterns). The input layer of a 2D-CNN receives the x-coordinate, y-coordinate, and brightness value of each pixel in the array. The spatial information associated with each pixel is therefore retained, which is critical for image classification.

In the convolutional layer, the 2D-CNN uses a kernel which acts as a mask that can be applied to small subsets (i.e., spatial regions) of the input data. The model computes the convolution between the kernel and the first subset of the pixel array, which is selected by a sliding window that starts in the upper left corner of the original image. The convolution between the kernel and the subset of the pixel array is defined by the dot product:

$$(f * g) = \sum_{i=1}^{N} f_i \, g_i \tag{49}$$

where f is the subset of the pixel array covered by the sliding window, g is the kernel array, and N is the number of elements in both f and g (which are the same size and shape, by definition).

To better illustrate, Figure 3.8 shows the first convolution step, in which a sample 4x4 kernel is convolved with the first 4x4 subset of a sample pixel array. The result of the convolution between the 4x4 subset of the pixel array and the 4x4 kernel becomes the first element of the convolutional layer output, which known as the *feature map*.

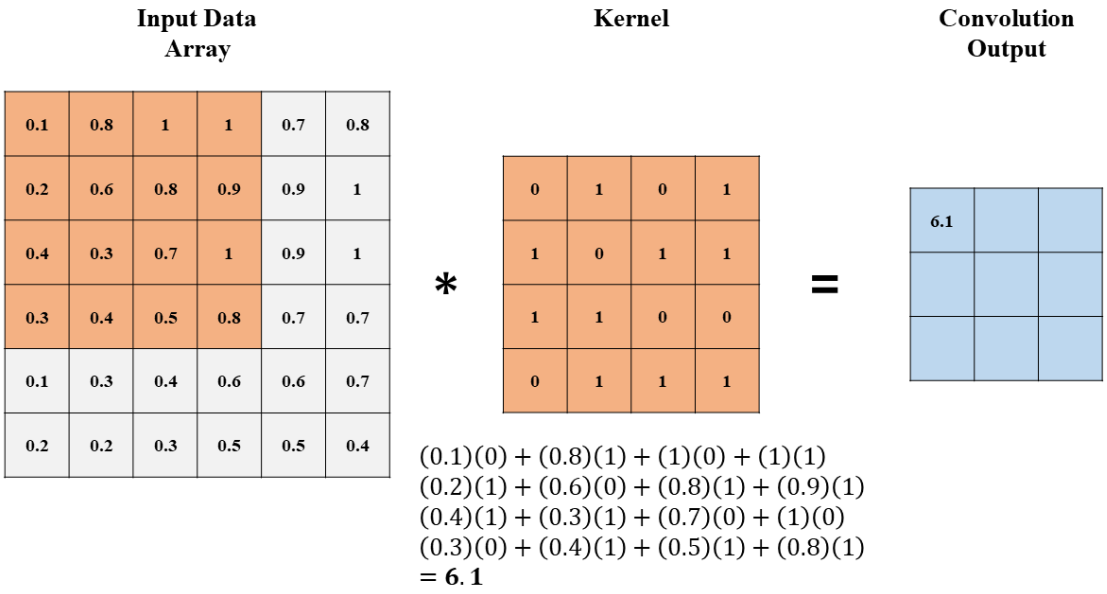

Figure 3.8: First convolution between a 2D array of data and a 4x4 kernel

The next 4x4 subset of the pixel array, which is found by shifting the sliding window to the right by one column, is then convolved with the kernel and the result is stored in the second element of the feature map, as shown in Figure 3.9.

The process of shifting the sliding window to the right and computing the convolution is repeated until the last column is reached. At this point, the sliding window is moved back to the left side of the pixel array and shifted down by one row, as shown in Figure 3.10. The process continues in this manner until all 4x4 subsets of the pixel array have been convolved with the kernel, generating a full feature map.

It is possible to add *padding* to the pixel array so that the kernel can scan the edges of the image more thoroughly, which may improve the model's performance. This concept is illustrated in Figure 3.11, in which an outer ring of zeros (i.e., padding = 1) is added to the array to allow the kernel to better identify patterns, or features, near the edges. Increasing the padding value adds more concentric rings of zeros around the edge.

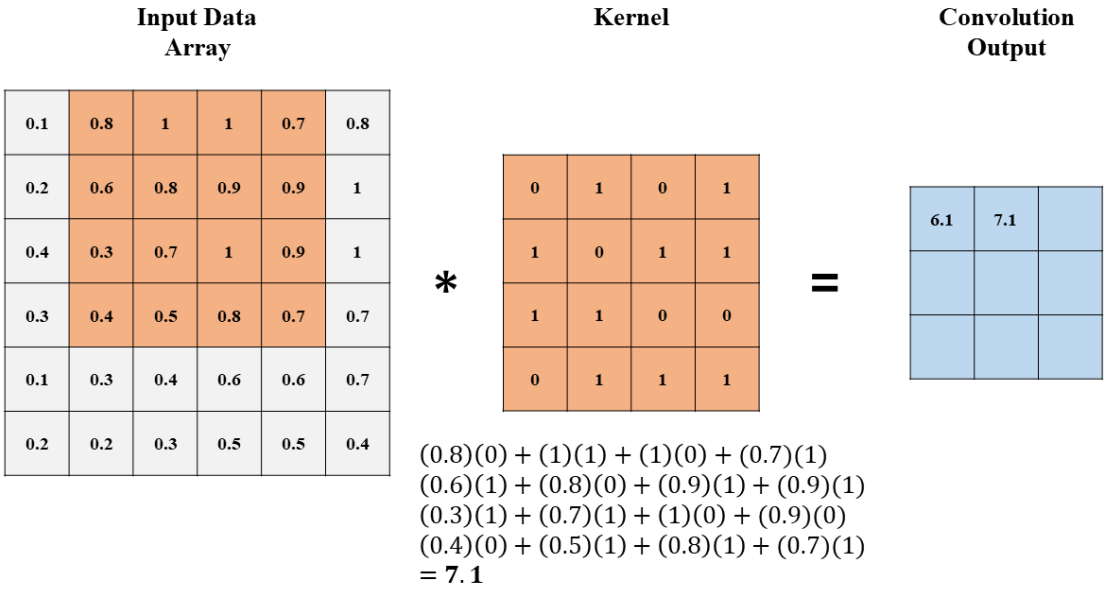

Figure 3.9: Second convolution between a 2D array of data and a 4x4 kernel

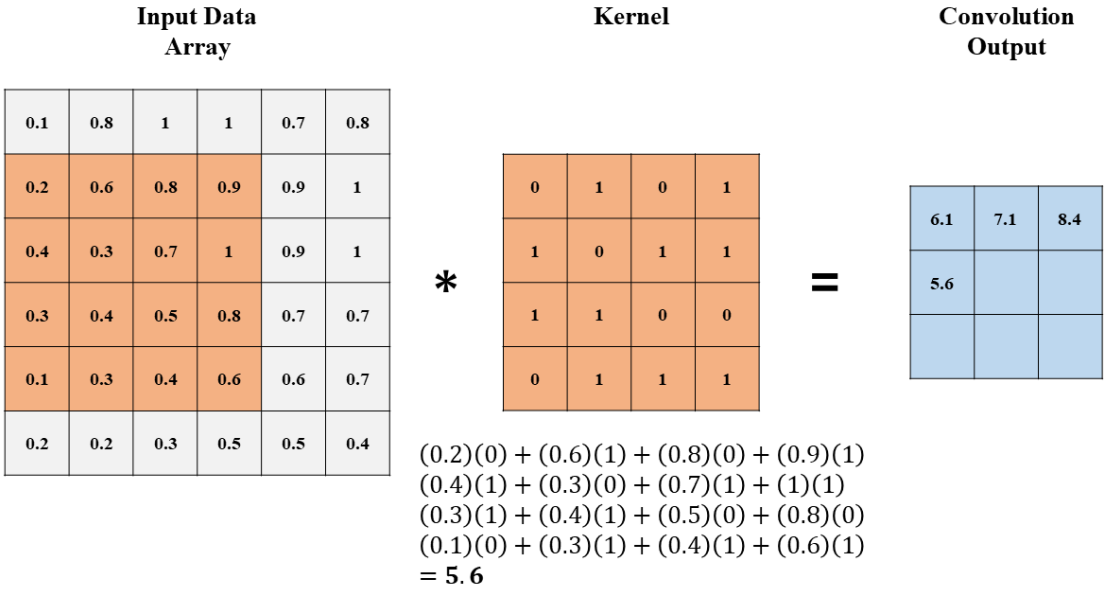

Figure 3.10: Fourth convolution between a 2D array of data and a 4x4 kernel

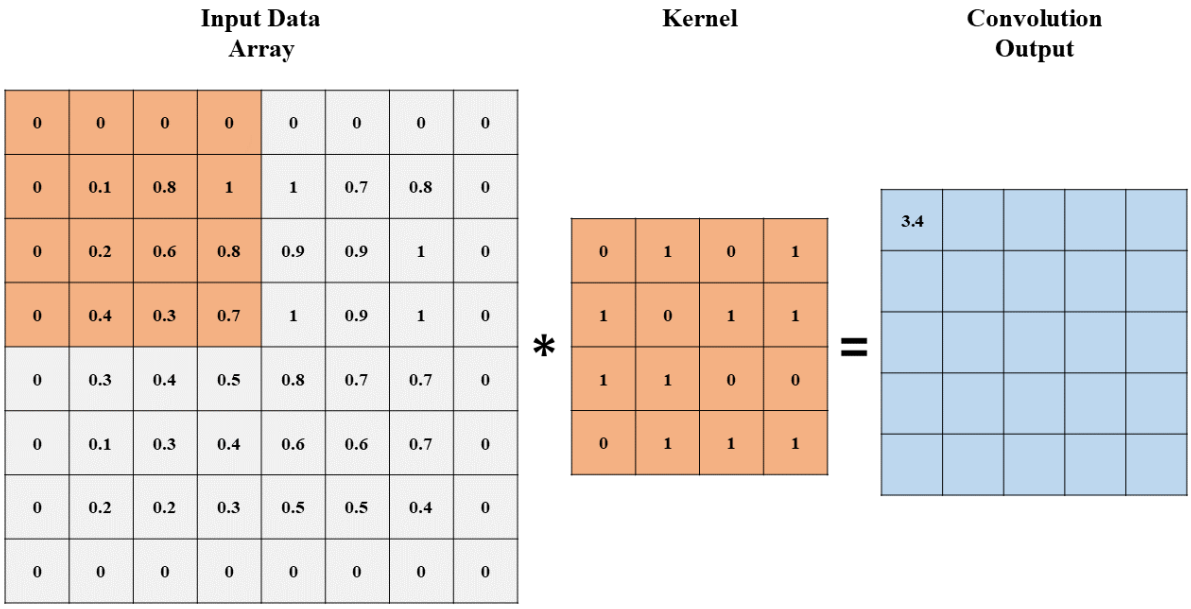

Figure 3.11: Effect of padding the 2D array of data before convolving with a 4x4 kernel

The shape of the feature map is determined by the input array and kernel sizes, the amount of padding, and the *stride* size (i.e., size of the steps taken when moving the sliding window around the input array). The height of the feature map is defined as:

$$H_{fm} = \frac{H_{input} - H_{kernel} + 2P}{S_{vertical}} + 1 \tag{50}$$

where H_{input} and H_{kernel} are the input array and kernel heights, respectively, P is the padding, and $S_{vertical}$ is the vertical stride or number of rows that the sliding window moves down after reaching the end of a row. Similarly, the width of the feature map is defined as

$$W_{fm} = \frac{W_{input} - W_{kernel} + 2P}{S_{horizontal}} + 1 \tag{51}$$

where W_{input} and W_{kernel} are the input array and kernel widths, respectively, and $S_{horizontal}$ is the horizontal stride or number of columns that the sliding window moves to the right at each step.

The feature map is then passed through a rectified linear (ReLU) activation function, which sets all negative values to zero and does not impact the positive values. The ReLU piecewise linear function can be formally expressed as:

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{52}$$

The rectified feature map is then passed to the pooling layer, where the size of the feature map is reduced to avoid overfitting and increase computational efficiency. The rectified feature map is divided into subsets of equal size and, typically, *max pooling* is applied. In max pooling, the maximum value in each subset is retained in the output as shown in Figure 3.12. The output of the pooling layer indicates how well each section (in this example, each quadrant) of the original pixel array matches the kernel that was used. A higher value indicates that the kernel had a higher correlation with the patterns found in that region of the pixel array, while a low value indicates the kernel had lower correlation. These higher and lower correlations between the kernel and the input array can be used to identify patterns that are characteristic of different image classes.

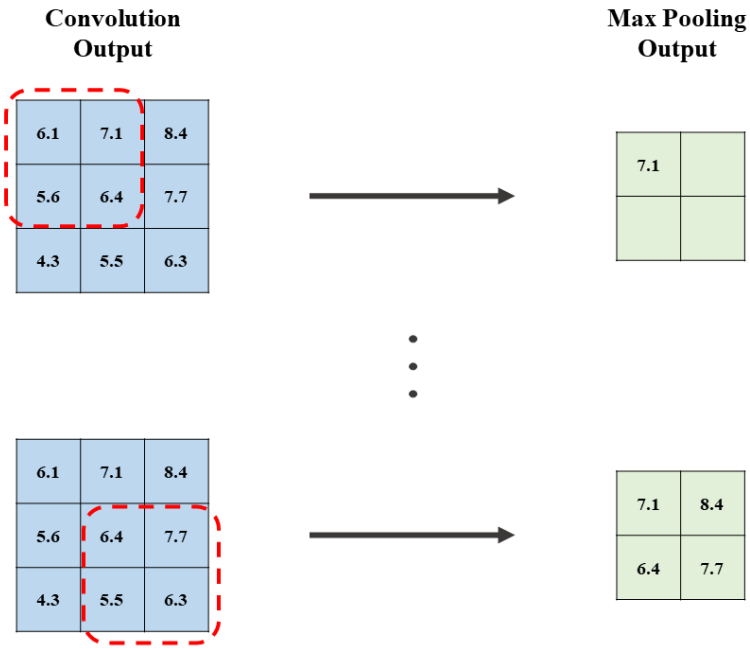

Figure 3.12: Illustration of max pooling

The output of the pooling layer can then be flattened into a 1-dimensional vector and fed to the input nodes of a standard artificial neural network. The neural network is then trained on the extracted (pooled) features of the original pixel data array to classify images. The values of each element in the kernel array are randomly initialized and then tuned during the training

process (analogous to the weights in an ANN). Thus, the learned parameters in a 2D-CNN are the optimal kernel values and neural network weights which result in the most accurate image classification.

3.4.2 1D-CNN

The 1D-CNN is a one-dimensional alternative to the 2D-CNN [55]. The 2D-CNN excels at solving problems involving 2-dimensional data, such as classifying images made of pixels with x and y coordinates (i.e., problems involving spatial information). In contrast, the 1D-CNN is better suited for 1-dimensional problems, such as making predictions with time series data (i.e., problems involving temporal information). The architecture of a typical 1D-CNN model is shown in Figure 3.13.

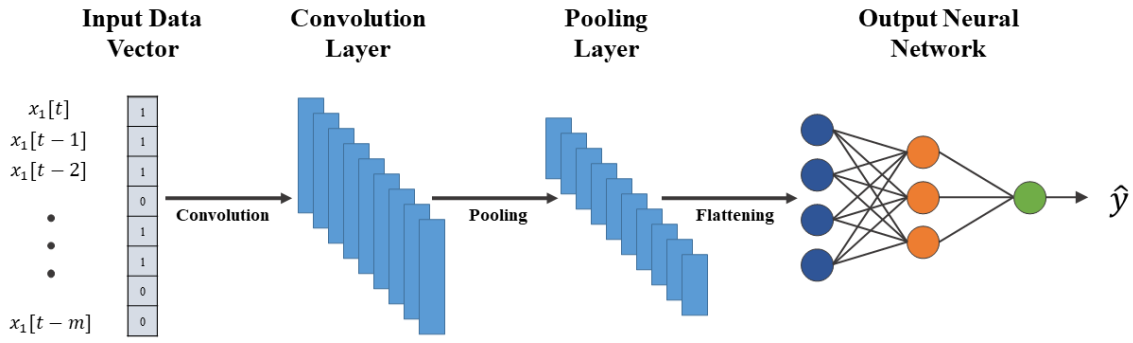

Figure 3.13: Architecture of a 1D-CNN model [55]

The first major difference between 1D-CNNs and 2D-CNNs is that the former uses discrete time convolution to convolve a 1-dimensional kernel with 1-dimensional time series data. Discrete time convolution is defined as:

$$(x * u)[n] = \sum_{m=a}^{b} x[m]\, u[n-m] \tag{53}$$

where x is a finite segment of the time series data which ranges from time a to time b, and u is a 1-dimensional kernel. The kernel can have any length, but it is generally shorter than the time series segment. The left hand side of Eq. (53) is illustrated in Figure 3.14.

The actual discrete time convolution computation from the right hand side of Eq. (53) is illustrated in Figure 3.15. Computing the convolution output, c, involves flipping the kernel

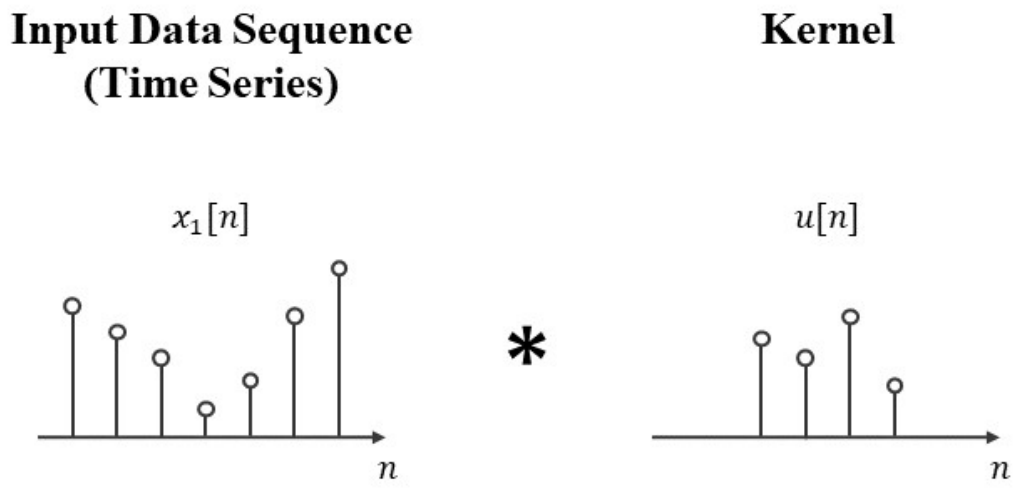

Figure 3.14: Sequence of time series data and 1D kernel

horizontally, shifting it by discrete time increments, and calculating the dot product between x and u at each step. It should be noted that adding zeros before the sample time series segment, $x_1[n]$, is called *causal padding* and is a 1D equivalent to the 2D padding earlier described. The output of the discrete time convolution is again a feature map. However, this feature map is 1-dimensional and is indicative of temporal patterns existing between neighbouring points, rather than spatial patterns in an image. Trends in the time series will therefore be identified and encoded in the feature map, provided that an appropriate kernel is used.

The second major difference between 1D-CNNs and 2D-CNNs is that the standard artificial neural network at the output of the 1D-CNN is trained to perform time series prediction, or *regression*, rather than classification. However, this difference does not change the structure or operation of the neural network; it merely changes its objective. Both CNN variants learn the optimal kernel values and neural network weights for their respective applications.

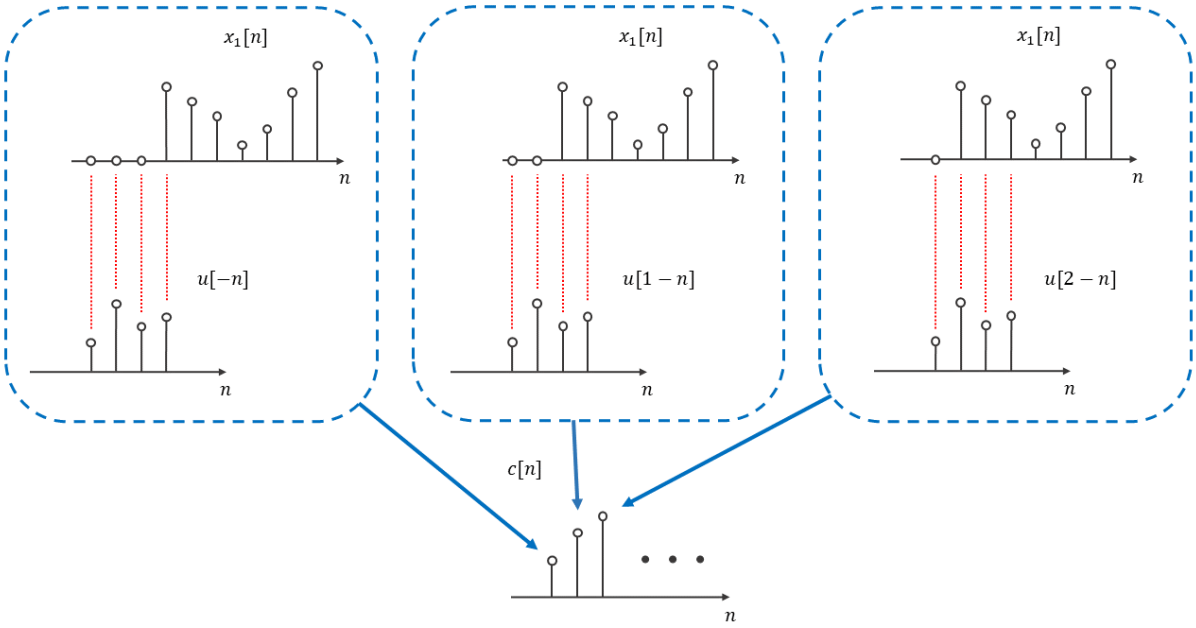

Figure 3.15: Visualization of discrete time convolution as a sliding dot product

Chapter 4: Spectral and Broadband Irradiance Variability

4.1 Assessment of Irradiance Variability Statistics

4.1.1 Scope and Impact

The following journal article manuscript assesses the spectral and broadband irradiance variability within the context of stationarity and solar forecasting applications. This paper, which uses one year (2022) of spectral and broadband irradiance from a custom Spectrafy SolarSIM-G, begins with an informative assessment of sky condition persistence and transition probabilities. Then, comparisons of seasonal and monthly clear-sky index distributions are presented, revealing longer-term temporal and spectral dependencies which violate the time invariance condition of strong stationarity. The seasonal and monthly means and variances are also analyzed, which indicate that assumptions of weak stationarity may be reasonable at these longer time scales. This paper also addresses short-term variability, which is of increasing concern in the context of operational solar nowcasting. An investigation of the short-term first-order statistics is performed using the spectral and broadband clear-sky index autocovariances. The autocovariances are fitted with the survivor function of the log-logistic distribution for the first time, to the best of the authors' knowledge. Finally, this manuscript provides an assessment of the second-order variability statistics using clear-sky index increment (forward difference) distributions. The distributions are found to be temporally dependent, exhibiting power law scaling (with distinct slope transitions) across time steps spanning several orders of magnitude. The distributions and scaling behaviours are also found to be spectrally dependent, with different spectral components probing different atmospheric scattering and absorption processes.

4.1.2 Author Contributions

Nick Anderson: All data analysis included in this article and lead-writing was performed by me, but was guided and supported by my co-authors.

Viktar Tatsiankou: As a collaborator from Spectrafy, Viktar has set up the instrumentation at the Ottawa site and provided me with the database used in this work. His previous research contributions have provided the foundations for this work, including the sky condition classification approach. He also assisted in the manuscript review process.

Karin Hinzer: As the director of the University of Ottawa's SUNLAB, Karin oversaw and guided my research. She also assisted in preparing and reviewing the final manuscript.

Richard Beal: As a collaborator from Spectrafy, Richard has set up the instrumentation at the Ottawa site and provided me with the database used in this work. He also assisted in the manuscript review process.

Henry Schriemer: As my supervisor and the associate director of the University of Ottawa's SUNLAB, Henry guided my research and assisted with interpreting results and preparing the final manuscript.

4.1.3 Publication - To be submitted 2023

Nick Anderson, Viktar Tatsiankou, Karin Hinzer, Richard Beal, and Henry Schriemer, "Statistical Assessment of Atmospheric Variability using Local Solar Spectral Irradiance", to be submitted in 2023.

Statistical Assessment of Atmospheric Variability using Local Solar Spectral Irradiance

Nick Anderson[a], Viktar Tatsiankou[a,b], Karin Hinzer[a], Richard Beal[b], Henry Schriemer[a]

[a] SUNLAB, University of Ottawa, Ottawa, ON, Canada

[b] Spectrafy, Ottawa, ON, Canada

Abstract

We perform a statistical assessment of a new broadband and spectral irradiance database containing nine spectral and one broadband irradiance time series with 250 ms resolution, collected by a custom Spectrafy SolarSIM-G located in Ottawa, Canada. The irradiance measurements are detrended of their deterministic diurnal and orbital trends using the clear-sky index. The first order statistics of the spectral and broadband clear-sky indexes are investigated on seasonal, monthly, and sub-hourly time scales, revealing nonstationary temporal dependencies. The sub-hourly spectral and broadband autocovariances are exceptionally well-fitted to the survivor function of the log-logistic distribution, which is, to the best of our knowledge, the first time this has been noted. The fitted log-logistic scaling and shape parameters are found to be spectrally dependent, the latter consistent with a simple extinction model and Ångström power law scaling. The second order statistics are assessed via clear-sky index increment distributions, which reveal spectrally dependent power law scaling in the peak probability densities on time scales consistent with nonlinear contributions.

Keywords: Spectral irradiance, irradiance variability, autocovariance, SolarSIM, clear-sky index increment

1. Introduction

The global installed capacity of solar photovoltaic (PV) energy resources has seen significant growth in recent years, reaching over 1 TW in 2022 [1]. However, solar irradiance variability caused by changing sky conditions leads to fluctuations in photovoltaic (PV) power, so variability considerations are increasingly a concern. Understanding irradiance variability is critical to quantifying uncertainty in its prediction, but despite this being a key stakeholder requirement, this area still suffers from a lack of research [2]. The salient features of the irradiance time series therefore need to be understood within a forecasting context [3].

There are two approaches to forecasting: physical and statistical. It is the fundamentals of the latter with which we are here concerned. Statistical methods encompass well-established approaches based on time series analysis as well as more recent machine learning techniques. Care, however, must be taken when applying statistical approaches to the time series output of a complex dynamical system. This is not just a question as to whether the data can be described by parametric or nonparametric statistics, it is a concern with how fundamental uncertainties are treated [4]. An operational consequence of this is the move from deterministic to the probabilistic treatments – from a single point-valued output to the likelihood of outcomes as a set of possibilities [5], [6].

Probabilistic prediction informs decision-making under uncertainty [7]. While typically expressed in terms of a single probability distribution, such predictions fail to distinguish between two inherently different sources of uncertainty, the aleatoric and the epistemic. Aleatoric uncertainties are those which have stationary statistical variation. Epistemic uncertainties are those that do not have such characteristics on the time scales used for model calibration and evaluation. The former is irreducible, caused by randomness; the latter is reducible, caused by ignorance of the system dynamics [8]. The distinction is also becoming increasingly important for machine learning in general, where uncertainty is a key element of the methodology [9]. It also motivates a more cautious statistical assessment of the solar irradiance time series as a precursor to solar forecasting.

Insight into epistemic uncertainty may be viewed by considering the nature of the processes that drive the irradiance dynamics. They are multiscale, encompassing celestial mechanics and complex

nonlinear atmospheric dynamics. In the former, predictability does not depend on initial state – diurnal and orbital variation can be described by simple deterministic representation, and the extraterrestrial solar irradiance can be known definitively at any time above any location, effectively without uncertainty.

By contrast, as the extraterrestrial irradiance is filtered through the atmosphere, its predictability becomes dependent on initial state [10]. This is because, while the evolution equations that describe atmospheric dynamics are formally deterministic, the system response – weather – is chaotic. Description ranges, in a computational sense, from the deterministic to the nondeterministic, and admixtures thereof, depending on conditions. For clear-sky irradiance – in a cloudless, static state with nominal atmospheric conditions – deterministic treatments are possible but complex, as calculation for any date, time and location requires a detailed physical model. Although some ambiguity is introduced by model choice due to methodological difficulties in treating climactic and atmospheric conditions [11], [12], epistemic uncertainty is largely eliminated. Deterministic processes may be described as known knowns.

Nondeterministic processes range from those which lack probabilistic quantification to those which can be described by a stochastic representation. Assuming the existence of an irradiance time series, we ignore the former and are thus forced to apply an epistemic lens to stochastic representations, which means we must begin with the presumption of nonstationarity. Because the solar irradiance is composed of both direct beam and diffuse components, variations under any circumstances other than near-clear sky conditions are markedly affected by the motion of clouds throughout the entire atmospheric hemisphere. Both *how* clouds interact with light and *where* they interact with light determine the statistical properties. These arise from the interplay of various processes: for the former, the narrowband absorption and scattering processes, and for the latter, the broadband distribution of diffuse and direct irradiance. Since weather is chaotic, its predictability may depend on initial state and time horizon. We should therefore not expect it, and thus certain detrended irradiance metrics, to be strictly stationary.

In what follows, we first provide a brief reminder of the fundamentals of time series analysis, the relevant statistical tools used for its quantification, and the decomposition procedures and metrics used to grant greater insight into the processes that drive the variation. Next is a review of the extant literature regarding the statistical treatments of irradiance time series in this context. We then describe our spectral irradiance measurement system, data acquisition, and validation. An extended statistical analysis is then given within a nonstationarity framework. As the literature below reveals, most approaches proceed within a stationary framework and consequently are forced to acknowledge nonstationarity without then having more carefully assessed its consequences. We consider some of these consequences. An ongoing discussion of results regarding the spectral dependence of irradiance variability is provided across extended time horizons. We conclude with some speculation regarding the relationship of these variations to the underlying physical processes and how a multi-spectral stochastic approach may be beneficial for PV forecasting.

2. Fundamentals of time series data and analysis

2.1 Stochasticity

A stochastic process is a collection of random variables, typically indexed by time, which have probabilistic quantification. As such, the fundamental algebraic definitions of stochastic processes are largely synonymous with those of temporally distributed random variables [13]. The *cumulative distribution function* (CDF), denoted as $F(x,t)$, for a stochastic process, $X(t)$, can be expressed as:

$$F(x,t) = P\{X(t) \leq x\} \tag{1}$$

where the right-hand side is the probability that the value of $X(t)$ will be below or equal to x. The *probability density function* (PDF) of a stochastic process, denoted as $f(x,t)$, represents the likelihood that a sample of the stochastic process will be equal to any given value. The PDF is defined as the partial derivative of the CDF with respect to x:

$$f(x,t) = \frac{\partial F(x,t)}{\partial (x)} \tag{2}$$

The expectation $E[X(t)]$, or statistical mean $\mu_X(t)$, of a stochastic process is defined as:

$$E[X(t)] = \mu_X(t) = \int_{-\infty}^{\infty} x f(x,t) dx \tag{3}$$

and the variance, $\sigma^2(t)$, which is the squared deviation from the mean, is:

$$\sigma^2 = E\left[\left(X(t) - \mu_X(t)\right)^2\right] \quad (4)$$

Besides the above fundamental definitions, an additional statistical quantity which is key for time series analysis is autocovariance. The autocovariance of a stochastic process, $K_{XX}(t_1, t_2)$, which measures the linear dependence of two samples at times t_1 and t_2, is expressed as:

$$K_{XX}(t_1, t_2) = E\left[\left(X(t_1) - \mu_X(t_1)\right)\left(X(t_2) - \mu_X(t_2)\right)\right] \quad (5)$$

In completely random, uncorrelated time series (e.g., white noise), the autocovariance will be very close to zero. However, temporal dependencies across different time scales will result in higher autocovariances at corresponding lags (i.e., separation between t_1 and t_2).

2.2 Time dependencies in stochastic systems

Time series generated by stochastic processes can be broadly categorized as either stationary or nonstationary. Stationary processes are those in which the statistical behavior does not change with time – they are absent of trends, periodicity, and seasonality. While strong stationarity requires the statistical distribution of the time series to be time-invariant, weak stationarity relaxes the requirements to constant mean, constant covariance, and finite variance. Stochastic processes in which significant time dependencies exist do not meet the requirements of strong or weak stationarity, and hence are considered nonstationary. While there are many ways in which nonstationarity can appear in a time series [14], the general forms are illustrated in Figure 1.

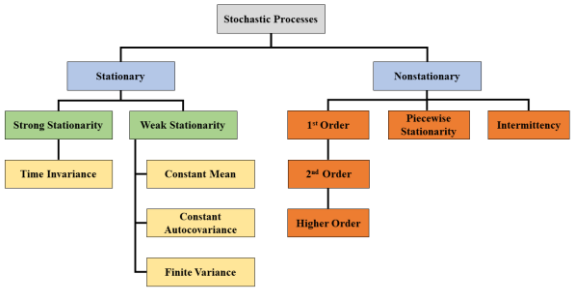

Figure 1. Classification of time series dependencies in stochastic systems

2.2.1 Nonstationary behavior

Nonstationarity can be the result of first order (first moment), second order (second moment), or higher order behavior. First order nonstationarity commonly appears as variations in the mean over time, which are caused by increasing or decreasing trends and seasonality. Second order nonstationarity behaviors include periodicity, changes in the covariance over time, and periods of varying volatility. Identifying higher order non-stationarity typically requires metrics pertaining to the shape of the system's statistical distributions, such as skewness and kurtosis [14].

Piecewise stationarity is a form of pseudo-stationarity which appears in some time series [15]. In piecewise stationary processes, the conditions of weak stationarity may be satisfied in local regions of the data, but not across the entire time series. In this case, the entire time series can be considered a collection of weakly stationary segments with instantaneous transitions. Similarly, a time series which alternates between periods of stability and variability is said to be intermittent, which is another form of non-stationary behavior. Despite meeting the conditions of weak stationarity in local segments, piecewise-stationary and intermittent time series are ultimately nonstationary and their consequent implications must be considered.

2.2.2 Weak stationarity

For a time series to be weakly stationary, its first moment (i.e., statistical mean) and autocovariance must be shift-invariant, and its second moment must be finite. Formally, a weakly stationary time series must meet the following conditions:

$$E[X(t)] = E[X(t+\tau)] = \mu_X \quad (6)$$

$$K_{XX}(t_1, t_2) = K_{XX}(t_1 - t_2, 0) = K_{XX}(\tau) \quad (7)$$

$$E[|X(t)|^2] < \infty \quad (8)$$

The first condition implies that the process must maintain a constant mean, regardless of location within the time series. The second condition implies that the autocovariance of a weakly stationary time series depends only on the distance between points (i.e., the lag, τ). This, in turn, also implies that the autocovariance of a weakly stationary process is constant for a given lag. Finally, the third property implies that weakly stationary time series have finite

variance at all locations, since the variance is the second moment about the mean. Hence, variations in the mean, variations in the autocovariance (for a fixed lag), and infinite variance are violations of the weak stationarity requirements.

2.2.3 Strong stationarity

A time series is strongly (strictly) stationary if its unconditional CDF time invariant. More formally, the CDF, $F(x,t)$, of a stationary process is unaffected when a sliding window probing the time series is shifted by some time step, τ:

$$F_X\big(x(t_1),\dots,x(t_n)\big) = F_X\big(x(t_1+\tau),\dots,x(t_n+\tau)\big) \quad (9)$$

All points in a strictly stationary time series are independent of each other, and any stochastic shocks to the system will quickly revert to the mean without causing permanent effects. The presence of trends, periodicity, or seasonality in the data – which would evolve the statistics and introduce temporal dependences between points – violates the requirements of strong stationarity.

2.2.4 Stationarity tests

Several methods have been proposed for testing stationarity, both qualitatively and quantitatively. Hyndman and Athanasopoulos [16] suggest several qualitative ways to test stationarity by visually identifying trends, seasonality, or periodicity. More quantitative approaches involving hypothesis tests, such as the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillis-Schmidt-Shin (KPSS) unit root tests, have been carried out by Ekstrom *et al.* [17] and Dong *et al.*[18]. In these studies, the unit root tests suggest that clear-sky index time series can be stationarized through careful data transformations. However, Yang [19] notes that these tests may be insufficient for identifying nonstationary attributes such as seasonality, since they only test whether the stochastic process contains a unit root. Instead, Yang proposes an approach to stationarity testing where conditional distributions of binned clear-sky index values are compared using the Kolmogorov-Smirnov (KS) test.

Rather than relying on hypothesis tests, we investigate long-term clear-sky index nonstationarity by comparing the seasonal clear-sky index empirical CDFs, or ECDFs. This comparison is done using the KS test, which provides a summary statistic determining whether two ECDFs are generated from the same distribution. The KS test summary statistic, $D_{n,m}$, corresponds to the largest absolute difference between two ECDFs:

$$D_{n,m} = \sup_{\kappa^*} \left| F_{1,n}(\kappa^*) - F_{2,m}(\kappa^*) \right| \quad (10)$$

where $F_{1,n}(\kappa^*)$ and $F_{2,m}(\kappa^*)$ are the two ECDFs being compared, with sample sizes n and m, respectively. For sufficiently large sample sizes, the two distributions are statistically different if:

$$D_{n,m} > \sqrt{-\frac{1}{2}\ln\left(\frac{\alpha}{2}\right)\left(\frac{n+m}{nm}\right)} \quad (11)$$

where the parameter α is used to determine what amount of difference between the ECDFs is statistically significant. As is common practice in the literature [19], we use $\alpha = 0.05$.

3. Data representations

3.1 Detrending: clear-sky indices

It is well understood that solar irradiance variability consists of deterministic and non-deterministic components. The nonstationary deterministic diurnal and orbital trends can be easily predicted by clear-sky models, whereas the non-deterministic cloud-induced variability is more challenging to characterize. The deterministic trends can be removed from the broadband global horizontal irradiance (GHI) time series via the clear-sky index, $\kappa^*(t)$:

$$\kappa^*(t) = \frac{I_{GHI}(t)}{I_{GHI,clr}(t)}\,, \quad (12)$$

where $I_{GHI}(t)$ is the measured broadband GHI and $I_{GHI,clr}(t)$ is the modelled broadband clear-sky GHI. As the deterministic trends are eliminated by this clear-sky normalization, the remaining variability in the clear-sky index time series is dominated by cloud dynamics. It is this clear-sky index time series, which is often presumed to be trend-stationary, that will be the foundation of the statistical assessments in the remaining sections of this work.

Similar to Eq. (12), which defines the broadband clear-sky index, the analogous spectral clear-sky index, $\kappa^*(\lambda;t)$, can be defined as:

$$\kappa^*(\lambda;t) = \frac{S_{GHI}(\lambda;t)}{S_{GHI,clr}(\lambda;t)} \quad (13)$$

where $S_{GHI}(\lambda;t)$ is the measured spectral GHI, $S_{GHI,clr}(\lambda;t)$ is the spectral clear-sky GHI, and λ is the center wavelength at which the spectral irradiance is measured. Using the spectral clear-sky index, the diurnal and orbital trends can also be removed from measured spectral GHI time series.

3.3 Differencing: index increments

To analyze higher-order statistical properties of cloud-induced variability, we quantify short-term irradiance variability using the clear-sky index increment, $\Delta\kappa_\tau^*(t)$, which is defined as the forward difference between two values in a clear-sky index time series:

$$\Delta\kappa_\tau^*(t) = \kappa^*(t+\tau) - \kappa^*(t) \quad (14)$$

where $\kappa^*(t)$ and $\kappa^*(t+\tau)$ are a given pair of clear-sky index values separated by a time step, τ. The clear-sky index increment evaluates the change in irradiance over a given time interval, which is set by changing τ. This differencing is also a data transformation known to promote stationarity. While the above definition is for the broadband clear-sky index increment, the spectral equivalent can be similarly defined as:

$$\Delta\kappa_\tau^*(\lambda;t) = \kappa^*(\lambda;t+\tau) - \kappa^*(\lambda;t) \quad (15)$$

for a given wavelength, λ.

4. Review of existing statistical approaches

The statistical properties of the irradiance time series have been extensively studied for over 60 years. Liu and Jordan [20] did formative assessments of the interrelationships and distributions of direct irradiance, diffuse irradiance, and total GHI using the clearness index. They found that the CDFs for monthly clearness indices with common averages were similar. They introduced parameterizations that allowed some generalization of results. This was followed up in more detail by Hollands and Huget [21] who more formally applied classical probability theory to their work, thus providing a general framework for analyzing solar data.

Gordon and Reddy [22] provided an analysis of the stationary and sequential properties of daily GHI on a monthly basis, using daily radiation data from locations of diverse climatic conditions. They showed the claim of previous studies, that climates with the same mean clearness index share similar PDFs, to be an artifact of analysis and limited sample size. Their approach was to produce what they described as "generalized stationary statistics" for multiyear data, normalizing the clearness index k with its monthly average clearness index $\bar{k}$ to produce $x = k/\bar{k}$ as a stochastic variable with daily, monthly and annual dependencies. They found that the erroneous insights arose because the multiyear average of monthly clearness indices and the corresponding average variance are correlated "in one way for temperate climates, and in a different way for tropical climates". Employing a heuristic technique, they developed as an empirical closed form:

$$P(x) = Ax^n[1 - (x/x_{max})] \quad (16)$$

where A, n, and x_{max} are interdependent fitting parameters determined from normalization, and knowledge of $\bar{k}$ and $\sigma^2(x)$. They also studied the persistence effects of the GHI with a day-lag autocorrelation approach, transforming x to a new stochastic variable z with zero mean and a unit standard deviation. They find wide variation in the dependence of both the one-day and two-day lag autocorrelation on average monthly clearness index, concluding via statistical tests that the former is distinguishable from zero whereas the latter is not. They followed up, in a subsequent paper [23], with an analysis of the hourly GHI. The dependencies of hour-lag correlations on the hour of the day were found to be strong, variable, nonlinear, nonmonotonic, and dependent on location; the general trend was a decreased correlation with increased lag. By comparison, the dependence of the daily average hourly autocorrelations on the lag hour was typically monotonic, although even more clearly locationally-dependent.

Aguiar and Collares-Pereira [24] undertook an approach initially similar to that of Gordon and Reddy. They too considered hourly and daily clearness indices, finding similar distribution results for the comprehensive ("mixed") dataset. However, once the hourly data was segmented by its dependence on daily indices and solar altitude, an *unmixed* analysis showed

different behavior. The unmixed hourly radiation does not possess stationarity, and if it is segmented by narrow daily clearness index bin ranges, the PDFs show no dependence on latitude and climate. They followed up on this approach in a companion paper [25] where a transformation similar to that of Gordon and Reddy was used to purportedly remove all trends. The dependence of the first- and second-order autocorrelation coefficients was determined. Note that, as with Gordon and Reddy, the autocorrelation was applied to a normalized *irradiance variability* time series (although referenced to different means and variances), not the normalized irradiance itself. It is not a test for stationarity but rather an enforcement of it for model application purposes.

Skartveit and Olseth [26] looked at one year of direct normal irradiance (DNI) and GHI data from measurements averaged across 5-minute intervals at three diverse locations. They showed observed and modelled PDFs of GHI and DNI based on their respective clear-sky indices, segmented by hourly averages. The PDFs were not unique functions of the hourly averages, but depended heavily on irradiance variability. Hourly analysis showed that data with low standard deviation had unimodal distributions, due to cloudless skies or uniform clouds, while increasing standard deviation gave bimodal distributions with increasing mode separation, due to increasingly broken skies. While the bimodal pattern was observed at all elevation angles, it was found to be more pronounced at high angles, especially for the DNI.

Woyte *et al.* [27] acquired multiyear GHI data from three sites with distinctly different climates at resolutions ranging from 1 to 5 seconds. CDFs for subsets of the clearness index with a pre-defined mean were determined for different values of the air mass, allowing regional comparison confirming previous findings. Interestingly, they note that to reduce the influence of air mass and local climate, the clear-sky index should be used to study the fluctuations introduced by passing clouds. They note that the clearness index, which detrends using extraterrestrial irradiance, could be interpreted as realizations of a stochastic process on the basis of having site-specific probability distributions determined by air mass and mean clearness index, with the order of the index sequence then a function of the autocorrelation of the stochastic process. They further note that, although the stochastic process is not stationary, a localized analysis limited to short time intervals, might be possible. They then demonstrated the applicability of a localized spectral analysis based on wavelet bases to decompose the fluctuating clearness index signal into a set of orthonormal subsignals, where each subsignal represents a specific fluctuation persistence scale.

Munkhammar and Widén [28] used the autocorrelation function of the clear-sky index to demonstrate how copulas could be used to model temporal variability. Data of an annual minute-resolution GHI dataset was restricted to a two-hour period centered on noon. Of interest here is the behavior of the autocorrelation. For this temporal restriction, the autocorrelation on the annual dataset showed a quasi-exponential decrease to zero over about 30 minutes, continuing negative until about 50 minutes and thence returning slowly to zero at 120 minutes. They also calculated the daily autocorrelations for the same temporal restrictions, which showed extensive diversity. The mean daily autocorrelation is seen to be visually indistinguishable from the annual autocorrelation. Presumably the former was calculated using daily mean values (which show order of magnitude variation over the year). The indistinguishability implies that their clear-sky index time series is indeed in some sense stationary.

Madanchi *et al.* [29] note that in a complex time series, $x(t)$, long-range correlations are usually characterized by scaling laws, where the scaling exponents characterize the underlying processes. Referencing [30], they state that for stationary data the correlation function $C(\tau)$ has short-range correlations, decaying as:

$$C(\tau) \sim \exp(-\tau/\tau_{max}) \tag{17}$$

where τ_{max} is a decay time. For long-range correlations, $C(\tau)$ follows a power law:

$$C(\tau) \sim \tau^{-\gamma} \tag{18}$$

with an exponent $0 < \gamma < 1$. Since the correlation function is a linear regression, they suggest that higher order statistical properties are needed to fully characterize the time series. If the increment of $x(t)$ over time τ is defined as:

$$\Delta x(\tau) = x(t+\tau) - x(t) \tag{19}$$

then the qth order absolute moment of $x(t)$ is:

$$S(q,\tau) = \langle |\Delta x(\tau)|^q \rangle \tag{20}$$

If the scaling behavior of this structure function for a certain range of τ is described by a power law such that:

$$S(q,\tau) \simeq C_q \tau^{\xi_q} \tag{21}$$

where C_q is a prefactor and ξ_q is the exponent of the power law with a linear dependence on q, then the process is called scale invariant. If ξ_q is a linear function of q, then $S(q,\tau)$ is called monofractal (linear); if a nonlinear function, then multifractal (nonlinear).

They used multifractal detrended fluctuation analysis to show that solar irradiance time series have strong nonlinear and nonstationary properties. Multiple GHI datasets, whose measurement duration ranged from 1 to ~10 years and whose resolution ranged from 1 s to 1 minute, were detrended using the clear-sky irradiance. Defining a generalized fluctuation function, they find scaling behaviour by means of a log-log plot. Their analysis shows gentle cross overs in scaling that are weakly location-dependent, at time scales $\tau_c \sim 450$ s. Deeper analysis suggests that multifractality is due to a combination of linear and nonlinear correlations and is a direct consequence of the frequency with which broken clouds occur.

5. The spectral irradiance measurement system

5.1 Instrumentation

We employ a custom Spectrafy SolarSIM-G spectral pyranometer, modified for rapid data acquisition, which collects spectral and broadband irradiance data in Ottawa, Canada. The SolarSIM-G measures the spectral GHI at nine wavelengths across the solar spectrum using nine channels of narrow bandpass filters paired with calibrated photodetectors. The center wavelengths of each channel within the solar spectrum are summarized in Table 1. Precise wavelength data for channels 1, 8 and 9 are not available; their centre wavelengths are specified by the manufacturer to lie within the given ranges. The spectral GHI is recorded at 250 ms resolution, which allows the irradiance variability to be sufficiently captured on all time scales relevant to PV generation. The SolarSIM-G is also equipped with a ventilator attachment to minimize the impacts of rain, dust, and snow on data quality.

The spectral GHI measurements, ambient pressure, relative humidity, and air temperature are combined using a radiative transfer model to self-consistently derive the atmospheric optical parameters and broadband GHI, which are also recorded at 250 ms resolution. The basic design principles of this spectral reconstruction process have been described in [31] and are summarized in section 5.2. Measurements from the infrared channels are used to classify the clouds by their water morphology, with a uniform cloud model employed to estimate cloud optical depth and to subsequently compute cloud transmittance in the 1000-4000 nm range. The efficacy of this approach has been validated against reference instruments under a full range of sky conditions [32], [33].

Table 1. SolarSIM-G spectral channels center wavelengths

Optical Channel	Center Wavelength (nm)	Targets
1	300 – 400	Aerosols, diffuse
2	420	Aerosols, diffuse
3	500	Aerosols, diffuse
4	610	Ozone
5	675	Aerosols, diffuse
6	880	Aerosols, diffuse
7	940	Water vapor
8	1000 – 1400	Aerosols, clouds
9	1400 – 1700	Aerosols, clouds

5.2 Spectral and broadband clear-sky GHI models

To compute the spectral clear-sky index, both the measured spectral GHI and modelled spectral clear-sky GHI are required – cf. Eq. (13). Since the spectral clear-sky GHI varies across the solar spectrum, it must be computed separately for each wavelength. For this, we begin by defining the spectral clear-sky GHI as:

$$S_{GHI,clr}(\lambda) = S_{DNI,clr}(\lambda) \cdot m^{-1} + S_{DHI,clr}(\lambda) \quad (22)$$

where $S_{\mathrm{DNI,clr}}(\lambda)$ and $S_{\mathrm{DHI,clr}}(\lambda)$ are the modelled spectral clear-sky DNI and diffuse horizontal irradiance (DHI), respectively, in the 280-4000 nm range, and m is the geometric air mass.

The spectral clear-sky DNI is obtained through a parameterized direct beam transmittance model as follows [31]:

1. Compute the zenith angle, air mass, and sun-earth distance using a solar position algorithm;
2. Apply sun-earth distance correction on the AM0 extraterrestrial solar spectrum;
3. Calculate Rayleigh scattering and transmittance functions for mixed and trace gases, e.g., CO_2, CH_4, O_2;
4. Determine aerosol transmittance through Ångström power law by fixing the aerosol optical depth (AOD) at 500 nm to 0.05 and with its spectral dependence defined by two Ångström exponents, 0.98 and 1.22, for wavelengths before and after 500 nm, respectively;

5. Compute the total column ozone and its spectral transmittance from the SolarSIM-G measurements;
6. Calculate the precipitable water amount from the SolarSIM-G measurements, then generate its transmittance profile.

The spectral clear-sky DHI is difficult to obtain using the parameterization techniques used to calculate the spectral clear-sky DNI. Instead, we use the methodology proposed by Bird [34] and later refined by Gueymard [35], where the diffuse horizontal irradiance is given as:

$$S_{\mathrm{DHI,clr}}(\lambda) = S_{\mathrm{R,dif}}(\lambda) + S_{\mathrm{a,dif}}(\lambda) + S_{\mathrm{g,dif}}(\lambda) \quad (23)$$

where $S_{\mathrm{R,dif}}(\lambda)$ is the Rayleigh scattering component, $S_{\mathrm{a,dif}}(\lambda)$ is the aerosol scattering component, and $S_{\mathrm{g,dif}}(\lambda)$ is the sunlight's back and forth scattering between the ground surface and the air. Finally, the broadband clear-sky GHI, required by Eq. (12), can be computed by integrating the spectral clear-sky GHI in the 280-4000 nm range:

$$I_{\mathrm{GHI,clr}} = \int_{280}^{4000} S_{\mathrm{GHI,clr}}(\lambda) \cdot \mathrm{d}\lambda \quad (24)$$

6. Data acquisition and validation

To ensure we develop a comprehensive perspective regarding the full scope of any non-deterministic variations, we need an extended dataset of the *spectral* irradiance. We have therefore created a new spectral irradiance database based on measurements by the aforementioned custom Spectrafy SolarSIM-G spectral pyranometer located in Ottawa, Canada. The climate diversity in Ottawa and the high temporal resolution of the database combine to enable a broadly representable evaluation of irradiance variability across the dominant sky conditions and time scales most relevant to PV generation. The broadband GHI is determined by integrating the spectral irradiance as described above; the direct normal irradiance (DNI) is determined by a complementary instrument to allow the statistical properties of local versus extended sky conditions to be compared.

This work uses one year of 250 ms resolution spectral and broadband GHI data, spanning the entirety of 2022. The database is filtered such that nighttime data and significant outliers are removed. An additional filter is used to remove periods when the solar elevation angle is less than 4° to eliminate any effects introduced by periods surrounding dusk and dawn. These 10- to 20-minute periods near sunrise and sunset typically contribute very little to daily PV generation, thus their obscuring effects on irradiance variability can be neglected.

To further ensure data quality control, periods of precipitation (e.g., rain, mist, snow), maintenance, and intermittent data outages related to power failure are also filtered. Data from the national weather station, located approximately 8 km away from the SolarSIM-G site at the Ottawa/MacDonald-Cartier International Airport, is used to determine the periods of precipitation. This station reports the sky conditions (e.g., sunny, cloudy, light rain, drizzle) at 15-minute to 1-hour resolution. Therefore, broader periods where the sky conditions imply precipitation are used to filter the SolarSIM-G database appropriately.

7. Data analysis and discussion

7.1 General observations regarding variability

We first consider irradiance variability using sky condition persistence and transition metrics to describe general conditions. We begin this analysis by defining our sky conditions. Table 2 below shows how the data is divided into seven sky conditions based on the spectral clear-sky indices of channels 1 and 9 [33]. While the infrared (IR) channel 9 alone can sufficiently classify several of the sky conditions, the ultraviolet (UV) channel 1 is required to properly distinguish between the very clear, clear, and hazy conditions. Since the channel 1 and 9 clear-sky indices are used to define the sky conditions, their variability is embedded in the sky condition persistence and transitions analysis.

Table 2. Sky condition definitions using channels 1 and 9 of the SolarSIM-G

Sky Condition	$\boldsymbol{\kappa(\lambda_{\mathrm{UV}})}$		$\boldsymbol{\kappa(\lambda_{\mathrm{IR}})}$	
	min.	max.	min.	max.
Lensing	-	-	1.05	-
Very Clear	1.0	-	0.75	1.05
Clear	0.8	1.0	0.75	1.05
Hazy	-	0.8	0.75	1.05
Thin Clouds	-	-	0.5	0.75
Thick Clouds	-	-	0.25	0.5
Overcast	-	-	-	0.25

Using these definitions, each timestamp in the dataset is labelled with a corresponding sky condition. Figure 2 shows the distribution of sky conditions throughout the year of 2022 in Ottawa, which highlights the diversity of its weather. It is important to consider that climate varies between locations, so this distribution may be significantly different for another location. However, given its diversity of sky conditions, the Ottawa data enables a general assessment of sky condition persistence and transitions.

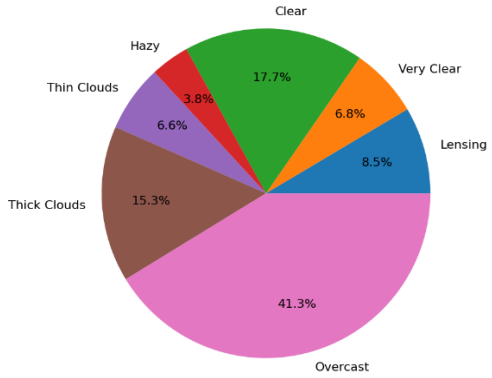

Figure 2. Distribution of sky conditions in Ottawa during 2022

Knowledge of the current sky condition, and for how long it has already persisted, can be highly informative in a forecasting context. For instance, if the sky condition has been overcast for several minutes, one might expect that the sky is likely to remain overcast in the next minute. In contrast, mixed sky conditions which include fast-moving thin clouds may be more volatile, reducing predictability. We aim to provide insight into the following two questions:

1. Given the current sky condition, *x*, what will most likely be the next sky condition, *y*?
2. If the sky transitions from *x* to *y*, for how long will *y* then last?

Given the probabilistic nature of cloud dynamics, the quantification of sky condition behavior should likewise take a probabilistic perspective. Figure 3 illustrates both the likelihood of certain sky conditions occurring and the probability that the latter condition will persist between 1 and 2 minutes. For example, there are 1450 instances within the 1-year database where the sky condition transitions from very clear (vertical axis) to clear (horizontal axis), with the clear condition then lasting at least 1 minute but no more than 2 minutes. The color pattern identifies the dominant transition combinations (darker blue) which are most likely to occur. The transition combinations which have very low counts (light blue or white) are either unlikely to occur or only occur during rapid irradiance ramping periods. These ramping periods are inherently filtered out in this analysis by the requirement that the latter sky condition must persist at least 1 minute following the transition.

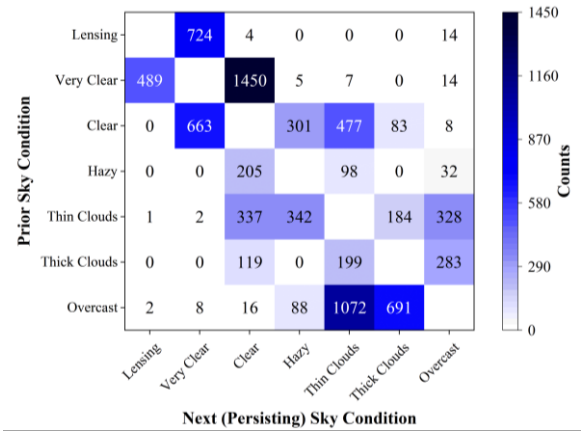

Figure 3. Instances of sky condition persistence lasting between 1 and 2 minutes following a transition. The persisting sky condition is given along the horizontal axis

In Figure 4, the analysis is repeated but for the latter sky condition persisting between 5 and 6 minutes. Interestingly, the dominant sky transition combinations, which are shown by the color patterns, do not change significantly from Figure 3. However, the impact of changing the persistence time scale is evident in the number of counts. The reduced instances of sky conditions persisting for longer periods of time indicates that as we look further into the future, the sky condition is more likely to change and less predictable.

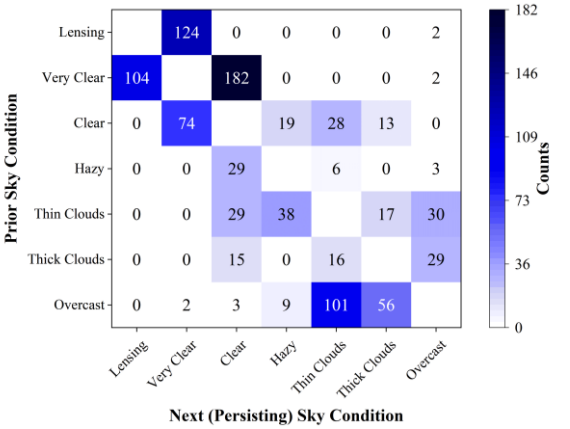

Figure 4. Instances of sky condition persistence lasting between 5 and 6 minutes following a transition. The persisting sky condition is given along the horizontal axis

The overall impact of increasing the persistence duration period is shown in Figure 5. This plot shows the post-transition persistence counts at each time interval for very clear to clear transitions, which is among the dominant transition combinations and can therefore illustrate the temporal distribution well.

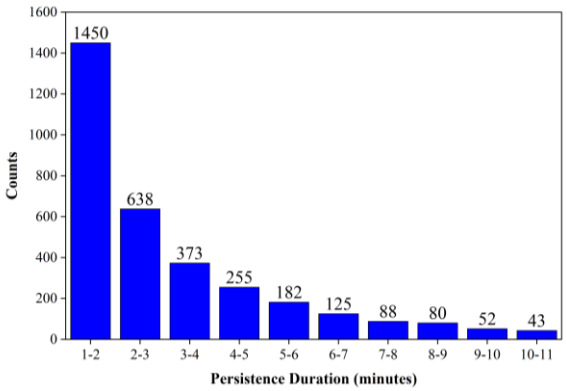

Figure 5. Temporal distribution of persistence duration counts following very clear to clear transitions

Returning to the questions posed earlier in this section, neglecting fast ramping periods in which sky conditions do not persist more than a few seconds, Figure 3 and Figure 4 show the dominant transition combinations which help answer the first question. The second question, pertaining to the persistence duration of the next sky condition, can be guided by the results shown in Figure 5, along with a similar analysis for all other transition combinations.

7.2. Seasonal Clear-Sky Index Statistics

We begin our investigation of the clear-sky index at the seasonal time scale. Regarding the time invariance condition for stationarity, the seasonal ECDFs can be compared visually using a *P-P* plot. Figure 6 shows seasonal *P-P* plots for the broadband GHI clear-sky index, with non-repeating season combinations. Here, each pair of seasonal ECDFs are plotted against one another on either axis. If two seasonal distributions are identical (time invariant), the resulting plot will fall along the diagonal line; however, differences between the ECDFs will result in deviations from the diagonal line. Figure 6 gives validity to seasonal KS test results, which have suggested that all seasonal ECDFs are statistically different. The *P-P* plots also are much more informative in terms of where the ECDFs differ, in contrast to the binary result of the KS test.

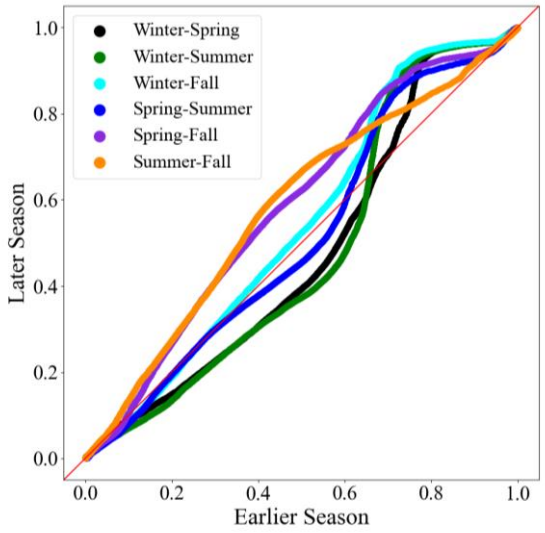

Figure 6. P-P plot for the seasonal ECDFs of the broadband clear-sky index

The seasonal *P-P* plots for the nine spectral channels are presented in Figure 7. While the P-P plots support the KS test results for each channel, they also illustrate that the clear-sky index and its variability are spectrally dependent. The differences between the spectral *P-P* plots show that clouds affect each of the spectral components differently, even at the seasonal time scale. Though some channels have similar *P-P* plots to each other (and to the broadband), others show more drastic differences, which is likely due to different cloud scattering and absorption processes.

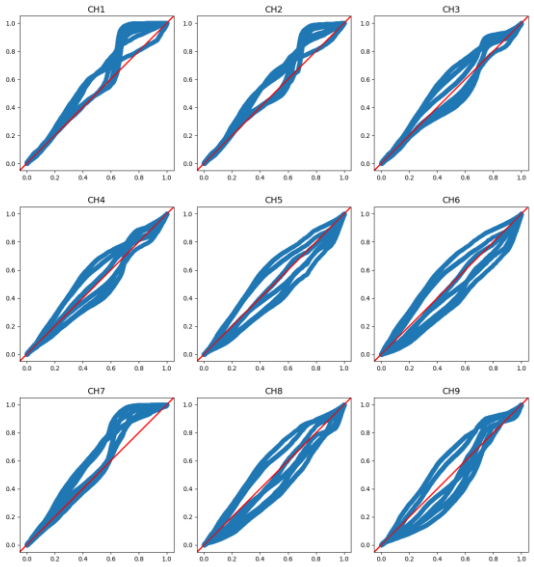

Figure 7. P-P plots for the seasonal CDFs of all nine spectral clear-sky indexes

It is evident from Figure 7 that the broadband and spectral clear-sky index time series do not meet the time invariance condition of strong stationarity on seasonal time scales. However, in the context of solar forecasting, it is typically beneficial if the time series is at least weakly stationary. To meet the requirements of weak stationarity, it is often sufficient for a time series to maintain a constant mean and variance over time. Figure 8 and Figure 9 show the seasonal means and variances, respectively, for all spectral and broadband clear-sky index time series. With the possible exception of channel 7 and perhaps the winter period, fluctuations in the mean and variance might be considered modest, since, from a sample population perspective, the distributions are of significant width. The large variances imply only relatively weak variation in the mean over time, and with one another; hence, the assumption that the clear-sky index is weakly stationary may be a reasonable one at (most) seasonal time scales.

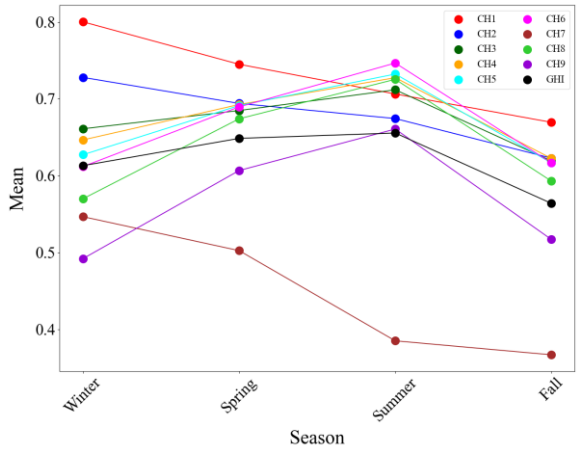

Figure 8. Seasonal clear-sky index means for all spectral and broadband irradiances

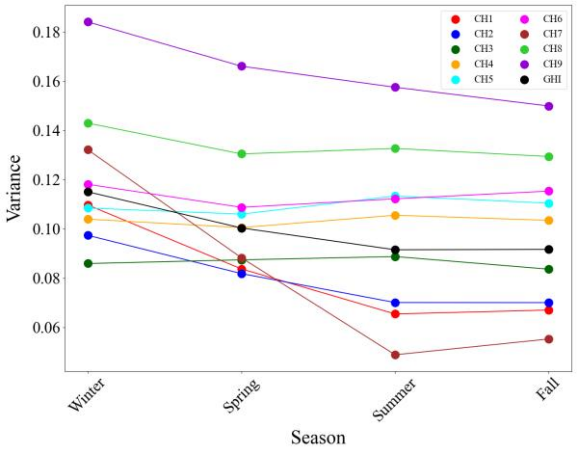

Figure 9. Seasonal clear-sky index variances for all spectral and broadband irradiances

7.3. Monthly Clear-Sky Index Statistics

Having established that the spectral and broadband clear-sky indexes may be considered weakly stationary at best on the seasonal time scale, we now turn our attention towards the monthly time scale. Since separating the high resolution time series into 1-month bins still permits sufficiently large sample sizes, representative monthly ECDFs can be generated. As such, the same metrics and techniques that were used to investigate seasonal stationarity can also be used to address monthly stationarity.

The monthly ECDFs of the broadband clear-sky index are all statistical different from one another, each unique monthly comparison exceeding the KS test summary statistic, $D_{n,m}$. Even at the level of this coarse metric, there are no months with even nominally similar distributions, not even the adjacent months. Figure 10 provides the monthly *P-P* plots for the broadband GHI clear-sky index, with non-repeating month combinations, that confirms this perspective. Due to the large number of curves, these are not individually identified. While further insights into monthly statistical variation can be extracted, it is here sufficient to note that the increasing departure from the diagonal, compared with seasonal dependence, is evidence of greater overall variability on a month-to-month basis.

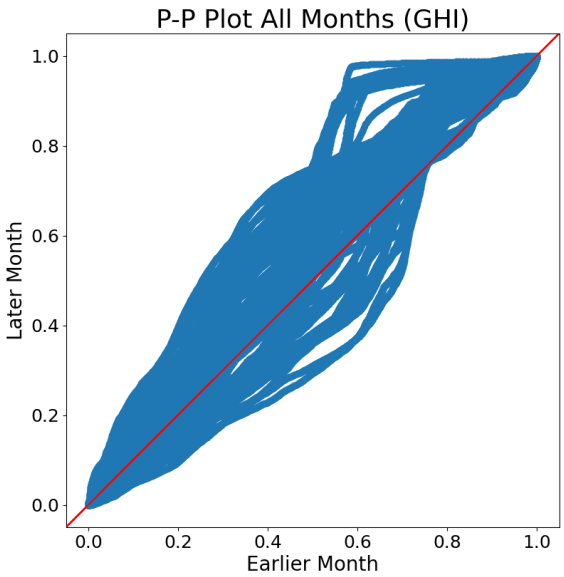

Figure 10. P-P plot for the monthly ECDFs of the broadband clear-sky index

The monthly KS test results for the nine spectral channels provide a more nuanced view of the irradiance variability. Presented in Figure 11 are the KS test results as a digital "heat map", where filled squares denote passing the test and unfilled failing the test. The diagonals compare the month with itself and thus pass by definition. Some specific month pairs for particular channels are similar, with an overall bias towards summer months and short wavelengths, but the broad trend is still a stationarity violation for nearly all month combinations.

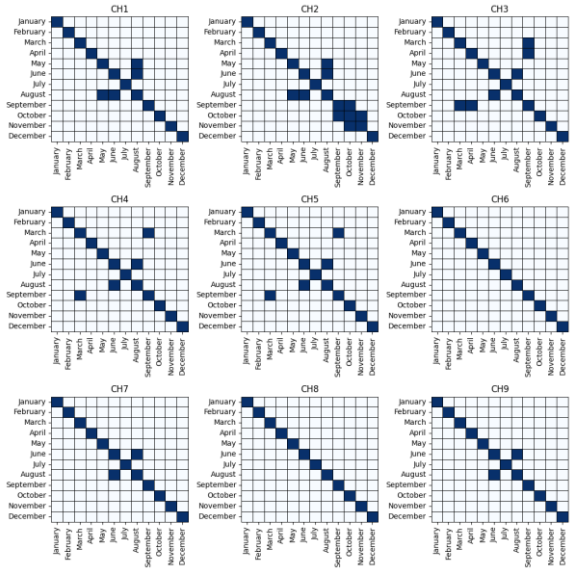

Figure 11. KS test results for the monthly ECDFs of all nine spectral clear-sky indexes

The seasonal *P-P* plots for the nine spectral channels are presented in Figure 12. While the *P-P* plots support the KS test results, they also illustrate that the clear-sky index and its variability are spectrally dependent on monthly time scales, similar to the seasonal spectral *P-P* plots. Again, some channels have *P-P* plots of broadly similar extent, while others show more drastic differences, likely due to different scattering and absorption processes.

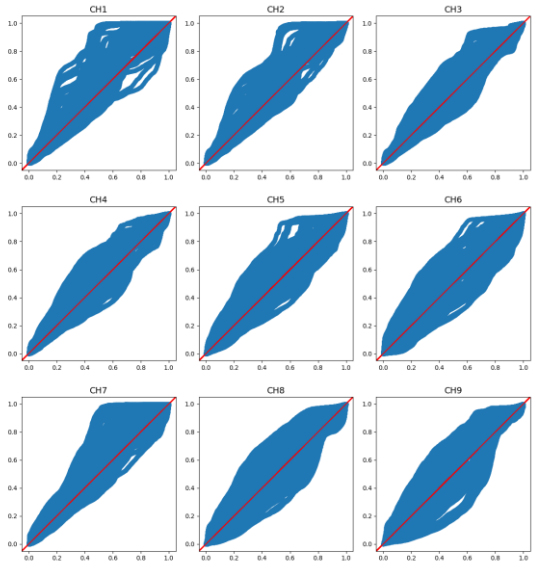

Figure 12. P-P plots for the monthly ECDFs of all nine spectral clear-sky indexes

For completeness, we also consider the monthly dependence of mean and variance, shown in Figure 13 and Figure 14, respectively. Again, the behaviour of channel 7 is anomalous, but the general trend is similar to that of the seasonal dependencies, with one exception. We see here that the variation in the means, excluding late winter, is less than that seen for the seasons, although the dependence of the variance is similar. This is indicative of a move toward stationarity as the averaging time is reduced, and is also consistent with what we see in the KS tests.

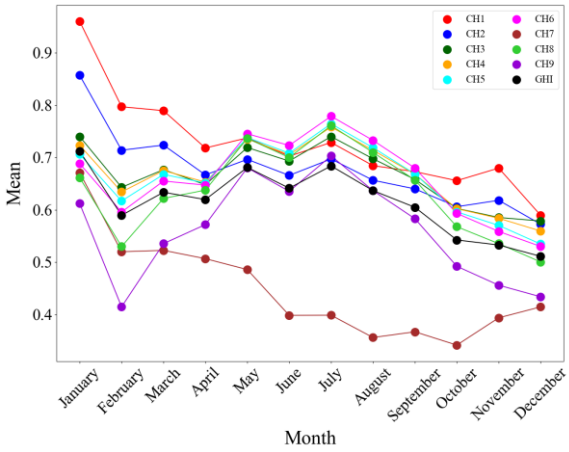

Figure 13. Monthly clear-sky index means for all spectral and broadband irradiances

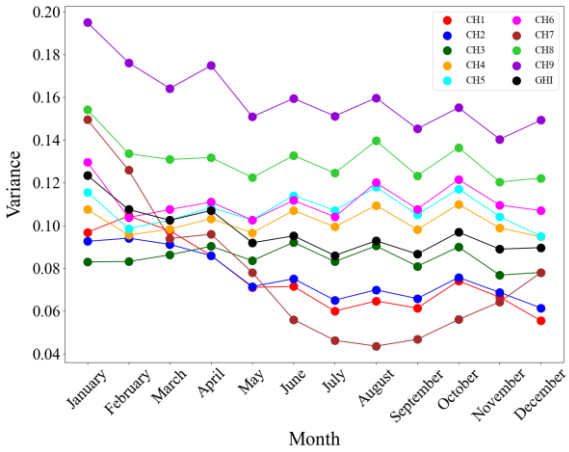

Figure 14. Monthly clear-sky index variances for all spectral and broadband irradiances

In the context of solar forecasting, it would be advisable to prioritize models that are less sensitive to nonstationarity (i.e., advanced machine learning rather than conventional statistics-based models) on monthly or even weekly time scales. Additionally, parsing training data by month or season may help reduce the non-stationarity present in the clear-sky indexes, though this will be at the cost of reduced training data volume and a possible need for month- or season-specific models.

7.4. Sub-Hourly Clear-Sky Index Statistics

As short-term solar forecasting has become an area of great interest in recent years, both for researchers and electricity grid operators, considering sub-hourly statistics is becoming increasingly important. Given the lack of high-resolution irradiance data, assessments on sub-hourly time scales are few. Though there have been many solar resource assessments using satellite data and numerical weather prediction (NWP) models, these are often limited to temporal resolutions greater than 5-15 minutes and spatial resolutions of a few square kilometres. As such, the fast local irradiance fluctuations are lost to spatiotemporal averaging. Other ground-based irradiance variability studies have been presented in the literature, but all high-resolution measurements appear limited to broadband data, neglecting the spectral components.

7.5 Irradiance variability distributions

In this section, we investigate the spectral and broadband irradiance variability on sub-hourly time scales, ranging from sub-second to one hour. To better discern any departure from stationarity – or representations of the type of stationarity – we consider irradiance variability via both first-order and higher-order statistical properties. As to the former, we have already seen the time dependencies of mean and variance for narrowband and broadband clear-sky indices at seasonal and monthly scales. What we lack is an understanding of the behaviour of the autocovariance. Is a comprehensive interpretation possible if time shift invariance is assumed? That is, might a form of weak stationarity exist? The same question might be asked regarding the nature of the index fluctuations themselves. Below we address how the autocovariance and the index increment behave under time shift.

7.5.1 Temporal dependencies of the autocovariance

Employing Eq. (5), we calculate the autocovariance for the nine individual narrowband channels and the broadband GHI for all pairs of daylight times, t_1 and t_2, in the annual dataset where the angle of elevation is in excess of 4° such that $t_2 = t_1 + \tau$, where τ (the lag) ranges from 1 s to 1 hour. The data are shown by the symbols in the linear-log plot of Figure 15. The data were fitted to the survivor function of the log-logistic distribution. The log-logistic distribution is the probability distribution of a random variable, here distributed in time, whose logarithm has a logistic distribution [36]. Similar in shape to the log-normal distribution, it has heavier tails; the PDF is given by:

$$f(\tau) = \frac{(\beta/\alpha)(\tau/\alpha)^{\beta-1}}{[1+(\tau/\alpha)^{\beta}]^2}, \quad (25)$$

for $\tau > 0, \alpha > 0$, and $\beta > 0$. Its CDF can be written in closed form as:

$$F(\tau) = \frac{(\tau/\alpha)^{\beta}}{1+(\tau/\alpha)^{\beta}}. \quad (26)$$

The survivor function:

$$S(\tau) = 1 - F(\tau) = \frac{1}{1+(\tau/\alpha)^{\beta}}, \quad (27)$$

is a measure of the likelihood that the status will persist beyond τ. The fitting parameters α and β are known as the scaling and shape parameters, respectively. For $\beta > 1$, the scaling parameter is the median of the random variable; for $\beta < 1$, the median does not exist and another interpretation must be found.

The survivor function was fit to the autocovariance data by incorporating the clear-sky index variance into its numerator as an independent fitting parameter – the normalization of the autocovariance by the variance produces the autocorrelation. Fits were done to all points except the first and last ones – the former as its measurement was deeply into the highly nonlinear ramping regime, and the latter as systemic notice of slightly worsening fit gave evidence of a departure from the short-range correlation regime. Fits of exceptional accuracy were determined, as shown in Figure 15, using the nonlinear Levenberg Marquardt fitting algorithm in Microcal Origin, with R^2 values better than 0.999. The suggestion that the short-range correlations decay exponentially, as given in Eq. (17), was thoroughly falsified – such a fit underestimates both short and long times, while overestimating the interval in between.

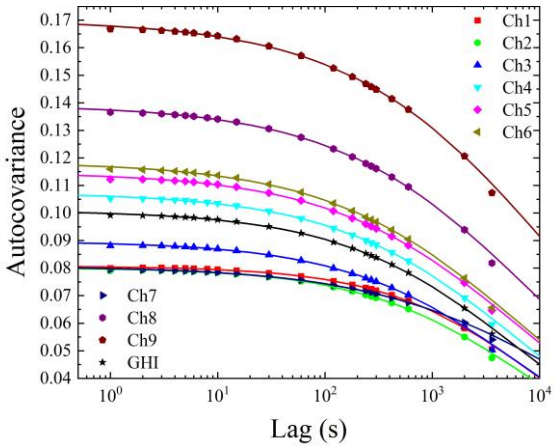

Figure 15. Dependence of the autocovariance on the lag time τ for the 9 narrowband channels (colour) and the broadband GHI (black) – symbols; the solid curves are fits to the data

As noted above, the variance of the clear-sky index distribution was determined as a fitting parameter (and independently confirmed by direct calculation). The results are shown in Figure 16 as a function of the channel nominal wavelength. Black squares are for narrowband channel data while the red bar is that of the broadband GHI. The large "uncertainties" in the wavelengths for channels 8 and 9 are for manufacturer intellectual property reasons. Uncertainties in the variances themselves are determined as standard errors in the fitting procedure and are too small to be seen on this scale. The following observations are worthy of note.

The overall variance is rather low, which presumably reflects the fact that, although Ottawa has highly diverse sky conditions, such conditions may tend to persist for some time, as earlier seen in Figures 3-5. The spectral variation exceeds a factor of two over the channel range and is mostly monotonic, with the highest variances in the infrared. Note that the short wavelength channels 1-6 (except 4) target aerosols and diffuse irradiance while the two IR channels target aerosols and clouds. For short wavelengths, the dependence is likely due to the interplay between two factors: the wavelength dependence of the aerosol extinction as modeled an Ångström power law [37], and the transition probabilities for moving from one sky state to another – cf. Figures 3-5. The lowest variances – for the UV and blue channels 1 and 2, and for the water vapor channel 7 – likely occur because the scattering (channels 1 and 2) and absorption (channels 1 and 7) are largely saturated for their dominate sky states. By contrast, the long wavelength channels show greater variances, possibly due to the

diversity of cloud droplet sizes. Further investigation is needed.

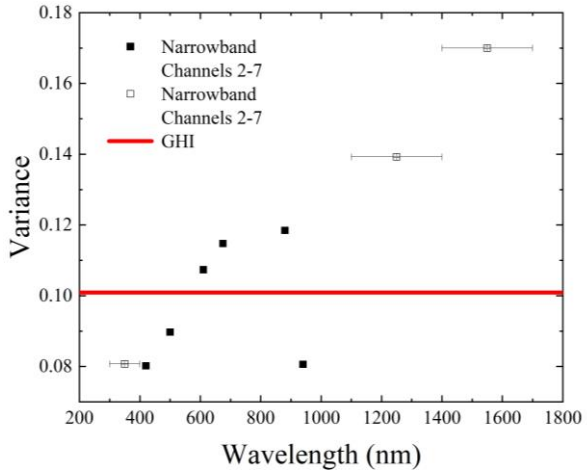

Figure 16. Dependence of the clear-sky index variance on wavelength for the 9 narrowband channels (data points) and the broadband GHI (red bar); error bars on the open symbols denote their respective regions of operation as specified by the manufacturer – see Table 1

Figure 17 shows the fitting results of the scaling parameter α as a function of the channel nominal wavelength. Black squares are for narrowband channel data while the red bar is that of the broadband GHI. Uncertainties in α are determined as standard errors in the fitting procedure and typically too small to be seen on this scale, except for channels 7 and 9. Formally, since $\beta < 1$ (see Figure 18), α cannot be considered as a median since the PDF is monotonically decreasing. For $0 < \beta < 1$, the mathematical significance of this parameter now lies in the scaling behaviour of the log-log dependence of the PDF, which transitions from weakly negative slope to more strongly negative slope, the intersection occurring at α. The physical interpretation of this is that α marks the transition from a short-range correlation regime, where correlations are strong, to a long-range correlation regime, where correlations are increasingly weak; that is, it marks the transition from correlated to a decorrelated behaviour and may thus be considered as a correlation time.

The correlation behaviour shown in Figure 15 is strongly dependent on wavelength and is distinct from that of the variance. Interestingly, the broadband GHI shows a correlation time that is typically less than that of its spectral components, likely as a consequence of its behaviour arising from a multiplicity of physical processes that are to first order independent and identically distributed. Tellingly, its value coincides with that of channel 4, which is basically at the peak of the solar spectrum. Spectral averaging is therefore likely to remove the longer-range correlations, retaining only their common times. There is a significant increase in correlation time with increasing blueness that may reflect the nature of the clear sky states, and likewise an increase into the IR that might demonstrate increasing cloud uniformity. Channel 7 shows a correlation time two to three times that of any other channel. Since this channel was positioned to be sensitive to precipitable water vapor, this indicates that memory of this state is retained far longer than those of other processes – equivalently, that its process evolution is slower. The spectral dependence of the scaling parameter is likely a characteristic dynamic signature of the local climate. Further investigation is needed.

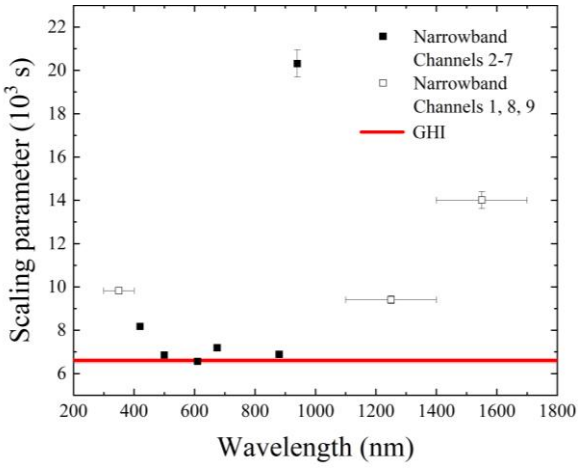

Figure 17. Dependence of the scaling parameter on the wavelengths for the 9 narrowband channels (colour) and the broadband GHI (red bar – width corresponds to uncertainty); error bars on the open symbols denote their respective regions of operation as specified by the manufacturer – see Table 1

Figure 18 shows the fitting results of the shape parameter β as a function of the channel nominal wavelength. Black squares are for narrowband channel data while the red bar is that of the broadband GHI. Uncertainties in β are determined as standard errors in the fitting procedure and are now visible but modest, similar for all channels and the broadband GHI (shown by the faint red lines above and below the red bar). We first note that the shape parameter for the GHI (0.515) agrees with the spectral-intensity-weighted average of the narrowband parameters (0.511) to within the standard error (0.005). This is in line with our earlier discussion regarding averaging over the multiplicity of physical processes, applied now to shape.

It is instructive to speculate on the physical origins of the wavelength dependence of the shape parameter. If we presume that the shape is governed by the persistence of the atmospheric state, then we can assume the dependence of the shape parameter to arise from that of an optical depth which we may parameterize into absorption and scattering components. Assuming the former to be constant and the latter to have Ångström power law behaviour yields as fitting function:

$$\beta(\tau) = a + \left(b/\lambda\right)^{c}, \tag{28}$$

where a, b, and c are fitting parameters. The result of this fit is shown by the black curve in Figure 18, which was determined using the nonlinear Levenberg Marquardt fitting algorithm in Microcal Origin, with a reduced χ^2 value of 1.33 and R^2 value of 0.99. Fitting parameters were found to be $a = 0.4543 \pm 0.0055$, $b = 143 \pm 13$, and $c = 2.15 \pm 0.24$. These values are in line with expectations.

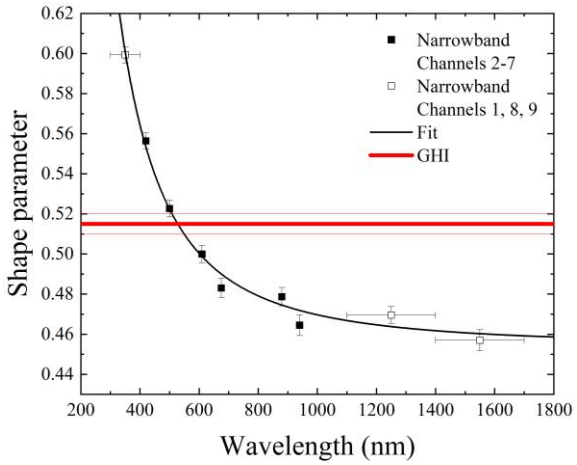

Figure 18. Dependence of the shape parameter on wavelength for the 9 narrowband channels (symbols – data; curve – fit) and the broadband GHI (red bar – range of uncertainty); error bars on the open symbols denote their respective regions of operation as specified by the manufacturer – see Table 1

7.5.2 Index increment distributions

Clear-sky index increment probability distributions have been used to assess irradiance variability at short time scales [38]. This forward difference has also been used in the definition of a structure function to show scaling behaviour – see Eq. (20) – and assess the time series for non-linear and non-stationary properties [29]. We use kernel density estimation as a nonparametric approach to determining these probability distributions. We do this for increment time shifts ranging from 0.75 to 2000 s, for all narrowband channels and for the broadband GHI.

Figure 19 shows clear-sky index increment distributions for the broadband GHI for select time shifts, plotted on a log-linear scale. There are certain observations regarding the general nature of the distributions that are of note. First, the shape, which describes the distribution of the size of the irradiance fluctuations, is very much non-Gaussian. Neglecting lensing effects, the abscissa ranges from -1 to 1, as it is possible to go from a completely opaque sky to full sun (or vice versa) in the time shift period (the range is closer to -1.5 to 1.5 due to lensing effects, but as the roll-off for $|\langle\Delta\kappa_\tau\rangle| > 1$ is rapid this region is not shown). Second, it is sharply peaked at zero, indicating the dominance of the persistent state for $|\langle\Delta\kappa_\tau\rangle| = 0$. The density then drops sharply as the size of the irradiance fluctuation increases.

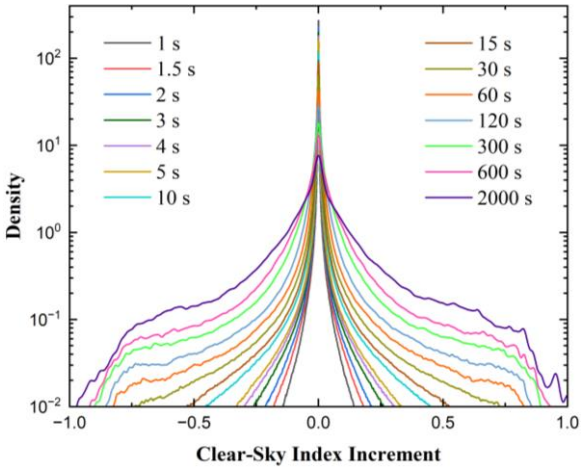

Figure 19. Clear-sky index increment distributions for various time steps using broadband GHI

Regarding the specific features of these distributions as compared to one another, Figure shows that the shortest time shifts are the most sharply peaked, with the highest densities for persistence – the likelihood of seeing extreme fluctuations on short time scales is remote. As the time shift increases, the peak height drops and the wings broaden – the likelihood of seeing extreme fluctuations on long time scales is increasingly likely but still not large. These trends appear monotonic.

Next, we illustrate in Figure 20 how the clear-sky index increment distributions depend on narrowband channels for a representative time shift of 10 s. The broadband GHI and DNI distributions are also shown. The same general observations apply as were

previously described. Specific features are more subtle. On this log scale, variations in peak height are not apparent, nor are any differences over three orders of magnitude in density. Differences become apparent in the wings of the distribution (from which we can then infer peak heights must also change). Channel 1 has the narrowest distribution and channel 9 the widest, but the trend is not monotonic. Broadband GHI and DNI also differ, but the log scale makes any definitive discussion problematic. Another metric must be pursued to interrogate and compare the nature of the fluctuations.

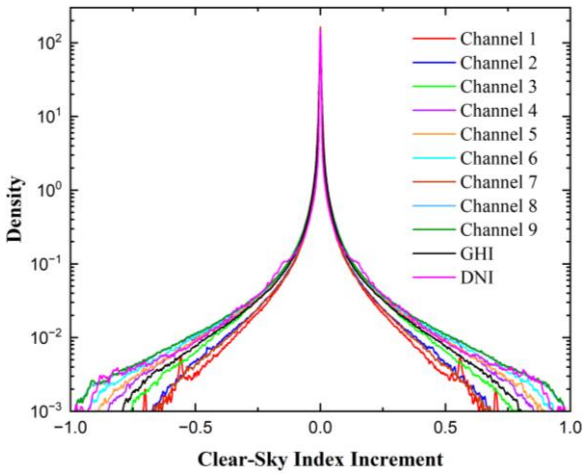

Figure 20. Clear-sky index increment distributions for a fixed time step of 10 seconds

7.5.3 Temporal scaling of peak probability densities

To assess the behaviour of the irradiance fluctuations over a range of time shifts, we need to consider the structure function given in Eq. (20), which is here given by the peak of kernel density estimate. For simplicity, we will presume $q = 1$ and see whether linear power law dependence exists. Figure 21 shows the dependence of the structure function with index increment for the broadband GHI and channels 3-6 (which are well within the visible range). We first note that well-defined power law scaling is indeed observed over at least three orders of magnitude in time shift. Moreover, at least qualitatively, the scaling can be described as linear over most of this range (as delineated by the straight lines). The behaviour of these five datasets is very similar – they have similar scaling behaviour. One interesting feature that all share is a transition in the scaling from a regime that begins ~1 s to one of lesser slope that begins rather abruptly at about 290 s. This is behaviour that is in general agreement with that found by Madanchi *et al.* [29]. Employing a detrended fluctuation analysis (DFA) method, they were the first to note such behaviour for the GHI by using a sophisticated generalized fluctuation function for the quantification of nonlinear scaling. For the very shortest time shifts, there may be some indication of nonlinearity, or at least differences in spectral dependence.

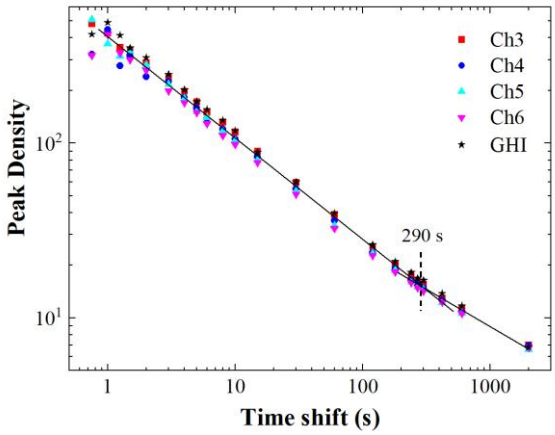

Figure 21. Peak probability densities for broadband GHI and its most similar channels across the investigated time scales

In Figure 22, the peak densities for channels 2 and 8 are shown, with the GHI for reference. Both have scaling transitions and slopes similar to those found in the previous figure, with channel 2 having an upward displacement and channel 8 a downward displacement, indicative of differences in process strength.

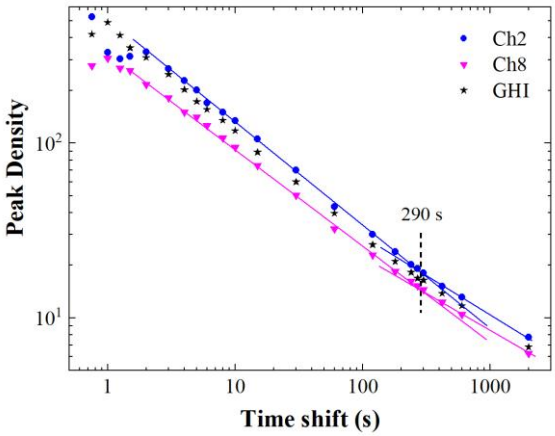

Figure 22. Peak probability densities for the channels with slopes similar to broadband GHI but with vertical offset.

In Figure 23, the spectral channels with scaling behavior that deviate markedly from the broadband are shown, namely, channels 1, 7, and 9; the broadband GHI is shown for reference. Channel 1 (UV) has a slope transition at the same time and scales similarly to the broadband GHI at longer times, but at shorter times its slope is gentler with pronounced early time nonlinearity below ~5 s. Channels 7 and 9 see pure linear scaling without any noticeable transition in the ~290 s region; both have early time nonlinearity below ~5 s. The differences between these three spectral channels and the broadband GHI are likely the result of the impacts of cloud scattering and absorption processes on ultraviolet and infrared wavelengths, whereas these spectral effects are averaged out, to some extent, in the broadband. However, it is for this reason that the ultraviolet and infrared channels are more sensitive to clouds and thus give greater insights into cloud dynamics and the resulting irradiance variability.

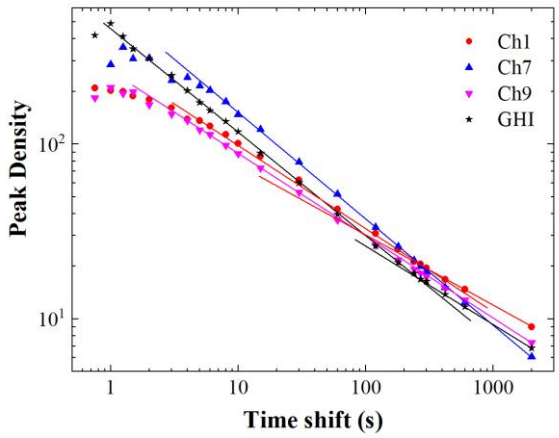

Figure 23. Peak probability densities across the investigated time scales for broadband and spectral irradiance variability distribution

8. Conclusions

In this work, a statistical assessment of atmospheric solar irradiance variability was conducted within the context of data stationarity – or rather nonstationarity – using one year of high temporal resolution spectral irradiance measurement data for Ottawa, Canada. The database, which includes nine narrowband irradiance measurement channels and derived broadband GHI, enables the quantification and intercomparison of multiscale spectral and broadband clear-sky index behavior. The data resolution allows the clear-sky index statistics to be assessed on time scales ranging from one second to monthly and seasonal.

Evidence of irradiance nonstationarity was revealed by first order clear-sky index statistical assessments. Though detrended of deterministic diurnal and orbital patterns, *P-P* plots and KS test results indicate that the monthly and seasonal clear-sky index distributions violate the time-invariance condition of strict stationarity. The requirements of weak stationarity, often simplified to constant mean and variance, are arguably violated as well; however, a multi-year assessment may be required to reduce the biasing effects of dominant short-term weather patterns.

Investigation of the sub-hourly spectral and broadband clear-sky index autocovariances revealed strong evidence of first order spectral and temporal dependencies. The sub-hourly autocovariances were fitted to the survivor function of the log-logistic distribution for the first time, to the best of our knowledge. The fits achieved remarkable accuracy, with R^2 values better than 0.999. The autocovariances revealed spectrally dependent correlation and decorrelation regimes, with the transition times between each being indicated by the fitted scaling parameter, α. Furthermore, the fitted shape parameters for the spectral autocovariances were found to follow Ångström power law behavior when plotted against the respective wavelengths of each spectral channel (under the assumption of constant absorption parameterization). The spectral dependencies of the autocovariances and their fitted log-logistic scaling and shape parameters are necessarily indicative of the underlying physical scattering and absorption processes; however, further investigation is needed to separate the impacts of either mechanism.

Finally, the second order (variability) statistical properties of the clear-sky index were probed via the clear-sky index increment. The distributions of the spectral and broadband clear-sky index increments were also found to have distinct spectral and temporal dependencies. The temporal dependencies are evidenced by distribution broadening and drops in peak probability density by over an order of magnitude across sub-hourly time scales. The peak probability densities were found to exhibit power law scaling, with varying slopes, vertical offsets, and slope transitions (variability regimes), depending on the spectral channel. Similar to the first order statistics, these second order behaviors reveal interesting characteristics and dependencies of the underlying scattering and absorption processes. However, the

slope transitions seen in the peak density scaling are evidently second order by nature, as they do not appear in the autocovariances.

References

[1] S. Rai-Roche, "The world installed 174GW of solar in 2021 and is on track to deploy 260GW by end of 2022 – IEA," PVTech, 25 October 2022. [Online]. Available: https://www.pv-tech.org/the-world-installed-174gw-of-solar-in-2021-and-is-on-track-to-deploy-260gw-by-end-of-2022-iea/. [Accessed 2023].

[2] R. Ahmed, V. Sreeram, Y. Mishra and M. Arif, "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization," *Renewable and Sustainable Energy Reviews,* vol. 124, 2020.

[3] M. Fliess, C. Join and C. Voyant, "Prediction bands for solar energy: New short-term time series forecasting techniques," *Solar Energy,* vol. 166, pp. 519-528, 2018.

[4] S. Ferson and L. R. Ginzburg, "Different methods are needed to propagate ignorance and variability," *Reliability Engineering & System Safety,* vol. 54, no. 2–3, pp. 133-144, 1996.

[5] D. Yang, "A guideline to solar forecasting research practice: Reproducible, operational, probabilistic or physically-based, ensemble, and skill (ROPES)," *Journal of Renewable and Sustainable Energy,* vol. 11, 2019.

[6] D. Salinas, V. Flunkert, J. Gasthaus and T. Januschowski, "DeepAR: Probabilistic forecasting with autoregressive recurrent networks," *International Journal of Forecasting,* vol. 36, no. 3, pp. 1181-1191, 2020.

[7] D. Sarewitz, R. Pielke and R. Byerly, Prediction: Science, Decision Making, and the Future of Nature, Washington, DC: Island Press, 2000.

[8] K. Beven, "Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication," *Hydrological Sciences Journal,* vol. 61, no. 9, pp. 1652-1665, 2016.

[9] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods," *Machine Learning,* vol. 110, p. 457–506, 2021.

[10] T. Palmer and P. Williams, "Introduction. Stochastic physics and climate modelling," *Phil. Trans. R. Soc. A,* vol. 366, pp. 2419-2425, 2008.

[11] J. A. Ruiz-Arias and C. A. Gueymard, "Worldwide inter-comparison of clear-sky solar radiation models: Consensus-based review of direct and global irradiance components simulated at the earth surface," *Solar Energy,* vol. 168, pp. 10-29, 2018.

[12] X. Sun, J. M. Bright, C. A. Gueymard, B. Acord, P. Wang and N. A. Engerer, "Worldwide performance assessment of 75 global clear-sky irradiance models using Principal Component Analysis," *Renewable and Sustainable Energy Reviews,* vol. 111, pp. 550-570, 2019.

[13] Z. Zhang, "Stochastic Representation and Modelling," in *Multivariate Time Series Analysis in Climate and Environmental Research*, Springer, 2018, pp. 149-178.

[14] C. Cheng, A. Sa-Ngasoongsong, O. Beyca, T. Le, H. Yang, Z. (. Kong and S. T. Bukkapatnam, "Time series forecasting for nonlinear and non-stationary processes: a review and comparative study," *IIE Transactions,* vol. 47, no. 10, pp. 1053-1071, 2015.

[15] A. Hasanzadeh, X. Liu, N. Duffield and K. R. Narayanan, "Piecewise Stationary Modeling of Random Processes Over Graphs With an Application to Traffic Prediction," *2019 IEEE International Conference on Big Data (Big Data)*, pp. 3779-3788, 2019.

[16] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and practices, OTexts, 2018.

[17] J. Ekström, M. Koivisto, I. Mellin, R. J. Millar and M. Lehtonen, "A Statistical Model for Hourly Large-Scale Wind and Photovoltaic Generation in New Locations," *IEEE Transactions on Sustainable Energy,* vol. 8, no. 4, pp. 1383-1393, 2017.

[18] Z. Dong, D. Yang, T. Reindl and W. M. Walsh, "Short-term solar irradiance forecasting using exponential smoothing state space model," *Energy,* vol. 55, pp. 1104-1113, 2013.

[19] D. Yang, "Choice of clear-sky model in solar forecasting," *J. Renewable Sustainable Energy,* vol. 12, no. 2, 2020.

[20] B. Y. Liu and R. C. Jordan, "The interrelationship and characteristic distribution of direct, diffuse and total solar radiation," *Solar Energy,* vol. 4, no. 3, pp. 1-19, 1960.

[21] K. Hollands and R. Huget, "A probability density function for the clearness index, with applications," *Solar Energy,* vol. 30, no. 3, pp. 195-209, 1983.

[22] J. Gordon and T. Reddy, "Time series analysis of daily horizontal solar radiation," *Solar Energy,* vol. 41, no. 3, pp. 215-226, 1988.

[23] J. Gordon and T. Reddy, "Time series analysis of hourly global horizontal solar radiation," *Solar Energy,* vol. 41, no. 5, pp. 423-429, 1988.

[24] R. Aguiar and M. Collares-Pereira, "Statistical properties of hourly global radiation," *Solar Energy,* vol. 48, no. 3, pp. 157-167, 1992.

[25] R. Aguiar and M. Collares-Pereira, "TAG: A time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation," *Solar Energy,* vol. 49, no. 3, pp. 167-174, 1992.

[26] A. Skartveit and J. Olseth, "The probability density and autocorrelation of short-term global and beam irradiance," *Solar Energy,* vol. 49, no. 6, pp. 477-487, 1992.

[27] A. Woyte, R. Belmans and J. Nijs, "Fluctuations in instantaneous clearness index: Analysis and statistics," *Solar Energy,* vol. 81, no. 2, pp. 195-206, 2007.

[28] J. Munkhammar and J. Widén, "An autocorrelation-based copula model for generating realistic clear-sky index time-series," *Solar Energy,* vol. 158, pp. 9-19, 2017.

[29] A. Madanchi, M. Absalan, G. Lohmann, M. Anvari and M. R. R. Tabar, "Strong short-term non-linearity of solar irradiance fluctuations," *Solar Energy,* vol. 144, pp. 1-9, 2017.

[30] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde and H. Stanley, "Multifractal detrended fluctuation analysis of nonstationary time series," *Physica A: Statistical Mechanics and its Applications,* vol. 316, no. 1–4, pp. 87-114, 2002.

[31] V. Tatsiankou, K. Hinzer, J. Haysom, H. Schriemer, K. Emery and R. Beal, "Design principles and field performance of a solar spectral irradiance meter," *Solar Energy,* vol. 133, pp. 94-102, 2016.

[32] V. Tatsiankou, K. Hinzer, H. Schriemer, S. Kazadzis, N. Kouremeti, J. Gröbner and R. Beal, "Extensive validation of solar spectral irradiance meters at the World Radiation Center," *Solar Energy,* vol. 166, pp. 80-89, 2018.

[33] V. Tatsiankou, K. Hinzer, H. Schriemer and R. Beal, "Improved Global Irradiance Decomposition by Sky Condition Classification from Measured Spectral Clearness Indices," *47th IEEE Photovoltaic Specialists Conference (PVSC),* pp. 0072-0076, 2020.

[34] R. E. Bird, "A simple, solar spectral model for direct-normal and diffuse horizontal irradiance," *Solar Energy,* vol. 32, no. 4, pp. 461-471, 1984.

[35] C. Gueymard, SMARTS2, a simple model of the atmospheric radiative transfer of sunshine: algorithms and performance assessment, Florida Solar Energy Center, 1995.

[36] N. Balakrishnan, Handbook of the Logistic Distribution, CRC Press, 2013.

[37] J. Martínez-Lozano, M. Utrillas, F. Tena and V. Cachorro, "The parameterisation of the atmospheric aerosol optical depth using the Ångström power law," *Solar Energy*, vol. 63, no. 5, pp. 303-311, 1998.

[38] G. Lohmann, "Irradiance Variability Quantification and Small-Scale Averaging in Space and Time: A Short Review.," *Atmosphere* , vol. 9, no. 7, 2018.

[39] B. Y. Liu and R. C. Jordan, "The interrelationship and characteristic distribution of direct, diffuse and total solar radiation," *Solar Energy*, vol. 4, no. 3, pp. 1-19, 1960.

Chapter 5: Solar Irradiance Nowcasting Models

In this chapter, several intra-hour solar irradiance forecasting (nowcasting) models are implemented and compared in terms of their training times, prediction times, and forecasting skill. The LSTM, XGBoost, and 1D-CNN algorithms outlined in Chapter 3 are used as the base models throughout this chapter. These models are implemented in Python, with Google's Keras library, which runs on top of TensorFlow, being used for LSTM and 1D-CNN and an open source library from Distributed (Deep) Machine Learning Common (DMLC) being used for XGBoost. The smart persistence model described in Chapter 2 is used as the reference model for calculating the skill scores. The nowcasting models are trained using 12 months of data (June 2021 to May 2022) and tested on 9 months of data not used in the training process (June 2022 to February 2023). This training and testing data is taken from the same Spectrafy SolarSIM-G spectral irradiance database that was used for the statistical analysis presented in Chapter 4. It should be noted that in this chapter, the term "irradiance" refers to the clear-sky index, which is detrended of the known deterministic diurnal and orbital patterns. After predictions are made using the clear-sky index, the predicted values can simply be mapped back to GHI using the same clear-sky model used for the detrending; however, this is beyond the scope of this thesis.

Three sets of nowcasting models are developed in this chapter to predict the future broadband irradiance at different sub-hourly forecast horizons. The first set of models, referred to as the *All-Sky (Broadband Only)* models, are trained using data from all sky conditions and take broadband irradiance as the only input. These first models reflect the simplest approach to solar irradiance forecasting using ground-based measurement data, and are thus used as benchmarks. The second set of models, referred to as the *All-Sky (Broadband & Spectral)* models, are also trained using data from all sky conditions but take both the broadband and spectral irradiance data as inputs. Since the Spectrafy SolarSIM-G database contains one broadband and nine spectral irradiance time series, these models take a total of ten inputs. The third set of models, referred to as the *Ramp Regime* sub-models, likewise take the broadband and spectral irradiance data as inputs. However, these models are not trained on all sky conditions; instead, the database is divided into four ramping (variability) regimes and a different specialized sub-model is trained on each. The hypothesis inspiring this approach is that each sub-model will outperform an otherwise equivalent all-sky model within its specialty ramp regime. This hypothesis is tested by comparing the ramp regime sub-models with the All-Sky (Broadband & Spectral) models within their respective ramp regimes. In practice, the ramp regime sub-models would need to be combined in a diverse and dynamic ensemble for making predictions in all sky conditions, though this remains beyond the scope of this thesis. The approach to identifying the

different ramp regimes, which leverages insights from the conference proceeding provided in the appendix of this thesis, is also described in this chapter.

5.1 All-Sky (Broadband Only) Models

This section presents the hyperparameter details, training times, and performances of the All-Sky (Broadband Only) LSTM, XGBoost, and 1D-CNN nowcasting models. These models take the broadband irradiance as the sole input and make predictions of the future broadband irradiance at various forecast horizons. The selected forecast horizons span the entire sub-hourly time scale, ranging from five seconds to one hour ahead. It was experimentally determined that the use of sub-second resolution data does not improve forecasting performance compared with 1 s resolution data, but it does significantly increase computational resource demand and training time. As such, the training and testing data was resampled from the original 250 ms resolution to 1 s resolution.

The All-Sky (Broadband Only) LSTM model was trained using one year of broadband irradiance training data. Using ten seconds of past observations (i.e., the lookback window) was found to be appropriate for balancing model performance and training time, as increasing the lookback window increases data pre-processing time. Since irradiance time series were found to exhibit nonstationary behaviour in Chapter 4, it is was determined that the LSTM model should contain multiple hidden layers to better handle the non-stationarity [56], [57], [58]. Therefore, the LSTM used in this section was designed as a deep learning model with 64 LSTM cells in the first hidden layer, 16 LSTM cells in the second hidden layer, and 8 standard neurons in the third hidden layer. The third hidden layer containing the standard neurons is used to help aggregate the output of the second hidden layer before passing it to the model's output layer. The design of the hidden layers was tailored to balance model complexity and computational resource demands with forecasting performance across all of the sub-hourly forecast horizons.

The All-Sky (Broadband Only) XGBoost model was trained using the same one year of broadband irradiance training data as the All-Sky (Broadband Only) LSTM model. Since the XGBoost algorithm does not use historical observations to make predictions, less data pre-processing was needed compared to the LSTM model. To limit the XGBoost model's complexity and computational resource demand, a maximum tree depth of 6 was used. The number of estimators (XGBoost trees) in the model was set to 200, which was found to balance underfitting and overfitting based on training loss curves.

The All-Sky (Broadband Only) 1D-CNN model was trained using the same one year of broadband irradiance training data as the All-Sky (Broadband Only) LSTM and XGBoost models. In order to facilitate a fair comparison between the 1D-CNN and LSTM, which are both neural networks, the same architecture that was used for the LSTM was used for the 1D-CNN. That is, the 1D-CNN was also designed as a deep learning model with 64 convolutional neurons in the first hidden layer, 16 convolutional neurons in the second hidden layer, and 8 standard neurons in the third hidden layer. Additionally, the 1D-CNN uses the same 10 s lookback window as the LSTM. By making the models similar, neither has an architectural advantage over the other; training time and performance differences between the two models are therefore due to the algorithms themselves.

5.1.1 All-Sky (Broadband Only) Model Training and Testing Times

Solar nowcasting model training and testing times, which are used as an indicator of computational resource demand in this thesis, must be low in order to be economical and practical in an operational context. Since model training can largely be done before deployment, training time is not a direct concern for real-time operational constraints. However, when using cloud computing for training, updating, and deploying nowcasting models [59], computational resource demand is correlated with cloud computing costs. Thus, as an indicator of resource demand, training time must be kept low for a nowcasting model to be cost-effective in practice. Regarding real-time forecasting, model testing (i.e., real-time prediction) must be fast in order to meet the operational forecasting time requirements. For instance, when making 5-minute-ahead forecasts every 30 seconds, a model which takes more than 30 seconds to generate a new prediction will not be suitable. This constraint is further tightened as the forecast horizon gets shorter, such as making 10-second-ahead forecasts every second. Additionally, when forecasting models are deployed using edge computing [60] – which relies on devices like smart meters, sensors, or inverters to handle the real-time forecasting – computational resources are very limited, meaning forecasting models must be lightweight.

To help assess the economic practicality of the All-Sky (Broadband Only) LSTM, XGBoost, and 1D-CNN models, their average training times are presented in Table 1. The values represent the average training times of the 14 models trained for each algorithm which correspond to the 14 sub-hourly forecasting horizons used in this chapter. From Table 1, it is clear that the XGBoost models are by far the fastest to train, followed by 1D-CNN. On the contrary, the LSTM models, which are popular in the literature, take a substantially longer time to train. In fact, when considering how much it would cost to train these LSTMs using cloud computing

through a third party provider, the LSTM training times may even be prohibitive.

Table 1: Average training times for the All-Sky (Broadband Only) nowcasting models

Model Category	**Model Training Time (s)**		
	LSTM	**XGBoost**	**1D-CNN**
All-Sky (Broadband Only)	5950	145	1150

There are ways in which the training times in Table 1 could be reduced. First, these models were trained using the central processing unit (CPU) of a computer with 32 GB of random-access memory (RAM). Using a computer with more RAM (e.g., 64 GB or 128 GB) and performing the computations with a graphics processing unit (GPU) instead of only using a CPU would help reduce the training time, for instance. However, these changes would not likely reduce cloud computing costs, since these are normally determined by the total resource (e.g., memory) demand. That is, whether it takes 30 minutes or 15 minutes to train a model, the total computational resource demand of each algorithm is the same. This illustrates that training time only works as an indicator of the *relative* computational requirements of each algorithm for a given computer. As such, the algorithm ranking from Table 1 likely remains a reasonable measure of each algorithm's economic practicality relative to one another; however, further investigation and testing with cloud computing services is needed to confirm this.

Next, to help assess the suitability of each model for real-time operation, Table 2 contains the average testing times of the All-Sky (Broadband Only) LSTM, XGBoost, and 1D-CNN models. These testing times represent the average time it takes each model to make forecasts for the entire 9 months of testing data. Specifically, the times in this table show how long it takes each model to make approximately 9.8 million predictions. The testing times reflect the same ranking of each algorithm as the training times from Table 1, with LSTM being the slowest, XGBoost being by far the fastest, and 1D-CNN falling in between.

Table 2: Average testing times for the All-Sky (Broadband Only) nowcasting models

Model Category	**Model Testing Time (s)**		
	LSTM	**XGBoost**	**1D-CNN**
All-Sky (Broadband Only)	540	4	190

Despite the comparatively long time taken by the LSTM to make all 9.8 million predictions, this testing time is by no means prohibitive. The LSTM takes less than 10 minutes to make these

predictions, which is sufficiently fast for real-time operation. In fact, it takes around 430 ms for the LSTM to make a single prediction, which is sufficient for most forecast horizons. However, even though the LSTM's prediction time is sufficient for meeting operational requirements, the 1D-CNN and XGBoost models are still much faster. For making a single prediction, 1D-CNN takes around 90 ms and XGBoost takes less than 10 ms. Although all three algorithms meet the operational time constraints of making real-time forecasts, their prediction times – like the training times – indicate their real-time computational resource demand. If the models are implemented with edge computing, the computational resource demand is a critical variable which must be taken into consideration. In this context, the XGBoost and 1D-CNN models are preferred over the most computation-intensive LSTM.

5.1.2 All-Sky (Broadband Only) Model Forecasting Performances

The All-Sky (Broadband Only) models are used in this thesis to provide forecasting performance benchmarks. As described in Chapter 2, deterministic forecasting performance should be evaluated using the RMSE-based skill score in accordance with the ROPES guideline. Therefore, to begin the forecasting performance evaluation, the RMSEs of the All-Sky (Broadband Only) LSTM, XGBoost, and 1D-CNN models across sub-hourly time scales are shown in Figure 5.1. These RMSE values were calculated using predictions made by each model on the nine months of testing data. The smart persistence model RMSEs for the same testing data are provided for reference.

The RMSEs in Figure 5.1 illustrate the strength of the smart persistence model at very short forecast horizons. For forecast horizons less than 10 seconds, the smart persistence model achieves very low RMSEs down to around 0.05. The smart persistence model is therefore a difficult baseline to beat at these horizons since there is little room for improvement; for a model to be considered skillful on very short time scales, its prediction error must approach zero. While the machine learning models are also able to achieve very low RMSEs on these short time scales, they overlap with the smart persistence model RMSEs which indicates they are unable to significantly improve upon its results. However, the smart persistence model shows its weakness at longer forecast horizons, where it separates from the machine learning models, which achieve much lower RMSEs. Figure 5.1 also shows that all three machine learning models perform similarly to each other in terms of their RMSEs; however, to properly evaluate each model the skill score must be used.

To better visualize their forecasting performances relative to the reference (smart persistence)

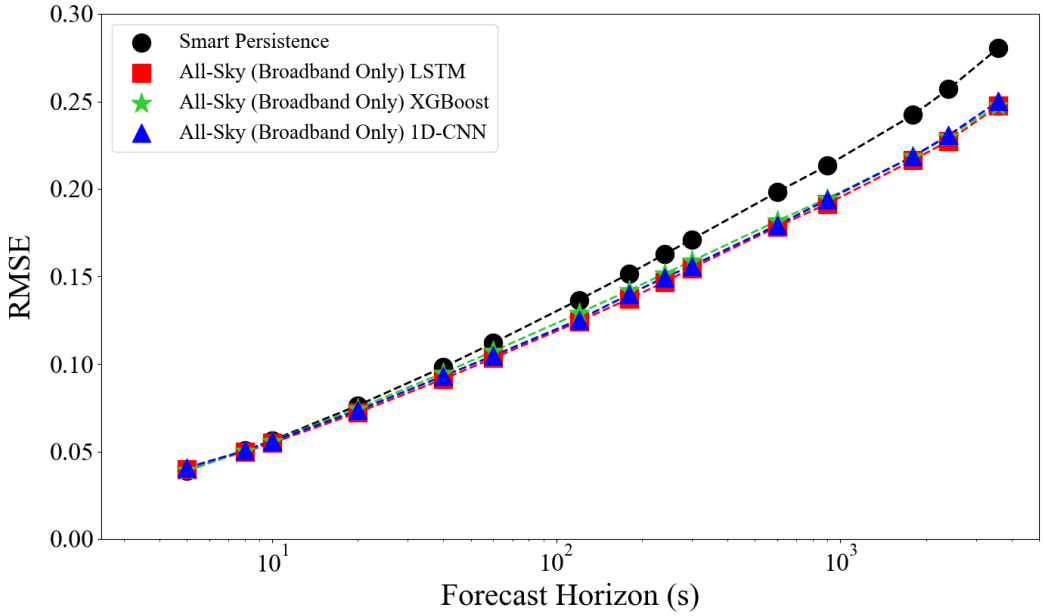

Figure 5.1: RMSEs of the All-Sky (Broadband Only) models using 9 months of testing data; persistence model provided for reference

model, Figure 5.2 shows the skills scores of the All-Sky (Broadband Only) models. Negative skill scores, which are not shown in these plots, indicate that a given model cannot outperform the no-skill smart persistence model at a particular forecast horizon. In contrast, positive skill scores mean that a model outperforms the smart persistence model, with a higher skill score indicating a greater relative improvement. It is evident that the LSTM model has the highest forecasting skill across all sub-hourly time scales, achieving positive skill scores down to 8 seconds ahead. The 1D-CNN also performs well and achieves positive skill scores beyond 8 seconds ahead, though it does not outperform the LSTM at any forecast horizon. The XGBoost performs slightly worse than the other two models for most forecast horizons, although it does close the performance gap at the shortest (i.e., 8-10 seconds) and longest (i.e., over 900 seconds) time scales. Interestingly, the models rank in the opposite order in terms of forecasting skill than they do training and testing time. Therefore, a tradeoff exists between computational (and financial) costs and forecasting accuracy, which must be considered in practice. Of the three models being compared, the 1D-CNN seems to offer the best balance between performance and resource demand; however, exhaustive fine tuning of hyperparameters may help offset the weaknesses of each model. It should be kept in mind that every machine learning model has strengths and weaknesses, and model selection depends upon the application requirements.

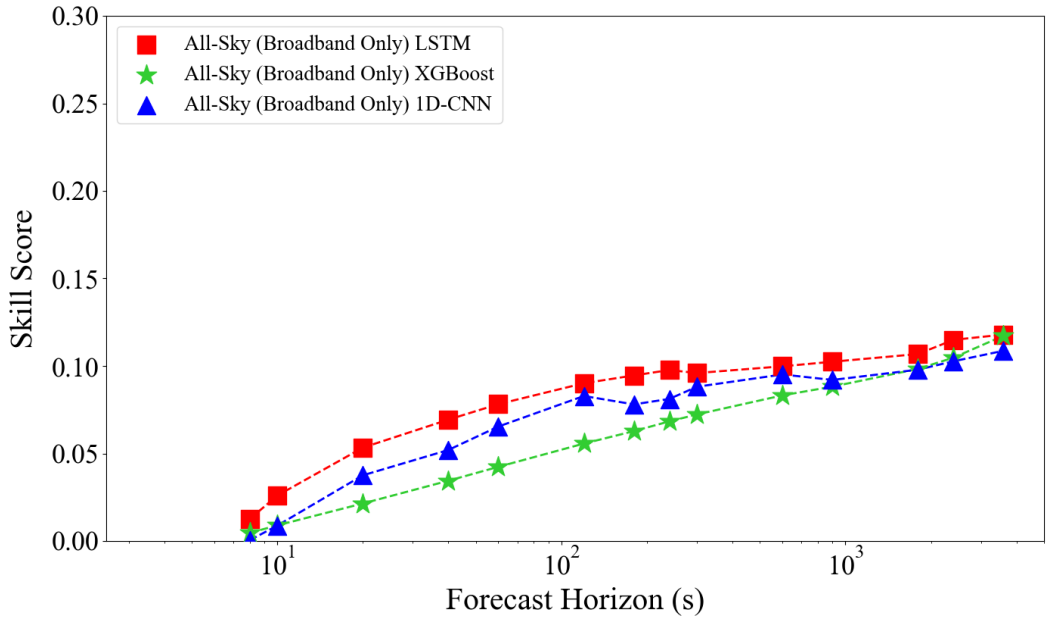

Figure 5.2: Skill scores of the All-Sky (Broadband Only) models using 9 months of testing data

All three models exhibit an upward trend in skill score as the forecast horizon increases. However, it is important to note that this upward trend is not due to the forecasts getting more accurate, but rather the smart persistence model getting progressively worse (see Figure 5.1) and more easily outperformed. Conversely, the low skill scores at shorter forecast horizons are not due to the models performing poorly, but rather the smart persistence model performing very well. Though small in magnitude, the positive skill scores achieved down to 10 second (and below) forecast horizons using data from a single ground-based irradiance sensor are a significant result. These time scales are often neglected, partly due to lack of high temporal resolution data, but Figure 5.2 shows that there is potential for machine learning models to achieve better forecasting results than smart persistence down to a few seconds ahead.

5.2 All-Sky (Broadband & Spectral) Models

This section presents the hyperparameter details, training times, and forecasting performances of the All-Sky (Broadband & Spectral) LSTM, XGBoost, and 1D-CNN nowcasting models. Like the All-Sky (Broadband Only) models, these models take the broadband irradiance as an input and make predictions of the future broadband irradiance at various forecast horizons.

However, as the name suggests, the All-Sky (Broadband & Spectral) models also take the nine spectral irradiance time series provided by the SolarSIM-G database as inputs, totalling 10 inputs. The same sub-hourly forecast horizons, training data, and testing data from Section 5.1 are used in this section.

The All-Sky (Broadband & Spectral) LSTM models were trained using data from the same one year period as was used in Section 5.1. The LSTM lookback windows for all forecast horizons were again set to 10 seconds. To ensure that only the effect of adding spectral irradiance input data is assessed, the LSTMs used in this section were designed with the same hyperparameters as those in Section 5.1.

The All-Sky (Broadband & Spectral) XGBoost models were also trained using data from the same one year period as was used in Section 5.1. Again, to isolate the effects of adding spectral irradiance input data, the same XGBoost hyperparameters from Section 5.1 were used for these models. Specifically, the All-Sky (Broadband & Spectral) models were allowed a maximum tree depth of 6 and contained 200 estimators (XGBoost trees).

The All-Sky (Broadband & Spectral) 1D-CNN models were similarly trained using data from the same one year period as was used in Section 5.1. A 10 second lookback window was again used in order to compare the 1D-CNN and LSTM models fairly. Finally, the same 1D-CNN model hyperparameters used in Section 5.1 were again used in for this set of models to isolate the effect of adding spectral irradiance input data.

5.2.1 All-Sky (Broadband & Spectral) Model Training and Testing Times

The All-Sky (Broadband & Spectral) models take an additional 9 input variables compared to the All-Sky (Broadband Only) models, which only take one. Theoretically, providing this additional information to the machine learning models has the potential of improving their forecasting performance. This is especially true since irradiance variability was found to be spectrally dependent in Chapter 4, meaning information relevant for forecasting may be embedded in the spectral measurements. However, before testing the hypothesis that the inclusion of spectral irradiance input data will improve the models' performances, it is important to first verify whether adding nine inputs jeopardizes the models' practicality. Specifically, it must be verified that the All-Sky (Broadband & Spectral) models are both economically feasible and able to make forecasts within a realistic time frame. Similar to the All-Sky (Broadband Only) models, this can again be done by assessing the training and testing times of the All-Sky (Broadband & Spectral) models.

The training times for the All-Sky (Broadband & Spectral) models are shown in Table 3. The training times for the All-Sky (Broadband Only) models from Section 5.1.1 are also included for comparative purposes. The training times are again used as indicators of each model's computational complexity, which is coupled with economic costs when using cloud computing to train the models. It is immediately evident from Table 3 that the All-Sky (Broadband & Spectral) LSTM, XGBoost, and 1D-CNN models rank the same way as their All-Sky (Broadband Only) counterparts in terms of training times. That is, XGBoost is remains the fastest model to train, LSTM takes the longest, and 1D-CNN lies in between. The more significant result from Table 3, however, is that the All-Sky (Broadband & Spectral) models take only slightly longer to train than the All-Sky (Broadband Only) models. This means that the addition of the 9 spectral irradiance inputs does not significantly increase the computational complexity or operating costs for any of the models, at least in terms of model training.

Table 3: Average training times for the All-Sky (Broadband Only) and All-Sky (Broadband & Spectral) nowcasting models

Model Category	**Model Training Time (s)**		
	LSTM	**XGBoost**	**1D-CNN**
All-Sky (Broadband Only)	5950	145	1150
All-Sky (Broadband & Spectral)	6030	180	1210

Next, to assess whether the All-Sky (Broadband & Spectral) models are suitable for real-time operation, Table 4 presents their average testing times. The testing times for the All-Sky (Broadband Only) models are also included in this table for comparison. As in Section 5.1.1, these testing times represent the average time it takes each of model to make forecasts for the entire 9 months of testing data (approximately 9.8 million predictions).

Table 4: Average testing times for the All-Sky (Broadband Only) and All-Sky (Broadband & Spectral) nowcasting models

Model Category	**Model Testing Time (s)**		
	LSTM	**XGBoost**	**1D-CNN**
All-Sky (Broadband Only)	540	4	190
All-Sky (Broadband & Spectral)	555	4	195

As expected, the testing times for the All-Sky (Broadband & Spectral) models reflect the same algorithm ranking as the training times from Table 3, with LSTM being the slowest, XGBoost

being the fastest, and 1D-CNN falling in between. While this ranking is an important factor when deciding which algorithm to use in practice, the more significant comparison is between the All-Sky (Broadband & Spectral) models and their All-Sky (Broadband Only) counterparts. Similar to the training times, the testing times do not increase by more than a few seconds when the 9 spectral irradiance inputs are added. This means that the type of algorithm is the limiting factor when it comes to meeting operational constraints, whereas the inclusion of spectral irradiance input data has a relatively small impact.

5.2.2 All-Sky (Broadband & Spectral) Model Forecasting Performances

Having established that adding spectral irradiance input data to the nowcasting models does not inhibit their operational and economically viability, this section evaluates the impact of using spectral irradiance input data in terms of forecasting performance. For this purpose, the All-Sky (Broadband & Spectral) model performances are compared with the benchmark All-Sky (Broadband Only) model performances. In Figure 5.3, the RMSEs of the All-Sky (Broadband & Spectral), All-Sky (Broadband Only), and smart persistence models are shown. While all of the models achieve similar RMSEs at the shortest time scales, the improvements are more evident at the longer forecast horizons. Comparing the All-Sky (Broadband & Spectral) and All-Sky (Broadband Only) models, it can be seen that using spectral irradiance input data improves forecasting accuracy by a relatively significant margin. In fact, the increasing separation between the All-Sky (Broadband & Spectral) and All-Sky (Broadband Only) RMSE curves indicates that the relative improvements obtained by the former will continue to increase as the forecast horizon extends beyond one hour.

To properly compare the All-Sky (Broadband & Spectral) and All-Sky (Broadband Only) models, Figure 5.4 shows their skill scores. This plot more clearly shows the performance improvements obtained from adding spectral irradiance inputs, with the LSTM, 1D-CNN, and XGBoost all benefiting from the extra data. The increasing separation between the All-Sky (Broadband & Spectral) and All-Sky (Broadband Only) skill scores further illustrates how the spectral irradiance data results in greater performance gains at longer horizons.

Comparing the individual machine learning algorithms, the skill score results in Figure 5.4 are mixed. Specifically for the All-Sky (Broadband & Spectral) models, LSTM achieves the highest skill at most of the shorter forecast horizons, whereas XGBoost dominates at the longer horizons. The 1D-CNN outperforms XGBoost at the very short time scales, though on average it has the lowest skill scores across the investigated forecast horizons. These mixed results, as

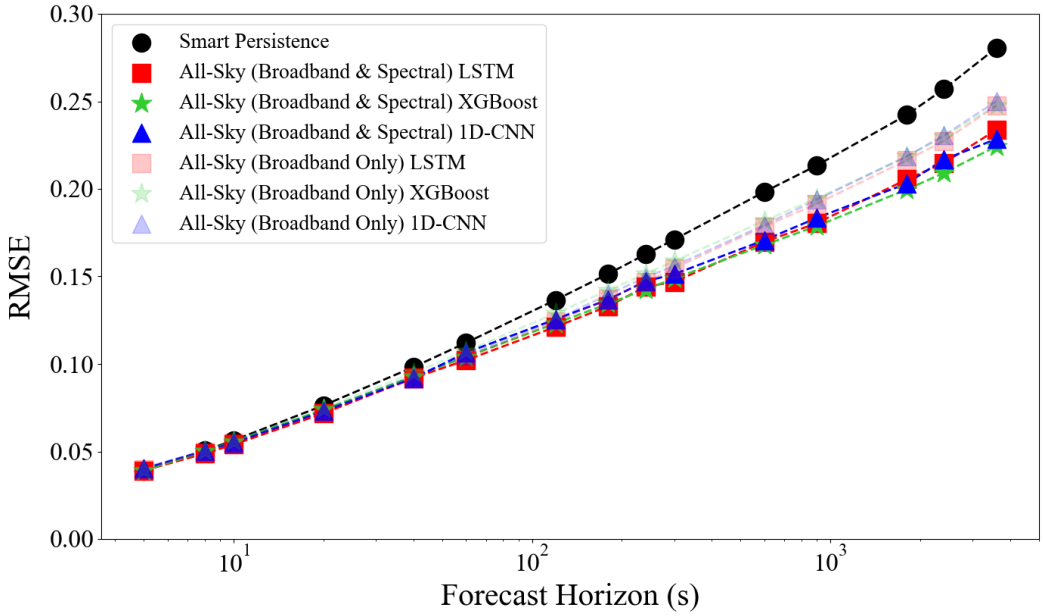

Figure 5.3: RMSEs of the All-Sky (Broadband & Spectral) models using 9 months of testing data. Smart persistence and All-Sky (Broadband Only) model RMSEs provided for reference

well as the computational complexity of each algorithm, imply that model selection depends on the application (i.e., the forecast horizon and operational constraints). Nonetheless, Figure 5.4 shows that whenever spectral irradiance data is available, it should be included as an input for making solar irradiance forecasts. In some cases, it may even be worthwhile to invest in instrumentation capable of measuring spectral irradiance given the forecasting performance gains that can be obtained.

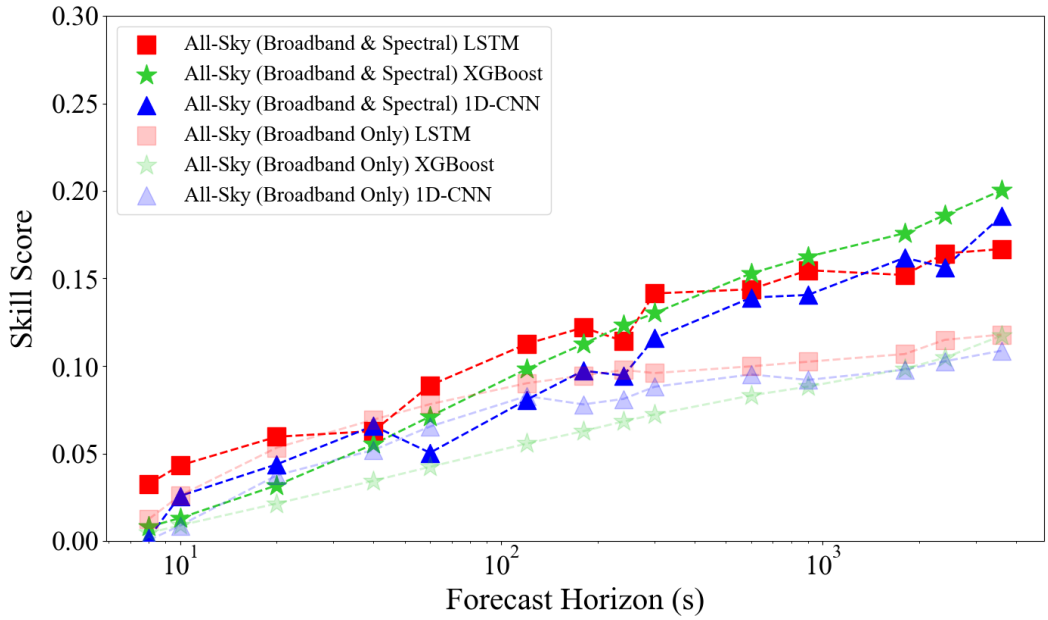

Figure 5.4: Skill scores of the All-Sky (Broadband & Spectral) models using 9 months of testing data. Skill scores of the All-Sky (Broadband Only) models are included for reference

5.3 Irradiance Ramp Regime Classification

As mentioned at the beginning of this chapter, the one-year training dataset was divided into four ramping (variability) regimes – persistent, slow ramping, moderate ramping, and fast ramping – in order to train different ramp regime sub-models. The development of the ramp regime sub-models is motivated by the hypothesis that each specialized sub-model will be better suited for making predictions during periods of its specialty ramping regime compared with the more generic all-sky models. The outputs of the specialized sub-models can then be combined in a diverse and dynamic ensemble for use in all sky conditions; however, this step is beyond the scope of this thesis. This section describes the ramp regime classification approach used to divide the training dataset.

Since irradiance variability consists of a wide range of ramp rates, defining the ramping classes is a non-trivial objective. Large, slow moving clouds create gradual but significant irradiance ramps in the clear-sky index. In contrast, faster moving clouds create sudden and steep irradiance ramps. An irradiance ramping regime classification method must therefore account for the slow ramps, fast ramps, and everything in between. A persistent class must also be defined to account for periods when sky conditions do not change (e.g., clear or overcast skies).

To separate the training dataset into the four ramping regimes, with each regime corresponding to a different range of ramp rates, we can first define the instantaneous clear-sky index slope over a time step, τ, as:

$$m = \frac{\Delta\kappa^*}{\Delta t} = \frac{\kappa^*(t+\tau) - \kappa^*(t)}{\tau} = \frac{\Delta\kappa^*_\tau(t)}{\tau} \tag{54}$$

where $\kappa^*(t)$ and $\kappa^*(t+\tau)$ are two points in along the clear-sky index time series separated by a time step, τ, and their difference is simply the clear-sky index increment. Since irradiance ramps can have a wide range of slopes, different values of τ are needed to capture different ramp rates. That is, smaller time steps (e.g., 1 second) will be more appropriate for extracting fast ramps, while longer time steps (e.g., 10 minutes) will be better for identifying more gradual ramps. Hence, the first challenge in defining ramping regimes is determining appropriate time steps over which to evaluate the slope. For the 1 second resolution training dataset used in this thesis, it was experimentally found that $\tau = 1s$ was effective for defining fast ramps, $\tau = 60s$ was effective for defining moderate ramps, and $\tau = 600s$ was effective for defining slow ramps. Since definitions of each ramp regime are application dependent and largely qualitative, the selected time steps may vary for other applications or datasets. However, it is generally advisable to use the shortest time step possible given the data resolution to define the fastest ramp regime. For instance, when using 30 second resolution data, $\tau = 30s$ should be used to capture the fastest ramps. Using sub-second time steps, however, should be avoided since this time scale mostly detects noise.

Once appropriate time steps have been selected for defining the different ramp regimes, a slope threshold must be determined in order to extract the ramps. If the slope between two points is sufficiently small (i.e., below the threshold), then those two points are not considered to be within a ramp. In contrast, if the slope is sufficiently large (i.e., above the threshold) then it is considered to be a ramp. While this approach to ramp identification is simple, determining appropriate slope thresholds is non-trivial. Considering how the variability statistics change depending on the time scale over which it is being evaluated (as revealed by the clear-sky index increment distributions in Chapter 4), it becomes clear that this question must be answered for each ramp regime's τ individually. For instance, an appropriate threshold for identifying fast ramps over $\tau = 1s$ will be too small for identifying moderate ramps over $\tau = 60s$ – it will not be able to differentiate ramps from noise on this time scale. Thus, for defining the four ramp regimes used in this thesis, three slope thresholds must be determined: one for identifying fast ramps ($\tau = 1s$), one for identifying moderate ramps ($\tau = 60s$), and one for identifying slow ramps ($\tau = 600s$).

The task of finding appropriate thresholds is further complicated by the fact that verifying whether ramps with different slopes are properly identified is a subjective and visual task. Testing different slope thresholds requires one to plot the classified time series and visually validating whether the ramps were properly identified using color coding, for example. This makes an exhaustive trial-and-error search for "optimal" thresholds at each τ impractical. Instead, in this thesis the analysis from the conference proceeding provided in the appendix is used to help guide the threshold selection. When observing the evolution (broadening) of the clear-sky index increment distributions across time scales, it is evident that the changes in irradiance are relatively small for short time scales and large for long time scales. Conveniently, the increasing scale of variability across sub-hourly time scales is captured by the distributions' full widths at half maximum (FWHM) (see appendix). With this in mind, it was experimentally revealed that when used as the slope thresholds for each ramping regime, the FWHM values at $\tau = 1s$, $\tau = 60s$, and $\tau = 600s$ result in excellent ramp identification. Therefore, as a heuristic approach, the FWHM values at each τ are used as the ramp classification thresholds.

Using time steps of 1 second, 60 seconds, and 600 seconds, and the FWHM values from the conference proceeding in the appendix as the slope thresholds, the clear-sky index time series was be separated into the four ramping regimes as shown for the broadband irradiance in Figure 5.5. All pairs of points with slopes that do not exceed any of the thresholds are considered to be from persistent conditions. Any points that are classified as more than one ramp regime are assigned to the faster, more volatile ramping class. For instance, a point that meets the conditions of both fast ramping and moderate ramping will be classified as fast ramping. To handle datapoints which are misclassified (often due to regime transitions), points are compared against neighbouring points to identify and correct outliers and noise. As evident in Figure 5.5, this ramp regime classification method is simple yet highly effective.

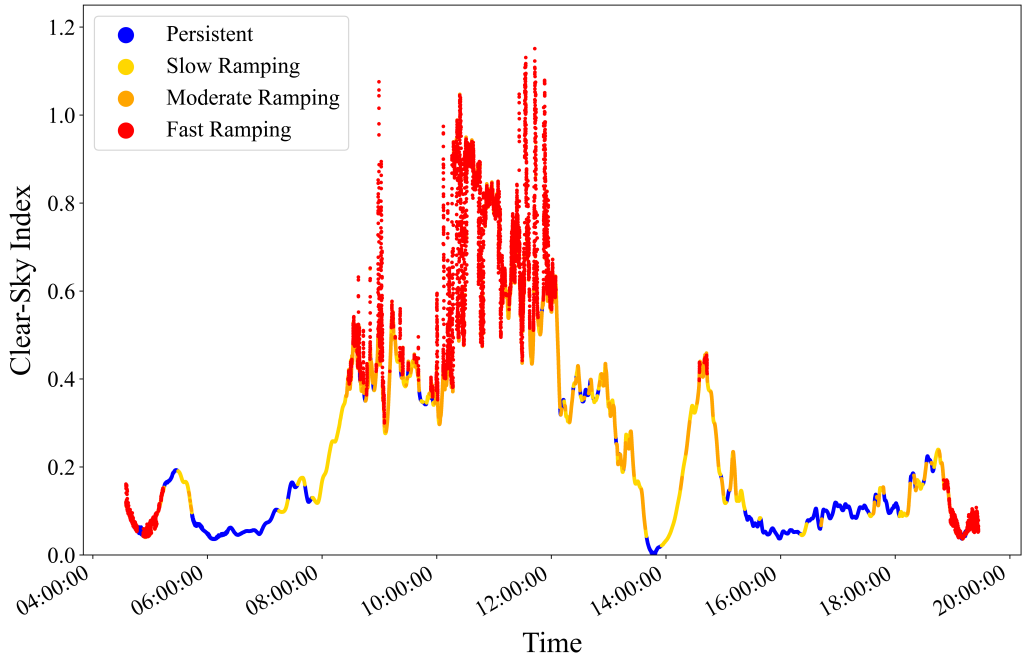

Figure 5.5: Result of implementing the proposed ramp classification method

5.4 Ramp Regime Sub-Models

This section presents the hyperparameter details, training times, and performances of the Ramp Regime LSTM, XGBoost, and 1D-CNN nowcasting sub-models. Since it was found in Section 5.2 that the inclusion of spectral irradiance data yields better nowcasting performance, these Ramp Regime sub-models take the broadband and nine spectral irradiance time series as inputs and make predictions of the future broadband irradiance. The same sub-hourly forecast horizons from Sections 5.1 and 5.2 are also used in this section.

The difference between the All-Sky (Broadband & Spectral) models and the Ramp Regime sub-models is that the former are trained using data from all sky conditions, whereas the latter are trained using only the individual ramp regime subsets of the training data. For instance, the Fast Ramping LSTM sub-model is an LSTM trained using only the training data classified as fast ramping. In a similar manner, the Moderate Ramping, Slow Ramping, and Persistent LSTM sub-models are trained on the moderate ramping, slow ramping, and persistent ramp regime subsets of the training data, respectively. The same process is done for the 1D-CNN and XGBoost algorithms, resulting in a total of 12 Ramp Regime sub-models for each forecast horizon (i.e., four sub-models for each of the three algorithms). Despite being trained with less data than the All-Sky (Broadband & Spectral) models, the Ramp Regime sub-models are

hypothesized to perform better within their specialty ramp regime. To test this hypothesis, the LSTM, XGBoost, and 1D-CNN Ramp Regime sub-models are compared to each other and the All-Sky (Broadband & Spectral) models using their respective ramp regime subsets of the nine month testing dataset. For example, the Fast Ramping sub-models and the All-Sky (Broadband & Spectral) models are tested and compared using the fast ramping subset of the testing data.

To facilitate a meaningful comparison, the Ramp Regime LSTM sub-models were designed with the same hyperparameters as the All-Sky (Broadband & Spectral) LSTM models. Specifically, the Ramp Regime LSTM sub-models contain a first hidden layer with 64 LSTM cells, a second hidden layer with 16 LSTM cells, and a third hidden layer with 8 standard neurons. These sub-models also use a lookback window of 10 seconds for all of the sub-hourly forecast horizons investigated. This way, the only difference between the All-Sky (Broadband & Spectral) LSTM models and the Ramp Regime LSTM sub-models is the training data used for each.

The Ramp Regime XGBoost sub-models were also trained using the same hyperparameters as the All-Sky (Broadband & Spectral) XGBoost models. Hence, the Ramp Regime XGBoost sub-models were allowed a maximum tree depth of 6 and contained 200 estimators (XGBoost trees). Like the LSTMs, the only difference between the All-Sky (Broadband & Spectral) XGBoost models and the Ramp Regime XGBoost sub-models is the training data used.

Lastly, the Ramp Regime 1D-CNN sub-models were also trained with the same hyperparameters as the All-Sky (Broadband & Spectral) 1D-CNN models. As such, each sub-model contains 64 convolutional neurons in the first hidden layer, 16 convolutional neurons in the second hidden layer, and 8 standard neurons in the third hidden layer. A 10 second lookback window was again used by each of the 1D-CNN sub-models for making predictions across all of the forecast horizons. Therefore, the only difference between the All-Sky (Broadband & Spectral) 1D-CNN models and the Ramp Regime 1D-CNN sub-models is the training data used.

5.4.1 Ramp Regime Sub-Model Training and Testing Times

Similar to the different All-Sky models presented in Sections 5.1 and 5.2, the evaluation of the Ramp Regime sub-models begins with an assessment of the training and testing times. Naturally, each individual Ramp Regime sub-model will take less time to train than an All-Sky model given that only a fraction of the entire training dataset will be used. However, the ultimate goal of training specialized Ramp Regime sub-models is to combine them in a diverse ensemble for use in all sky conditions. Each Ramp Regime sub-model should therefore be

considered as part of a larger ensemble, where all four Ramp Regime sub-models are needed to complete the ensemble for each algorithm. Hence, the total training time of all four Ramp Regime sub-models should be compared against the All-Sky models' training times. To this end, Table 5 shows the total training times of all Ramp Regime sub-models for each algorithm. The All-Sky (Broadband Only) and All-Sky (Broadband & Spectral) model training times from Sections 5.1 and 5.2 are also included for comparative purposes.

Table 5: Average training times for the All-Sky (Broadband Only), All-Sky (Broadband & Spectral), and Ramp Regime nowcasting models

Model Category	**Model Training Time (s)**		
	LSTM	**XGBoost**	**1D-CNN**
All-Sky (Broadband Only)	5950	145	1150
All-Sky (Broadband & Spectral)	6030	180	1210
Ramp Regimes	4475	485	1130

Interestingly, the training times in Table 5 show mixed results. For the LSTM, the total training time decreases by nearly 25% when the data is divided into the ramp regime subsets. In contrast, the total XGBoost training time increases by roughly 270% and the total 1D-CNN training time remains reasonably constant. These results reflect each algorithm's efficiency when dealing with large volumes of training data – XGBoost is most efficient with larger volumes of data, LSTM is most efficient with smaller volumes of data, and 1D-CNN is relatively unaffected by data volume. Hence, when considering training time as an indicator of computational demand, training Ramp Regime sub-models instead of All-Sky models decreases demand for LSTM, increases demand for XGBoost, and does not impact demand for 1D-CNN. While these results can have practical implications in terms of cloud computing costs, the algorithm ranking does not change between the All-Sky and Ramp Regime approaches. That is, despite the changes in training times, XGBoost is still the fastest, LSTM is still the slowest, and 1D-CNN still lies between.

Next, to assess whether the Ramp Regime sub-models are suitable for real-time operation, Table 6 presents their total testing times. The testing times for the All-Sky (Broadband Only) and All-Sky (Broadband & Spectral) models are also included in this table for comparison. Each Ramp Regime sub-model is tested using only data from its specialty ramp regime (e.g., a Fast Ramping sub-model is testing using only the fast ramping portion of the test dataset). Therefore, to ensure a fair comparison with the All-Sky models, Table 6 shows the total testing time for all Ramp Regime sub-models, which is the same number of predictions made by the All-Sky models.

Table 6: Average testing times for the All-Sky (Broadband Only), All-Sky (Broadband & Spectral), and Ramp Regime nowcasting models

Model Category	Model Testing Time (s)		
	LSTM	**XGBoost**	**1D-CNN**
All-Sky (Broadband Only)	540	4	190
All-Sky (Broadband & Spectral)	555	4	195
Ramp Regimes	425	4	180

The results in Table 6 are mostly expected, with the exception of LSTM. Since these testing times reflect how long it takes to make approximately 9.8 million predictions using each approach, it was expected that the results would not differ significantly. This is the case for XGBoost and 1D-CNN, where the testing times are roughly the same. For LSTM, however, the Ramp Regime sub-models are able to make the 9.8 million predictions nearly 25% faster than the All-Sky LSTMs. Overall, Table 6 shows that the Ramp Regime sub-model approach is at least equally viable for real-time operation compared to the more traditional All-Sky models.

As a final note on the Ramp Regime sub-model testing times, an additional consideration is the extra time needed to combine the sub-models in an ensemble. Depending on the sophistication of the ensembling method, this can range from a small increase (e.g., simply averaging the sub-model predictions) to a more significant increase (e.g., dynamic ensembling). However, since combining the sub-models in an ensemble is beyond the scope of this thesis, so, too, is this consideration.

5.4.2 Ramp Regime Sub-Model Forecasting Performances

Having confirmed that the Ramp Regime sub-models are no more computationally (or economically) expensive than the All-Sky models, this section evaluates the Ramp Regime sub-model forecasting performances within their respective specialty ramp regimes. The Ramp Regime sub-models are compared against the All-Sky (Broadband & Spectral) models, since these were found to outperform the original benchmark All-Sky (Broadband Only) models. To clarify, if the Ramp Regime sub-models are able to outperform the All-Sky (Broadband & Spectral) models, then they will also outperform the All-Sky (Broadband Only) models.

Beginning with the persistent regime, Figure 5.6 shows the RMSEs of the Persistent Ramp Regime sub-models, All-Sky (Broadband & Spectral) models, and smart persistence model

using the subset of testing data that is classified as persistent. These results show how well each model performs within the persistent regime; all fast, moderate, and slow ramping periods are filtered out of the testing dataset used here. While there are only minor differences between the Persistent Ramp Regime sub-models and the All-Sky (Broadband & Spectral) models for each machine learning algorithm, the former appear to show slight improvements.

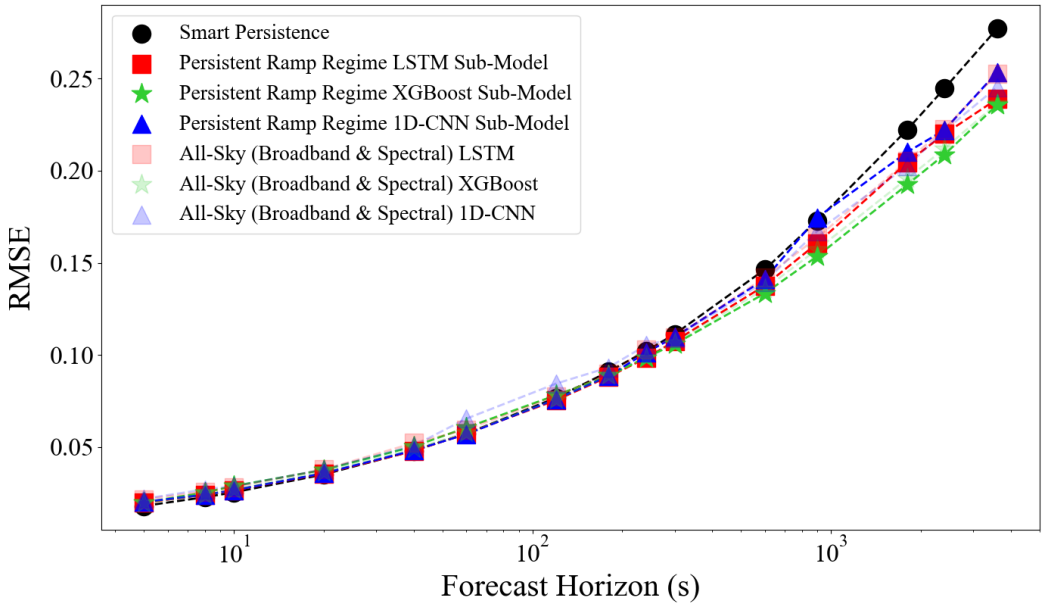

Figure 5.6: RMSEs of the Persistent Ramp Regime sub-models, All-Sky (Broadband & Spectral) models, and smart persistence model within the persistent regime

To better visualize how the two sets of models compare with each other and the smart persistence model, the skill score is once again needed. Figure 5.7 shows the skill scores of the Persistent Ramp Regime sub-models and the All-Sky (Broadband & Spectral) models when evaluated on persistent test data. While there are some inconsistencies across the sub-hourly forecast horizons, these results show several improvements with the Ramp Regime sub-models. For instance, the LSTM and 1D-CNN Ramp Regime sub-models are able to achieve positive skill scores at shorter forecast horizons that the All-Sky models. Though small in magnitude, these positive skill scores at very short horizons are a significant result given that the reference smart persistence model is especially well suited for short time scales and persistent conditions. Next, comparing the machine learning algorithms to each other, there is again some inconsistency across the selected forecast horizons which indicates that model selection may be application dependent. For instance, neglecting computational demands, LSTM and 1D-CNN may be best suited for forecasting on very short horizons (e.g., under 2 minutes), whereas XGBoost is better for longer horizons. However, it should be noted that fine tuning each model's hyper-

parameters for each individual forecast horizon may affect these results and the consistency of each algorithm.

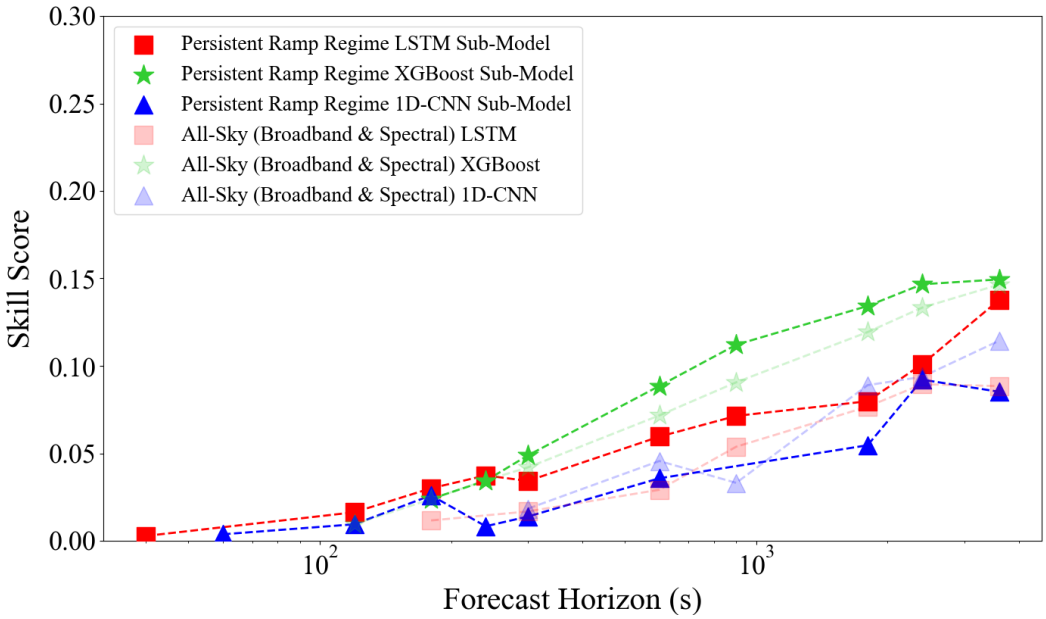

Figure 5.7: Skill scores of the Persistent Ramp Regime sub-models and All-Sky (Broadband & Spectral) models within the persistent regime

While the Persistent Ramp Regime sub-model performances are good compared to the All-Sky (Broadband & Spectral) models, the persistent regime is the least interesting in the context of PV nowcasting. When the sky conditions are stable (persistent), there is little threat posed to PV power stability. In contrast, more variable conditions like slow, moderate, and fast ramping can have detrimental impacts on PV power. With this in mind, Figures 5.8 and 5.9 show the RMSEs and skill scores, respectively, for the Slow Ramping Ramp Regime sub-models and All-Sky (Broadband & Spectral) models when evaluated using slow ramping test data.

Although some inconsistencies can be found in Figures 5.8 and 5.9, the Slow Ramping Ramp Regime sub-models generally outperform the All-Sky models in slow ramping conditions. The different machine learning algorithms perform similarly with each other, though XGBoost has a slight advantage on average. The higher skill scores in Figure 5.9 (slow ramping conditions) compared to Figure 5.7 (persistent conditions) are, in part, due to how the smart persistence model is better suited for persistent conditions. Nonetheless, these higher skill scores are meaningful as they represent how significant of an improvement the Slow Ramping Ramp Regime sub-models make compared to the smart persistence model. Moreover, the Ramp Regime sub-models show a clear improvement when compared with the All-Sky models for slow ramping

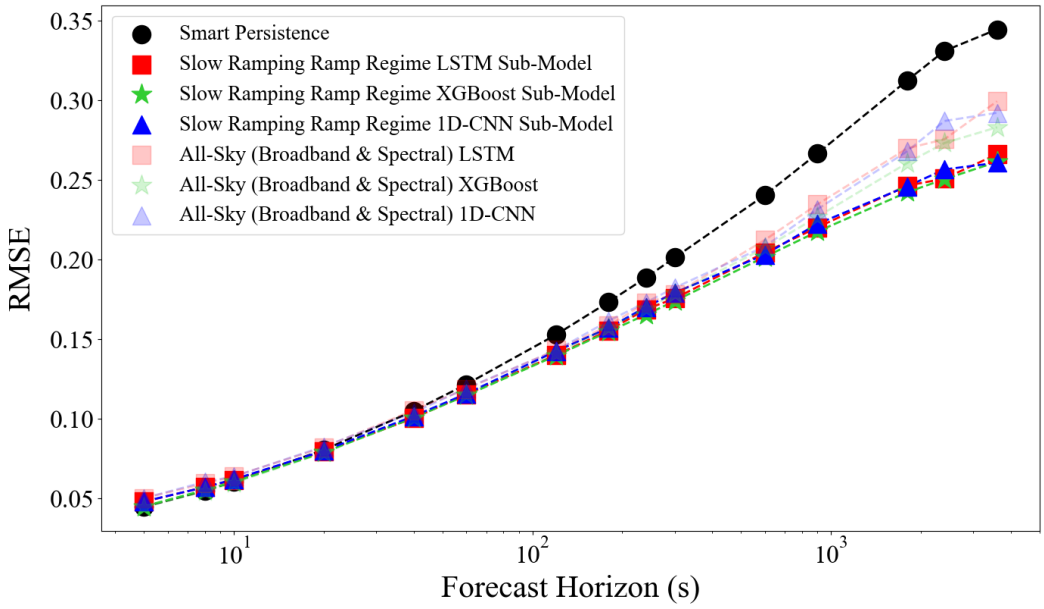

Figure 5.8: RMSEs of the Slow Ramping Ramp Regime sub-models, All-Sky (Broadband & Spectral) models, and smart persistence model within the slow ramping regime

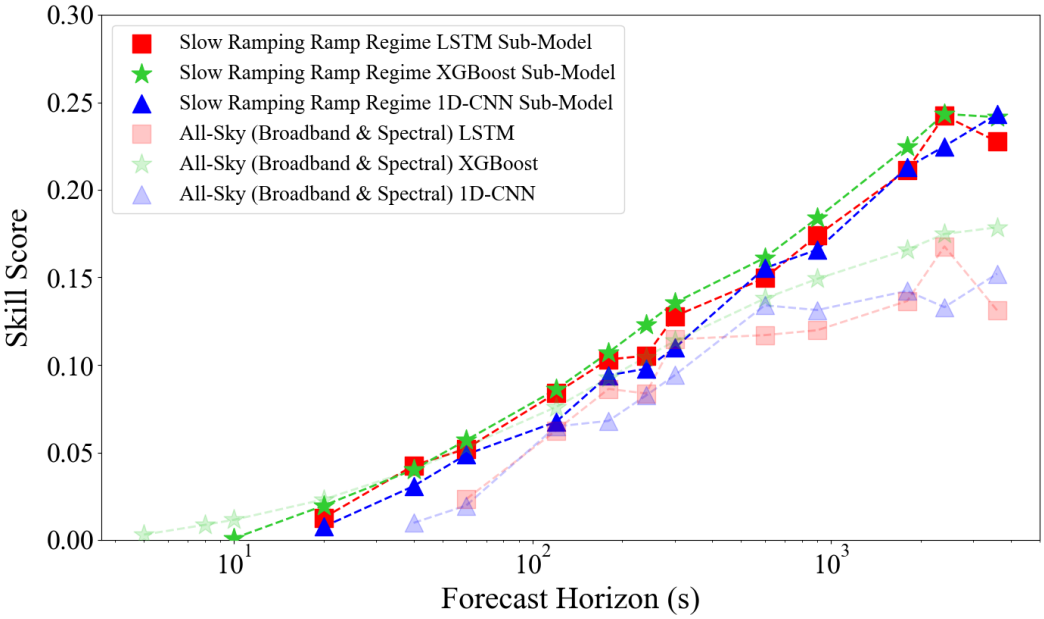

Figure 5.9: Skill scores of the Slow Ramping Ramp Regime sub-models and All-Sky (Broadband & Spectral) models within the slow ramping regime

conditions, unlike for persistent conditions where the results were mixed. This implies that the more variable the ramp regime, the more important it is to have specialized models.

Continuing the analysis, Figures 5.10 and 5.11 present the RMSEs and skill scores, respectively, for the Moderate Ramping Ramp Regime sub-models and All-Sky models in moderate ramping conditions. Similar to the persistent and slow ramping cases, these plots show the performances of each model in only the moderate ramping portion of the testing dataset.

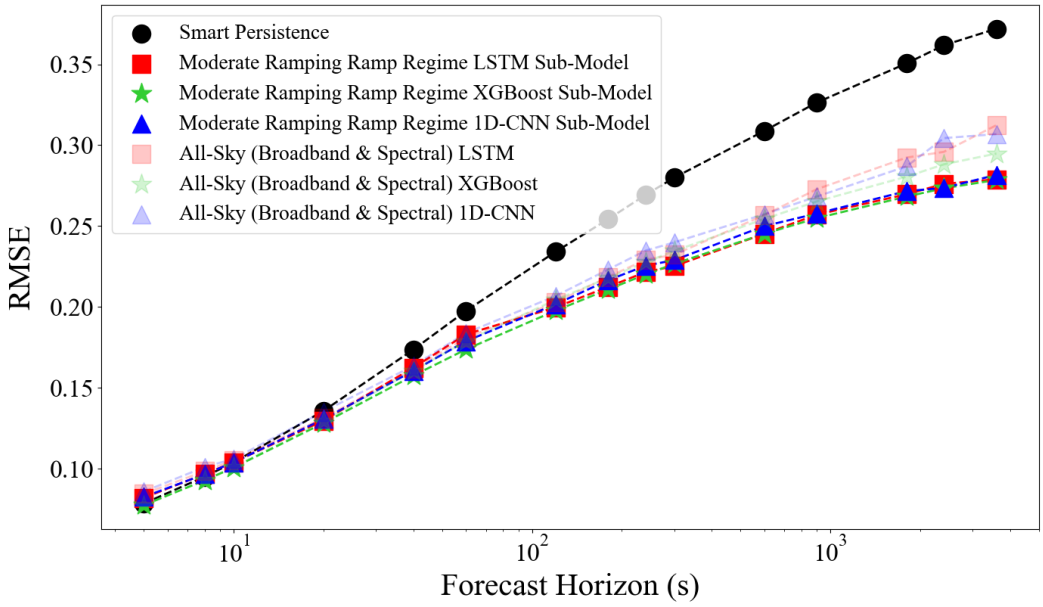

Figure 5.10: RMSEs of the Moderate Ramping Ramp Regime sub-models, All-Sky (Broadband & Spectral) models, and smart persistence model within the moderate ramping ramp regime

The moderate ramping RMSEs and skill scores support the trends seen by comparing the persistent and slow ramping cases. Specifically, the smart persistence model RMSEs are generally higher for moderate ramping than both persistent and slow ramping, resulting from the smart persistence model's struggle with higher variability. This leads to greater separation between the smart persistence RMSEs and both the All-Sky and Ramp Regime sub-models' RMSEs. This increased RMSE separation corresponds to generally higher skill scores, which can be seen in Figure 5.11. Comparing the All-Sky and Moderate Ramping Ramp Regime sub-models, there are a few inconsistencies but the specialized Ramp Regime sub-models again outperform the All-Sky models. The individual machine learning algorithms have similar performances, though the advantage belongs to XGBoost at shorter forecast horizons. Considering how computationally lightweight XGBoost is compared to LSTM and 1D-CNN, XGBoost is an attractive option for making very short term forecasts in variable conditions. However, it should again be noted that fine tuning the hyperparameters of each model for every forecast horizon may affect these results.

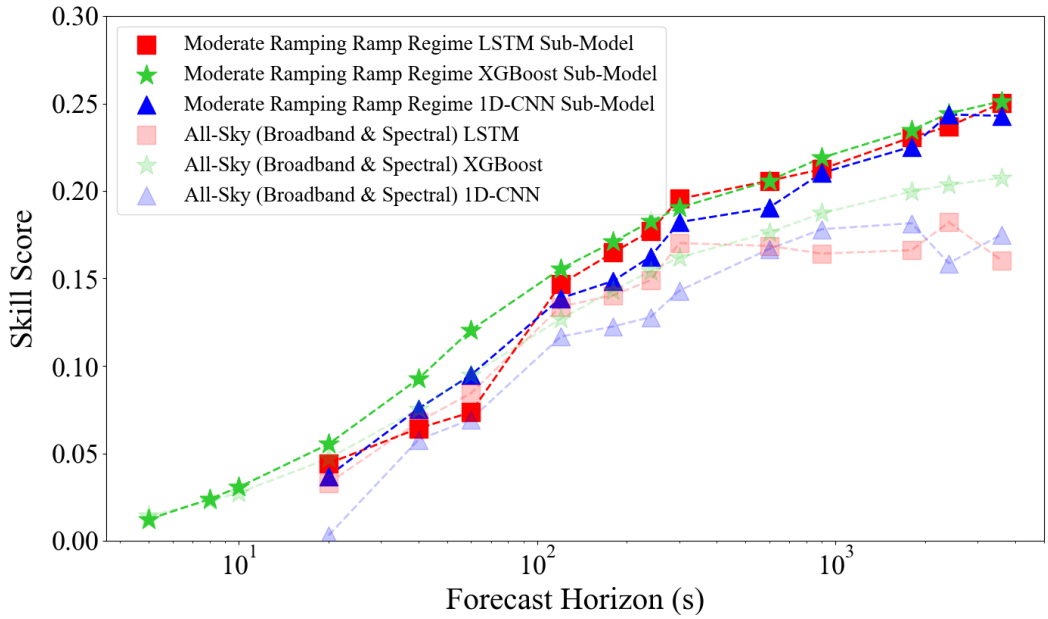

Figure 5.11: Skill scores of the Moderate Ramping Ramp Regime sub-models and All-Sky (Broadband & Spectral) models within the moderate ramping ramp regime

Finally, as arguably the most interesting case for PV nowcasting, Figures 5.12 and 5.13 present the RMSEs and skill scores, respectively, for the Fast Ramping Ramp Regime sub-models and All-Sky models in fast ramping conditions. The fast ramping regime can be considered the most interesting as it is the most volatile, threatening to PV power stability, and difficult to predict according to the smart persistence model. However, the challenge of making accurate forecasts in fast ramping conditions also offers a greater room for improvement. This is illustrated by Figure 5.12, in which the smart persistence model RMSEs are higher than in any of the previous ramping regimes. These high RMSEs for the smart persistence model allow for higher skill scores to be achieved by the All-Sky and Ramp Regime models, as shown in Figure 5.13.

Again, with only a few exceptions, the Fast Ramping Ramp Regime sub-models outperform the All-Sky models across the sub-hourly forecast horizons. The performance difference between the Ramp Regime and All-Sky models is especially evident at longer time horizons, with the specialized Fast Ramping Ramp Regime sub-models achieving up to almost 28% skill scores. In agreement with the other ramp regime test cases, the machine learning algorithms show similar performances to each other, though XGBoost is superior on average. While it should again be noted that the model hyperparameters are not optimized for each forecast horizon, XGBoost does appear to dominate in terms of both forecasting performance and computational efficiency.

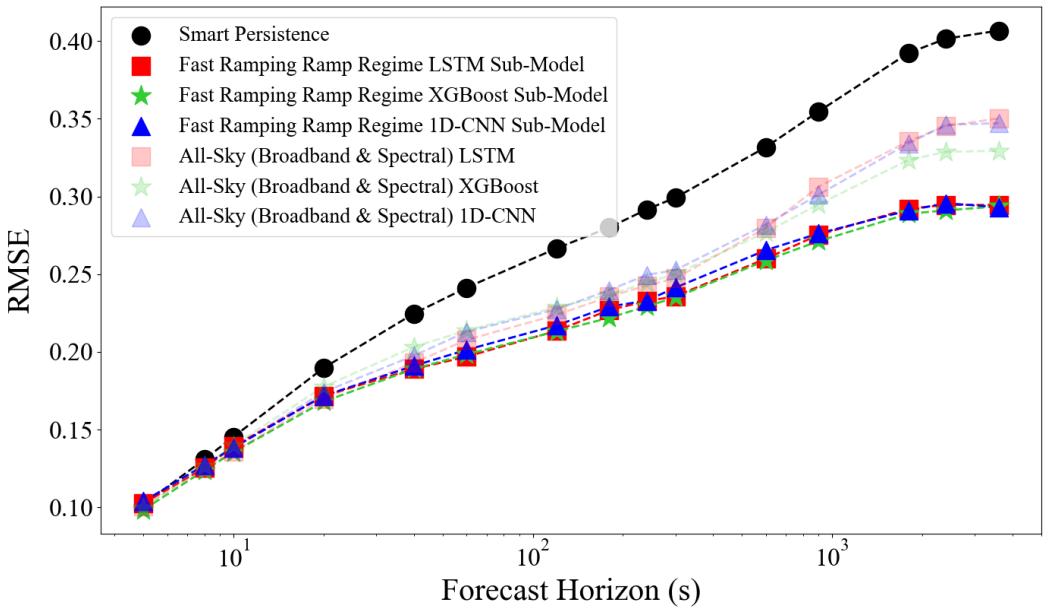

Figure 5.12: RMSEs of the Fast Ramping Ramp Regime sub-models, All-Sky (Broadband & Spectral) models, and smart persistence model within the fast ramping ramp regime

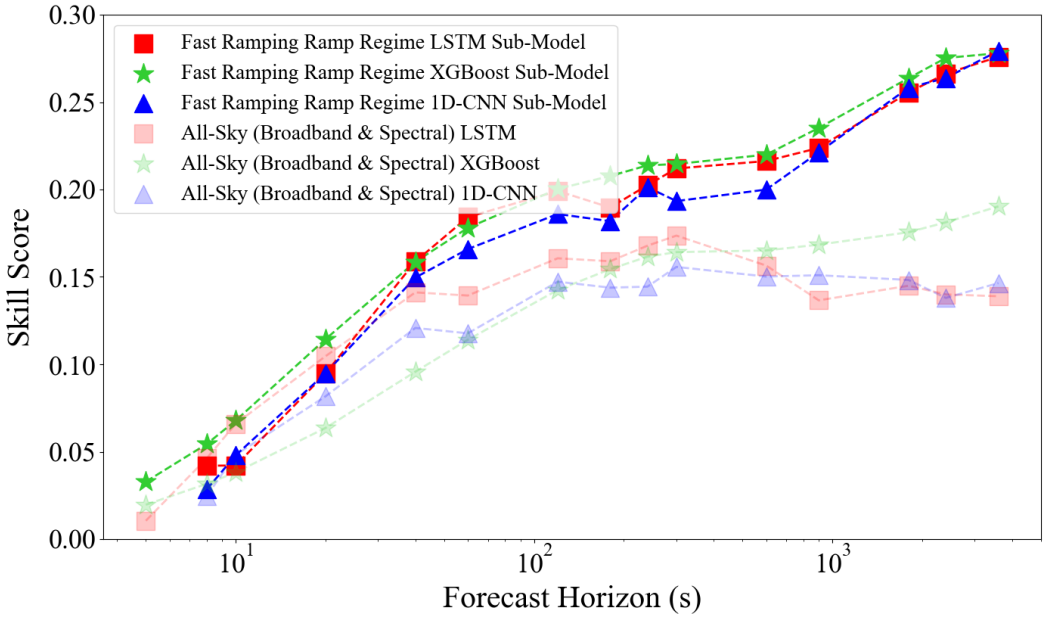

Figure 5.13: Skill scores of the Fast Ramping Ramp Regime sub-models and All-Sky (Broadband & Spectral) models within the fast ramping ramp regime

Overall, the Ramp Regime sub-models show performance improvements over the All-Sky (Broadband & Spectral) models within their specialty ramp regimes, which confirms the hy-

pothesis stated at the beginning of this section. While the Ramp Regime sub-models only slightly outperformed the All-Sky models in the persistent regime, the differences are more pronounced as the variability increases to slow, moderate, and fast ramping conditions. Conveniently, this trend correlates with the significance level of each ramp regime in the context of PV nowcasting – the more volatile the ramp regime is, the more of a threat to PV stability it poses. Hence, the large performance improvements over the All-Sky models in fast, moderate, and slow ramping conditions are of greater significance.

As a final note in this section, the test cases for the Ramp Regime sub-models only considered test data corresponding to each sub-model's specialty. However, in practice it is not likely that such ramp regime isolation will occur for more than a few seconds or minutes at a time. In contrast, real-world solar irradiance data contains many transitions from one ramp regime to another, which is not accounted for in this study. Therefore, for fully operational nowcasting, a sophisticated Ramp Regime Ensemble would be needed to navigate regime transitions while fully harnessing the potential of each sub-model.

Chapter 6: Conclusions

6.1 Future Work

This section provides suggestions for future extensions of this project, with particular emphasis further developing the nowcasting models for real-world deployment.

Although machine learning algorithms are praised for their ability to learn complex data patterns without explicit instructions from programmers, they do require their hyperparameters to be defined. Hyperparameters include, but are not limited to, the number of nodes in a layer, the number of hidden layers in a network, lookback window size, and learning rates. Given the wide range of hyperparameters that can be defined for a model – and the many possible values for each – hyperparameter tuning requires investigating a vast search space. Moreover, it was found from an early attempt at optimizing the LSTM, XGBoost, and 1D-CNN hyperparameters that different forecast horizons require different hyperparameters, which further expands the search space. As such, extensive hyperparamter tuning for each of the nowcasting models presented in this thesis is left as a possible future work.

In accordance with the ROPES guideline, solar forecasting models should be reproducible, operational, probabilistic and/or physically-based, an ensemble, and evaluated using the skill score. The nowcasting models presented in Chapter 5 of this thesis meet several of these requirements, but not all. For instance, the nowcasting models in this work are deterministic rather than probabilistic, since they produce single point forecasts rather than probability distributions. Additionally, the Ramp Regime sub-models presented in Chapter 5 provide the building blocks for a sophisticated ensemble; however, the creation of a dynamic ramp regime ensemble is beyond the scope of this thesis. Therefore, it is left as future work to both combine the Ramp Regime sub-models in ensembles and to convert the deterministic forecasts to probabilistic ones.

Finally, the Ramp Regime sub-models presented in this thesis yield promising results with the 1 second resolution irradiance data. However, high temporal resolution pyranometers are costly and not widely available at different locations. While the use of pyranometers may be practical for large PV plants where their costs are easily absorbed, they are far less practical for use in distributed (e.g., rooftop) PV at the grid edge. As a more widely applicable alternative, it is left as a future work to investigate how the ramp regime approach to solar forecasting would perform if PV power data is used instead of solar irradiance measurements.

6.1.1 Conclusion

Overall, this thesis provides a detailed investigation of spectral and broadband solar irradiance variability statistics on all time scales relevant to PV generation and forecasting. The variability induced by cloud dynamics was found to be spectrally dependent, with some regions of the solar spectrum being more highly variable than others. Additionally, the solar variability was found to be temporally dependent, with variability increasing as the time step gets longer. Significant differences were found between the clear-sky index increment distributions across sub-hourly time scales in particular. This short term variability was found to be limited by the different speeds at which clouds can move and sky conditions can change. The temporal dependence of the clear-sky index increment distributions shows evidence of power law scaling, which is indicative of chaotic processes.

The irradiance variability statistical analysis results were leveraged to develop a new irradiance ramp regime classification method. The temporal dependence of the variability statistics (namely, the FWHM) enabled a heuristic approach to be developed for tuning a slope threshold parameter used for deciding whether points an the irradiance time series belonged to a given ramping regime. Since faster, more volatile irradiance ramps pose a greater threat to PV power stability than slower ramps, the fastest ramp regimes are prioritized when points are assigned to multiple regimes.

Three sets of solar nowcasting models were developed and evaluated under a variety of test conditions. Within each set, the LSTM, XGBoost, and 1D-CNN machine learning algorithms were used as the base prediction algorithms and compared in terms of computational demand and forecasting performance. The first set of models, referred to as the All-Sky (Broadband Only) models, were used as benchmarks. The second set of models, referred to as the All-Sky (Broadband & Spectral) models, improved upon the performance of the All-Sky (Broadband Only) models by incorporating nine spectral irradiance inputs. Demonstrating that the inclusion of spectral irradiance can improve forecasts of the broadband irradiance is the first contribution to the field of solar forecasting from this thesis. Finally, the third set of models, referred to as the Ramp Regime sub-models, were trained and tested using data that was divided into four different ramp regimes by the ramp regime classification method proposed in this work. The resulting Ramp Regime sub-models outperformed the All-Sky (Broadband & Spectral) models within their respective specialty ramping regimes, which is the second contribution to the field of solar forecasting from this thesis.

The nowcasting models presented in this thesis were developed under the guidance of the

ROPES principles. While not meeting all of the criteria, the presented nowcasting models aim to lay the foundations for full ROPES adherence. In terms of reproducibility, it is the intention to make the dataset used in this work – or some partial and/or synthetic version of it – publically available. Typical operational requirements for nowcasting models have been accounted for through an investigation and discussion of computational demand and prediction times. The models inherently account for the physical processes (scattering and absorption) driving irradiance variability, though their exact origins need to be resolved. It remains as future work to convert the nowcasting model outputs from deterministic to probabilistic predictions, though there are existing techniques for this [42]. The building blocks for a diverse nowcasting ensemble have been created with the Ramp Regime sub-models, though it also remains as future work to realized a highly effective, dynamic Ramp Regime Ensemble. Finally, all model performances in this thesis have been evaluated using the skill score with smart persistence as the reference model. With the completion of the few remaining future works identified above, the models presented in this thesis show great potential for advancing the field of solar nowcasting and our understanding of the solar resource.

Chapter 7: Appendix

7.0.1 Scope and Impact

The following publication investigates the fluctuations in spectral and broadband irradiance cause by clouds from a statistical perspective. A high resolution spectral and broadband irradiance database from a Spectrafy SolarSIM-G containing 6 months of high resolution (250 ms) ground-based measurements taken in Ottawa, Canada is described used for the analysis in this work. The known diurnal and orbital irradiance trends are removed from the data using clear-sky normalization (i.e., the clear-sky index), thus isolating the variability due strictly to cloud motion. The database is parsed into 7 different sky conditions based on the ultraviolet and infrared spectral irradiance measurements, and the clear-sky index distributions of each sky condition are investigated. This article then assesses the spectral and broadband irradiance variability through the clear-sky index increment distributions across sub-hourly time scales. This assessment revealed a spectral dependence of the variability, indicating that the clouds impact the various spectral irradiance components in notably different ways. It was also found that the variability is temporally dependent, following power law scaling behaviour at the sub-hourly time scales that were probed.

7.0.2 Author Contributions

Nick Anderson: All data analysis included in this article and lead-writing was performed by me, but was guided and supported by my co-authors.

Viktar Tatsiankou: As a collaborator from Spectrafy, Viktar has set up the instrumentation at the Ottawa site and provided me with the database used in this work. His previous research contributions have provided the foundations for this work, including the sky condition classification approach. He also assisted in the manuscript review process.

Karin Hinzer: As the director of the University of Ottawa's SUNLAB, Karin oversaw and guided my research. She also assisted in preparing and reviewing the final manuscript.

Richard Beal: As a collaborator from Spectrafy, Richard has set up the instrumentation at the Ottawa site and provided me with the database used in this work. He also assisted in the manuscript review process.

Henry Schriemer: As my supervisor and the associate director of the University of Ottawa's SUNLAB, Henry guided my research and assisted with interpreting results and preparing the final manuscript.

7.0.3 Publication - Photonics West 2022

Nick Anderson, Viktar Tatsiankou, Karin Hinzer, Richard Beal, and Henry Schriemer, "Probabilistic description of short-term cloud dynamics from rapid sampling of the solar spectral irradiance", Proc. SPIE 11996, Physics, Simulation, and Photonic Engineering of Photovoltaic Devices XI, 119960B (4 March 2022); doi: 10.1117/12.2616231.

Probabilistic Description of Short-Term Cloud Dynamics from Rapid Sampling of the Solar Spectral Irradiance

Nick Anderson[a], Viktar Tatsiankou[a,b], Karin Hinzer[a], Richard Beal[b], Henry Schriemer*[a]
[a]SUNLAB, University of Ottawa, 800 King Edward Ave, Ottawa, ON, CAN, K1N 6N5
[b]Spectrafy, 4 Florence St, #204, Ottawa, ON, CAN, K2P 0W7

ABSTRACT

Solar irradiance variability due to stochastic cloud dynamics can cause unwanted fluctuations in the output voltage of photovoltaic (PV) modules. These dynamics must in particular be understood at very-short and short time scales if grid interconnection and generation/load balance requirements are to be maintained for PV distributed across the grid edge. Using a recently-created database for Ottawa, Canada, a 6-month longitudinal study was conducted with a specific focus on cloud dynamics. A spectral pyranometer was used to derive full-range spectral and broadband global horizontal irradiance under all sky conditions every 250 ms. Exploiting the infrared (IR) measurement channel of this software-augmented multi-filter radiometer allowed the cloud dynamics to be probed across time scales ranging from the sub-second to ~30 minutes. Seven distinct sky conditions were self-consistently determined without sky imaging. Probability distributions, established via kernel density estimates (KDE), allowed the statistical dependence of these conditions on the spectral clear-sky index to be found. The stochastic nature of the spectral irradiance variability was probed using spectral clear-sky index increments, over time steps that were found to span three distinct variability regimes.

Keywords: Solar irradiance variability, cloud dynamics, spectral pyranometer, probability density, index increments

1. INTRODUCTION

Photovoltaics (PV) is a key enabler of the global energy transformation.[1, 2] Within the emerging smart grid context,[3] this sustainable generation technology may be regionally capable of supporting the entire electrical load with a high penetration of distributed PV fleets if one follows a firm forecast overbuild/curtail/storage strategy to address the intermittent nature of the solar resource.[4] In this strategy, the short-term forecast uncertainty associated with weather-driven solar irradiance fluctuations is reduced by exploiting geographic dispersion to realize firm power generation. However, for PV systems as distributed energy resources (DER) at the edge of the distribution grid, the interconnection constraints[5] require "nodal approaches for the computation of power grid congestion constraints and power flow solutions".[6] To facilitate the energy transition, effective DER participation in local energy markets must be addressed as a fully transactive energy system under uncertainty.[7] For many grid services, in both settlement and control, this requires intra-hour forecasting (also known as "nowcasting"), which ranges from a few seconds to an hour and encompasses very-short-term and short-term categories.[8] A recent review has focused on the often neglected temporal and spatial resolution aspects of solar irradiance and power forecasting methods,[9] but it is safe to say that probabilistic solar forecasting capabilities are still immature.[10]

Maturation of probabilistic solar forecasting is a process that requires advances in instrumentation, accessibility to data, application, and assessment. A definitive approach to solar forecasting research practice "intended to facilitate comparison, comprehension, and communication within the solar forecasting field, and [to] speed up its development",[11] is the "ROPES" guideline (which stands for reproducible, operational, probabilistic/physical, ensemble, and skill) proposed by Dazhi Yang. We take this formalism as our goal for an operational implementation of PV generation nowcasting at the grid edge. Here we address its enabling requirements with an instrumentation and data analysis focus, employing a new spectral irradiation database for Ottawa, Canada. This database provides an ongoing comprehensive meteorological and solar spectral irradiance record, with measurements every 250 ms, beginning 1 June 2021 and here ending 30 November 2021. We describe the novel spectral pyranometer that acquired this data, and which enables the

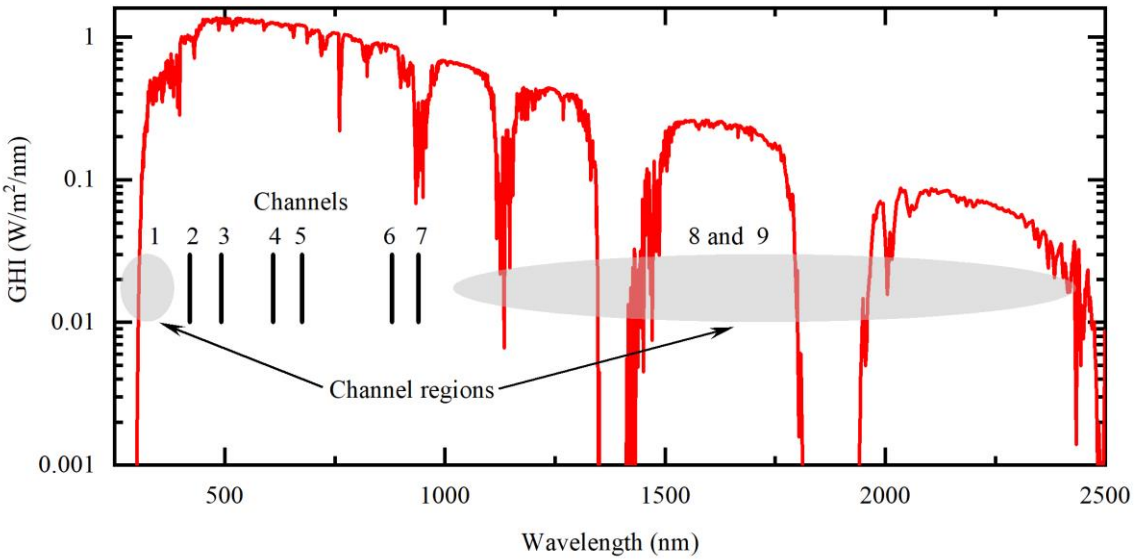

Figure 1. Spectral irradiance at 1:03 pm on 3 August 2021; the wavelength range from 2500 to 4000 nm is not shown. The measurement channel locations are as indicated. The precise locations of channels 1, 8 and 9 are not given for proprietary reasons.

ongoing self-consistent determination of sky conditions. Our preliminary data analysis first addresses the distribution of sky conditions over this 6-month period, and then statistically describes cloud dynamics as a function of clear-sky index increments for time scales less than 10 minutes.

2. INSTRUMENTATION AND SKY CLASSIFICATION

The spectral pyranometer employed in this work is a customized version of Spectrafy's SolarSIM-G, whose general properties have been described elsewhere.[12] It is a software-augmented multi-filter radiometer that measures the spectral global horizontal irradiance (GHI) using narrow bandpass filters paired with calibrated photodetectors. A multiplexer sequentially selects the voltage signal from each spectral channel and feeds it to an analog-to-digital converter sampling at 60 Hz, which is sufficient to accurately capture the fastest of transients.[13] The nine wavelength channels are illustrated in Figure 1 with respect to the spectral GHI for Ottawa at 1:03 pm on 3 August 2021; for proprietary reasons, the locations of the ultraviolet (UV) and infrared (IR) channels cannot be precisely disclosed. In combination with measurements of ambient pressure, relative humidity, and air temperature, this information is employed by a radiative transfer model to self-consistently derive atmospheric optical parameters for the reconstruction in real-time of spectral GHI from 280 nm to 4000 nm. The basic design principles have been previously described.[14] Measurements from the infrared channels are used to classify the clouds by their water morphology, with a uniform cloud model employed to estimate cloud optical depth and to subsequently compute cloud transmittance in the 1000-4000 nm range. The efficacy of this approach has been validated against reference instruments under a full range of sky conditions.[15]

3. RESULTS AND ANALYSES

3.1 The clear-sky index

Denoting the spectral GHI time series measured at wavelength channel λ by $E(\lambda;t)$, the clear-sky spectrum $E_{clr}(\lambda;t)$ is determined as part of the spectral reconstruction process and derives from the same set of underlying atmospheric optical parameters as the actual spectrum, but without the correction for clouds. This self-consistent determination allows the sky condition to be accurately modeled by a spectral clear-sky index

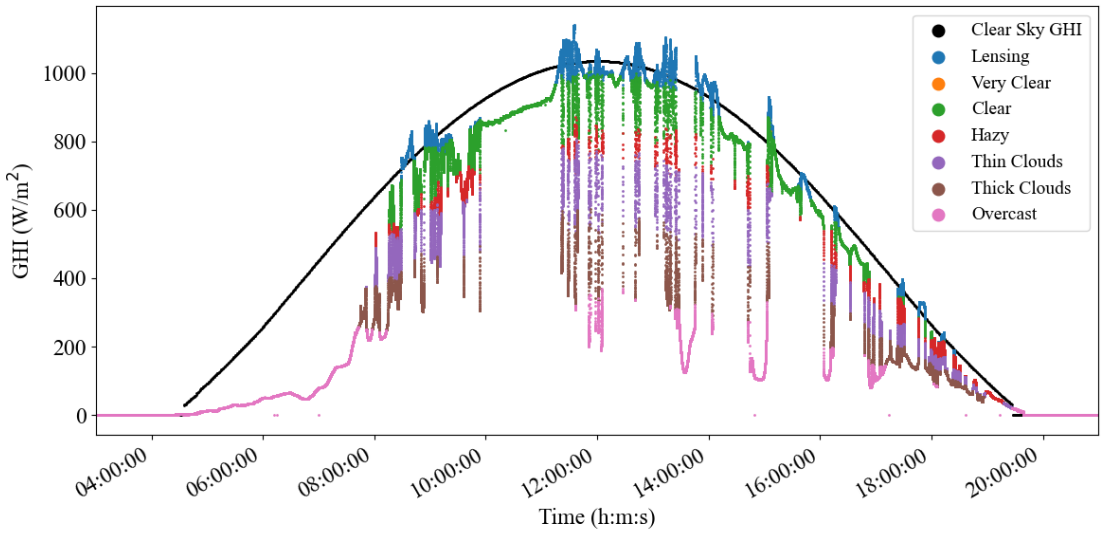

Figure 2. The GHI observed on 1 June 2021 in Ottawa, Canada, color coded for sky condition; the smooth black curve is the GHI that would be observed under clear sky conditions.

$$\kappa^*(\lambda;t) = E(\lambda;t)/E_{clr}(\lambda;t) \tag{1}$$

because seasonal and diurnal effect are now removed. We have elsewhere described the self-consistent classification of sky conditions based on this clear-sky index as part of the decomposition algorithm for deriving the broadband direct normal (DNI) and diffuse horizontal (DHI) irradiances from the spectral GHI.[15] Seven distinct sky conditions are then identified via clear-sky index ranges for measurement channels 1 and 9, as noted in Table 1. Figure 2 presents the broadband GHI observed on 1 June 2021, color coded for sky condition; the smooth black curve is the GHI that would be observed under equivalent clear sky conditions. Note that the clear-sky index is $k^*(t) = \mathrm{GHI}(t)/\mathrm{GHI}_{clr}(t)$, which is the broadband version of equation (1). Our focus, however, will be on the spectral clear-sky index.

Table 1. Classification of sky conditions based on channels 1 and 9.[15]

Sky condition	**$\kappa^*(\lambda_1)$**		**$\kappa^*(\lambda_9)$**	
	Min	*Max*	*Min*	*Max*
Lensing	–	–	1.05	–
Very clear	1.0	–	0.75	1.05
Clear	0.8	1.0	0.75	1.05
Hazy	–	0.8	0.75	1.05
Thin clouds			0.5	0.75
Thick clouds			0.25	0.5
Overcast			–	0.25

3.2 Distribution of sky conditions

The probability distributions for all sky conditions were determined by kernel density estimation (KDE); the data of channel 9 were used as this channel is the primary metric for sky classification. The results are shown in Figure 3 as a plot of probability densities versus clear-sky indices. The black curve shows the density for the sum total of all sky conditions, while the colored curves show those for the individual sky conditions. This enables us to visualize the

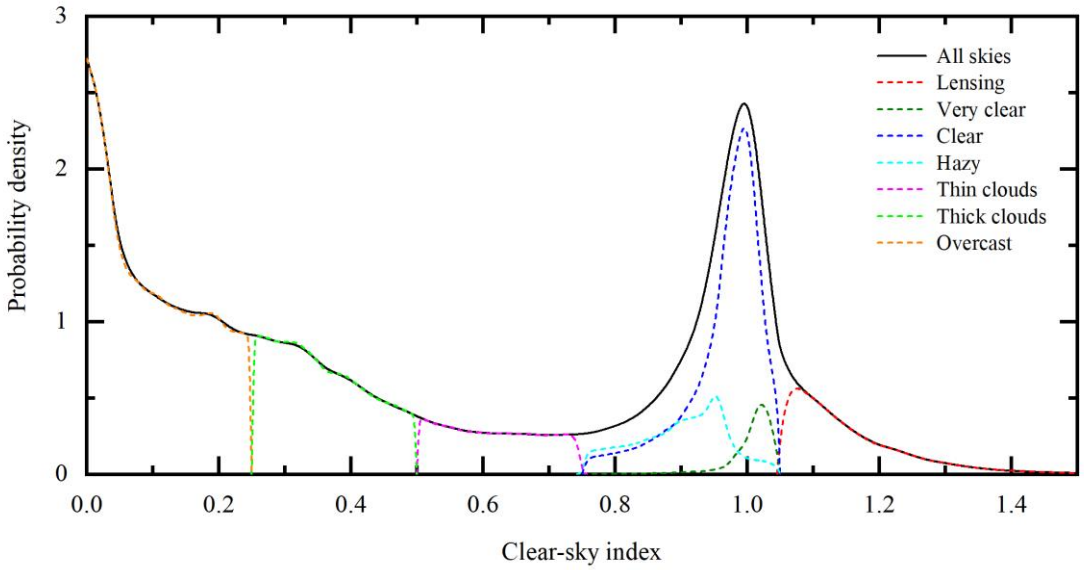

Figure 3. Statistical dependence of sky condition on spectral (channel 9) clear-sky index; the black curve gives the probability density for all sky conditions while the colored curves show those for the individual sky conditions, as labeled.

relative time spent in each cloud state. As described in Table 1, the thin, thick, and overcast distributions are independent of one another, while hazy, clear and very clear distributions overlap but with distinctly different dependencies on spectral clear-sky index.

3.3 Irradiance variability

We quantify irradiance variability using increments in the spectral clear sky index as

$$\Delta\kappa_\tau^*(\lambda;t) = \kappa^*(\lambda;t+\tau) - \kappa^*(\lambda;t), \tag{2}$$

where τ refers to the increment time step. The probability distribution of these increments describes the likelihood that sky conditions will persist for the given increment. Thus, for any increment, the probability distribution peaks in the limit as $\Delta\kappa_\tau^* \to 0$; that is, relatively small changes in the clear-sky index have the highest probability, with greater changes tailing off with decreasing probability. Figure 4 shows the dependence of the probability density, determined via kernel density estimates, on index increment for the different measurement channels; Figure 4(a) is for an increment time step of 600 s, while Figure 4(b) is for an increment time step of 1 s; both figures span the same range of densities. For the long time step, which spans the entire range of variability, spectral sensitivity is observed, which is easiest seen at large index increments (note the log scale). However, for the short time step, the spectral sensitivity is about twice as strong, as the variation in peak heights reveal, although only small index increments can be probed. Collectively, these demonstrate the spectral dependence of the probability density, with the IR channel (channel 9 – purple curve) typically spanning the greatest irradiance variability at all time steps.

The probability density is known to narrow as the increment time step is reduced, with the tails becoming more pronounced as the time step is increased.[16] This behavior is illustrated using the channel 9 (IR) spectral clear-sky index for increment time steps ranging from 1 s to 600 s. Channel 9 is chosen so as to be consistent with the sky classification procedure, as it is the channel most sensitive to cloud absorption and scattering.[15] Figure 5(a) shows the monotonic increase of the peak density, and the concomitant narrowing of the distribution, as the time step is decreased. This behavior is quantified in Figure 5(b), where the peak densities and their corresponding full widths at half maximum (FWHM) are plotted as a function of the increment time step; the former is shown on the right axis (blue squares) as a log-log plot, while the latter is shown on the left axis (red circles) as a lin-log plot.

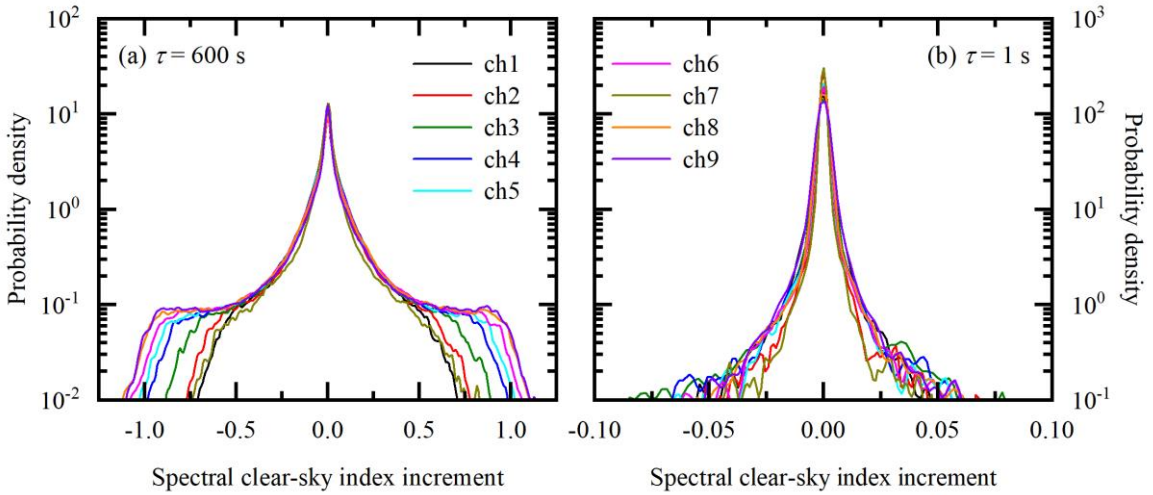

Figure 4. Channel dependence of the probability densities vs spectral clear-sky index increments for (a) a 600-s time step, and (b) a 1-s time step.

The kernel density estimation used to accurately describe the probability densities is a nonparametric method. Given the fat-tailed behavior of these distributions, the FWHM is therefore not expected to provide a definitive signature of the variability.[16] The peak density, however, is well-defined and is a highly sensitive metric. The peak density in Figure 5(b) is seen to increase monotonically with decreasing time step, but clear evidence of power law scaling is shown across three distinct temporal regimes within the ~30-minute variability period studied. Atmospheric variability is known to be dominated by scaling processes[17], and power law scaling has been shown to describe cumulus cloud size distribution.[18] The slope is therefore characteristic of the underlying variability. The short-time high-slope regime is very likely a sign of ramping behavior (i.e., when the sun's disk becomes occluded). A distinct transition happens at ~7 s to a state with reduced variability, which may indicate evolving intra-cloud dynamics. This is followed by a more abrupt transition at about 3 minutes to an intermediate variability state (with a slope between that of the other two), which may be indicative of a long-time regime dominated by transitions from one cloud type to another. The interruption in the monotonic increase in FWHM with increasing time step at this latter transition is consistent with this interpretation.

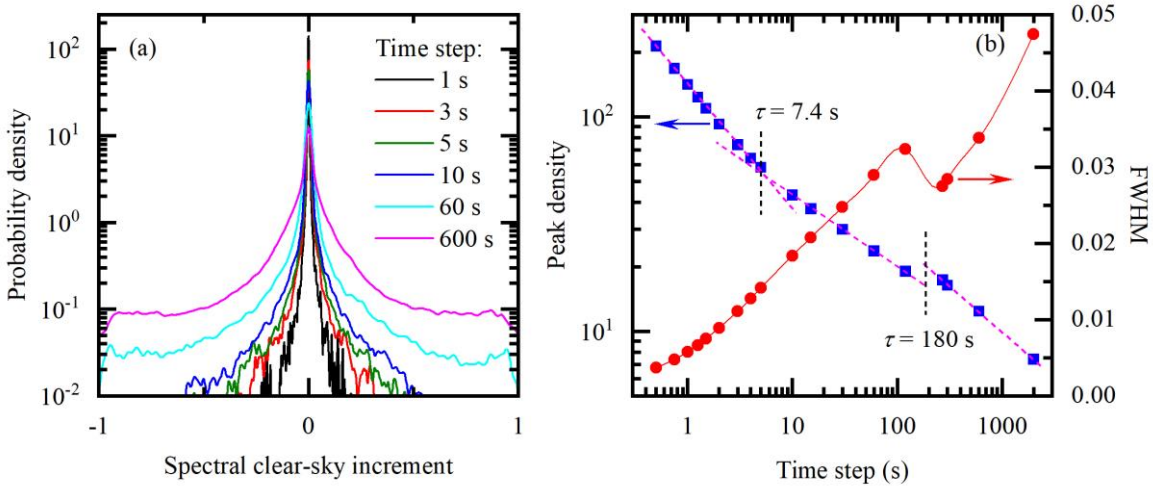

Figure 5. (a) Probability densities vs channel 9 spectral clear-sky index increments for the indicated increment time steps; and (b) their corresponding peak densities and full widths at half maximum (FWHM).

4. CONCLUSIONS

A novel spectral pyranometer was used in the creation of a spectral irradiance database for Ottawa, Canada. A preliminary analysis of the first six months of data, beginning 1 June 2021, from this software-augmented multi-filter radiometer allowed the cloud dynamics to be probed across time scales ranging from the sub-second to ~30 minutes. Sky conditions were self-consistently determined for each time interval without sky imaging. Exploiting the pyranometer's IR measurement channel, the statistical dependence on spectral clear-sky index of these seven distinct sky conditions were established via kernel density estimates of their probability densities. The irradiance variability was quantified using increments in the spectral clear-sky index. The resulting probability densities for a range of time steps spanning sub-second to ~30 minutes were quantified by their peak positions and FWHM. Power law scaling was found across three distinct temporal regimes within this range. They were speculatively associated with ramping behavior, intra-cloud dynamics, and transitions between cloud types. Since the identification of these time scales has implications for the nowcasting of distributed PV resources, further investigation is clearly warranted.

REFERENCES

[1] Gielen, D., Boshell, F., Saygin, D., Bazilian, M.D., Wagner, N., and Gorinia, R., "The role of renewable energy in the global energy transformation," Energy Strategy Reviews 24, 38–50 (2019)

[2] Bogdanov, D., Ram, M., Aghahosseini, A., Gulagi, A., Solomon Oyewo, A., Child, M., Caldera, U., Sadovskaia, K., Farfan, J., De Souza Noel Simas Barbosa, L., Fasihi, M., Khalili, S., Traber, T., and Breyer, C., "Low-cost renewable electricity as the key driver of the global energy transition towards sustainability," Energy 227 (2021)

[3] NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 3.0, NIST Special Publication 1108r3, National Institute of Standards and Technology (2014)

[4] Perez, R., Perez, M., Schlemmer, J., Dise, J., Hoff, T.E., Swierc, A., Keelin, P., Pierro, M., and Cornaro, C., "From firm solar power forecasts to firm solar power generation an effective path to ultra-high renewable penetration a New York case study," Energies, 13(17):4489, (2020)

[5] IEEE Standards Coordinating Committee 21 on Fuel Cells, Photovoltaics, Dispersed Generation, and Energy Storage, IEEE SStd 1547-2018 - IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces, IEEE (2018)

[6] Bachoumis, A., Andriopoulos, N., Plakas, K., Magklaras, A., Alefragis, P., Goulas, G., Birbas, A., and Papalexopoulos, A., "Cloud-edge interoperability for demand response-enabled fast frequency response service provision," IEEE Transactions on Cloud Computing (early access) (2021)

[7] Lezama, F., Soares, J., Hernandez-Leal, P., Kaisers, M., Pinto, T., and Vale, Z., "Local energy markets: paving the path toward fully transactive energy systems," IEEE Transactions on Power Systems, vol. 34, no. 5, pp. 4081-4088 (2019)

[8] Ahmed, R., Sreeram, V., Mishra, Y., and Arif, M.D., "A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization," Renewable and Sustainable Energy Reviews 124 (2020)

[9] Yang, B., Zhu, T., Cao, P., Guo, Z., Zeng, C., Li, D., Chen, Y., Ye, H., Shao, R., Shu, H., and Yu, T., "Classification and summarization of solar irradiance and power forecasting methods: a thorough review," CSEE Journal of Power and Energy Systems (Early Access) (2021)

[10] Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., and Hyndman, R.J., "Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond," Int J Forecasting 32, 896–913 (2016)

[11] Yang, D., "A guideline to solar forecasting research practice: reproducible, operational, probabilistic or physically-based, ensemble, and skill (ROPES)," J Renew Sustain Ener 11, 022701 (2019)

[12] Tatsiankou, V., Hinzer, K., Schriemer, H., and Beal, R., "Efficient, real-time global spectral and broadband irradiance acquisition," 7th World Conference on Photovoltaic Energy Conversion (2018)

[13] Yordanov, G. H., Saetre, T. O., and Midtgård, O., "Optimal temporal resolution for detailed studies of cloud-enhanced sunlight (overirradiance)," 2013 IEEE 39th Photovoltaic Specialists Conference (PVSC), 0985-0988 (2013)

[14] Tatsiankou, V., Hinzer, K., Schriemer, H., Emery, K., and Beal, R., "Design principles and field performance of a solar spectral irradiance meter," Journal of Solar Energy, vol. 133, 94-102 (2016)

[15] Tatsiankou, V., Hinzer, K., Schriemer, H., and Beal, R., "Improved global irradiance decomposition by sky condition classification from measured spectral clearness indices," 47th IEEE Photovoltaic Specialists Conference (PVSC) (2020)
[16] Lohmann, G.M., "Irradiance variability quantification and small-scale averaging in space and time: a short review," Atmosphere 9, 264-285 (2018)
[17] Lovejoy, S., "Spectra, intermittency, and extremes of weather, macroweather and climate," Sci Rep 8, 12697 (2018)
[18] Neggers, R. A. J., Griewank, P. J., and Heus, T, "Power-law scaling in the internal variability of cumulus cloud size distributions due to subsampling and spatial organization," J Atmos Sci 76, 1489-1503 (2019)

References

[1] IPCC, *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, H. O. Pörtner, D. C. Roberts, M. Tignor, *et al.*, Eds. 2022, ISBN: 9781009325844. DOI: 10.1017/9781009325844.

[2] IPCC, *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Summary for Policymakers), P. R. Shukla, J. Skea, A. Reisinger, *et al.*, Eds. 2022, ISBN: 9789291691609. DOI: 10.1017/9781009157926.

[3] A. Månsson, "Energy, conflict and war: Towards a conceptual framework," *Energy Research and Social Science*, vol. 4, pp. 106–116, C Dec. 2014, ISSN: 22146296. DOI: 10.1016/j.erss.2014.10.004.

[4] IEA, "World energy outlook 2022," International Energy Agency, Nov. 2022. [Online]. Available: https://www.iea.org/reports/world-energy-outlook-2022.

[5] E. Ela, V. Diakov, E. Ibanez, and M. Heaney, "Impacts of variability and uncertainty in solar photovoltaic generation at multiple timescales," National Renewable Energy Laboratory, 2013. DOI: 10.2172/1081387. [Online]. Available: http://www.osti.gov/bridge.

[6] P. Denholm, T. Mai, R. W. Kenyon, B. Kroposki, and M. O'malley, "Inertia and the power grid: A guide without the spin," National Renewable Energy Laboratory, May 2020. [Online]. Available: https://www.osti.gov/biblio/1659891.

[7] D. Yang, W. Wang, C. A. Gueymard, *et al.*, "A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality," *Renewable and Sustainable Energy Reviews*, vol. 161, Jun. 2022, ISSN: 18790690. DOI: 10.1016/j.rser.2022.112348.

[8] W. Liu, Y. Liu, X. Zhou, *et al.*, "Use of physics to improve solar forecast: Physics-informed persistence models for simultaneously forecasting ghi, dni, and dhi," *Solar Energy*, vol. 215, pp. 252–265, Feb. 2021, ISSN: 0038092X. DOI: 10.1016/j.solener.2020.12.045.

[9] D. Yang, "A guideline to solar forecasting research practice: Reproducible, operational, probabilistic or physically-based, ensemble, and skill (ropes)," *Journal of Renewable and Sustainable Energy*, vol. 11, 2 Mar. 2019, ISSN: 19417012. DOI: 10.1063/1.5087462.

[10] G. Kopp and J. L. Lean, "A new, lower value of total solar irradiance: Evidence and climate significance," *Geophysical Research Letters*, vol. 38, 2011. DOI: 10.1029/2010GL045777.

[11] N. Anderson, V. Tatsiankou, K. Hinzer, R. Beal, and H. Schriemer, "Probabilistic description of short-term cloud dynamics from rapid sampling of the solar spectral irradiance," *Proc. SPIE 11996, Physics, Simulation, and Photonic Engineering of Photovoltaic Devices XI*, Mar. 2022. DOI: 10.1117/12.2616231.

[12] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby, "The national solar radiation data base (nsrdb)," *Renewable and Sustainable Energy Reviews*, vol. 89, pp. 51–60, Jun. 2018, ISSN: 18790690. DOI: 10.1016/j.rser.2018.03.003.

[13] M. Šúri, T. A. Huld, and E. D. Dunlop, "Pv-gis: A web-based solar radiation database for the calculation of pv potential in europe," *International Journal of Sustainable Energy*, vol. 24, pp. 55–67, 2 Jun. 2005, ISSN: 1478646X. DOI: 10.1080/14786450512331329556.

[14] V. Tatsiankou, K. Hinzer, H. Schriemer, and R. Beal, "Improved global irradiance decomposition by sky condition classification from measured spectral clearness indices," in *47th IEEE Photovoltaic Specialists Conference (PVSC)*, 2020, pp. 72–76. DOI: 10.1109/PVSC45281.2020.9300629.

[15] G. M. Lohmann and A. H. Monahan, "Effects of temporal averaging on short-term irradiance variability under mixed sky conditions," *Atmospheric Measurement Techniques*, vol. 11, pp. 3131–3144, 5 2018. DOI: 10.5194/amt-11-3131-2018.

[16] Y.-C. Chen, "A tutorial on kernel density estimation and recent advances," *Biostatistics & Epidemiology*, vol. 1, no. 1, pp. 161–187, 2017. DOI: 10.1080/24709360.2017.1396742.

[17] B. W. Silverman, "Density estimation for statistics and data analysis," *Monographs on Statistics and Applied Probability*, 1986.

[18] B. K. Das and F. Zaman, "Performance analysis of a pv/diesel hybrid system for a remote area in bangladesh: Effects of dispatch strategies, batteries, and generator selection," *Energy*, vol. 169, pp. 263–276, 2019, ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2018.12.014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544218323806.

[19] L. Micheli, M. Theristis, D. L. Talavera, *et al.*, "The economic value of photovoltaic performance loss mitigation in electricity spot markets," *Renewable Energy*, vol. 199, pp. 486–497, Nov. 2022, ISSN: 18790682. DOI: 10.1016/j.renene.2022.08.149.

[20] J. Lin, M. Pipattanasomporn, and S. Rahman, "Comparative analysis of auction mechanisms and bidding strategies for p2p solar transactive energy markets," *Applied Energy*, vol. 255, Dec. 2019, ISSN: 03062619. DOI: 10.1016/j.apenergy.2019.113687.

[21] S. Karimi-Arpanahi, S. A. Pourmousavi, and N. Mahdavi, "Quantifying the predictability of renewable energy data for improving power systems decision-making," *Patterns*, vol. 4, 4 Apr. 2023, ISSN: 26663899. DOI: 10.1016/j.patter.2023.100708.

[22] H. Verbois, Y. M. Saint-Drenan, A. Thiery, and P. Blanc, "Statistical learning for nwp post-processing: A benchmark for solar irradiance forecasting," *Solar Energy*, vol. 238, pp. 132–149, May 2022, ISSN: 0038092X. DOI: 10.1016/j.solener.2022.03.017.

[23] G. Zhang, D. Yang, G. Galanis, and E. Androulakis, "Solar forecasting with hourly updated numerical weather prediction," *Renewable and Sustainable Energy Reviews*, vol. 154, Feb. 2022, ISSN: 18790690. DOI: 10.1016/j.rser.2021.111768.

[24] P. Lauret, H. M. Diagne, and M. David, "A neural network post-processing approach to improving nwp solar radiation forecasts," *Energy Procedia*, vol. 57, pp. 1044–1052, 2014. DOI: 10.1016/j.egypro.2014.10.089Ãŕ. [Online]. Available: www.sciencedirect.com.

[25] L. M. Aguiar, B. Pereira, P. Lauret, F. Díaz, and M. David, "Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting," *Renewable Energy*, vol. 97, pp. 599–610, Nov. 2016, ISSN: 18790682. DOI: 10.1016/j.renene.2016.06.018.

[26] P. Blanc, J. Remund, and L. Vallance, "Short-term solar power forecasting based on satellite images," in Elsevier Inc., Jun. 2017, pp. 179–198, ISBN: 9780081005057. DOI: 10.1016/B978-0-08-100504-0.00006-8.

[27] S. D. Miller, M. A. Rogers, J. M. Haynes, M. Sengupta, and A. K. Heidinger, "Short-term solar irradiance forecasting via satellite/model coupling," *Solar Energy*, vol. 168, pp. 102–117, Jul. 2018, ISSN: 0038092X. DOI: 10.1016/j.solener.2017.11.049.

[28] W. C. Skamarock, J. B. Klemp, J. Dudhia, *et al.*, "A description of the advanced research wrf version 2," University Corporation for Atmospheric Research, 2005. DOI: 10.5065/D6DZ069T.

[29] F. Grazzini and A. Persson, "User guide to ecmwf forecast products," *Meteorological Bulletin*, vol. 3, 153 pp. Jan. 2007. [Online]. Available: https://www.researchgate.net/publication/255267254.

[30] A. Al-Lahham, O. Theeb, K. Elalem, T. A. Alshawi, and S. A. Alshebeili, "Sky imager-based forecast of solar irradiance using machine learning," *Electronics (Switzerland)*, vol. 9, pp. 1–14, 10 Oct. 2020, ISSN: 20799292. DOI: 10.3390/electronics9101700.

[31] C. W. Chow, B. Urquhart, M. Lave, *et al.*, "Intra-hour forecasting with a total sky imager at the uc san diego solar energy testbed," *Solar Energy*, vol. 85, pp. 2881–2893, 11 Nov. 2011, ISSN: 0038092X. DOI: 10.1016/j.solener.2011.08.025.

[32] A. Al-Lahham, O. Theeb, K. Elalem, T. A. Alshawi, and S. A. Alshebeili, "Sky imager-based forecast of solar irradiance using machine learning," *Electronics (Switzerland)*, vol. 9, pp. 1–14, 10 Oct. 2020, ISSN: 20799292. DOI: 10.3390/electronics9101700.

[33] R. Chauvin, J. Nou, S. Thil, and S. Grieu, "Cloud motion estimation using a sky imager," in *AIP Conference Proceedings*, vol. 1734, American Institute of Physics Inc., May 2016, ISBN: 9780735413863. DOI: 10.1063/1.4949235.

[34] S. Dev, F. M. Savoy, Y. H. Lee, and S. Winkler, "Estimating solar irradiance using sky imagers," *Atmospheric Measurement Techniques Discussions*, 2019. DOI: 10.5194/amt-2019-141. [Online]. Available: https://doi.org/10.5194/amt-2019-141.

[35] D. Yang, S. Alessandrini, J. Antonanzas, *et al.*, "Verification of deterministic solar forecasts," *Solar Energy*, vol. 210, pp. 20–37, Nov. 2020, ISSN: 0038092X. DOI: 10.1016/j.solener.2020.04.019.

[36] D. Yang, G. M. Yagli, and D. Srinivasan, "Sub-minute probabilistic solar forecasting for real-time stochastic simulations," *Renewable and Sustainable Energy Reviews*, vol. 153, Jan. 2022, ISSN: 18790690. DOI: 10.1016/j.rser.2021.111736.

[37] P. G. Kosmopoulos, S. Kazadzis, M. Taylor, *et al.*, "Estimation of the solar energy potential in greece using satellite and ground-based observations," in *Perspectives on Atmospheric Sciences*, 2017, pp. 1149–1156. DOI: 10.1007/978-3-319-35095-0_165.

[38] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 65–76, 2013, ISSN: 13640321. DOI: 10.1016/j.rser.2013.06.042.

[39] P. Kumari and D. Toshniwal, "Deep learning models for solar irradiance forecasting: A comprehensive review," *Journal of Cleaner Production*, vol. 318, Oct. 2021, ISSN: 09596526. DOI: 10.1016/j.jclepro.2021.128566.

[40] A. Alzahrani, P. Shamsi, C. Dagli, and M. Ferdowsi, "Solar irradiance forecasting using deep neural networks," in *Procedia Computer Science*, vol. 114, Elsevier B.V., 2017, pp. 304–313. DOI: 10.1016/j.procs.2017.09.045.

[41] G. M. Yagli, D. Yang, and D. Srinivasan, "Automatic hourly solar forecasting using machine learning models," *Renewable and Sustainable Energy Reviews*, vol. 105, pp. 487–498, May 2019, ISSN: 18790690. DOI: 10.1016/j.rser.2019.02.006.

[42] D. Yang and D. van der Meer, "Post-processing in solar forecasting: Ten overarching thinking tools," *Renewable and Sustainable Energy Reviews*, vol. 140, Apr. 2021, ISSN: 18790690. DOI: 10.1016/j.rser.2021.110735.

[43] A. D. Dongare, R. R. Kharde, and A. D. Kachare, "Introduction to artificial neural network," 2008, pp. 2277–3754.

[44] L. Medsker and L. Jain, *Recurrent Neural Networks: Design and Applications*. CRC Press, 2001.

[45] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent advances in recurrent neural networks," Dec. 2018. [Online]. Available: http://arxiv.org/abs/1801.01078.

[46] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: Lstm cells and network architectures," *Neural Computation*, vol. 31, pp. 1235–1270, 7 Jul. 2019, ISSN: 1530888X. DOI: 10.1162/neco_a_01199.

[47] P. Le and W. Zuidema, "Quantifying the vanishing gradient and long distance dependency problem in recursive neural networks and recursive lstms," Mar. 2016. [Online]. Available: http://arxiv.org/abs/1603.00423.

[48] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," Mar. 2016. DOI: 10.1145/2939672.2939785. [Online]. Available: http://arxiv.org/abs/1603.02754%20http://dx.doi.org/10.1145/2939672.2939785.

[49] C. D. Sutton, "Classification and regression trees, bagging, and boosting," *Handbook of Statistics*, vol. 24, pp. 303–329, 2005, ISSN: 01697161. DOI: 10.1016/S0169-7161(04)24011-1.

[50] T.-H. Lee, A. Ullah, and R. Wang, "Bootstrap aggregating and random forest," in *Macroeconomic Forecasting in the Era of Big Data*. 2019, vol. 52, pp. 389–429.

[51] R. E. Schapire, "The boosting approach to machine learning an overview," in *Nonlinear Estimation and Classification*, D. D. Denison, M. H. Hansen, C. C. Holmes, B. Mallick, and B. Yu, Eds. 2002, pp. 149–171. DOI: 10.1007/978-0-387-21579-2_9.

[52] L. Breiman, "Stacked regressions," *Machine Learning*, vol. 24, pp. 49–64, 1996. DOI: 10.1007/BF00117832.

[53] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in Neurorobotics*, vol. 7, DEC 2013, ISSN: 16625218. DOI: 10.3389/fnbot.2013.00021.

[54] J. Wu, "Introduction to convolutional neural networks," 2017.

[55] S. Huang, J. Tang, J. Dai, and Y. Wang, "Signal status recognition based on 1dcnn and its feature extraction mechanism analysis," *Sensors*, vol. 19, May 2019, ISSN: 14248220. DOI: 10.3390/s19092018.

[56] B. M. Pavlyshenko, "Forecasting of non-stationary sales time series using deep learning," May 2022. [Online]. Available: http://arxiv.org/abs/2205.11636.

[57] M. Almuammar and M. Fasli, "Deep learning for non-stationary multivariate time series forecasting," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 2097–2106, ISBN: 9781728108582. DOI: 10.1109/BigData47090.2019.9006192.

[58] Z. Zhu, G. Xu, Z. Zhang, Y. Jiang, and M. Liu, "Very short-term solar irradiance forecasting at a sub-minute scale based on wt-cnns," in *Journal of Physics: Conference Series*, vol. 1659, IOP Publishing Ltd, Oct. 2020. DOI: 10.1088/1742-6596/1659/1/012042.

[59] S. E. Haupt and B. Kosovic, "Big data and machine learning for applied weather forecasts: Forecasting solar power for utility operations," in *2015 IEEE Symposium Series on Computational Intelligence, SSCI 2015*, Institute of Electrical and Electronics Engineers Inc., 2015, pp. 496–501, ISBN: 9781479975600. DOI: 10.1109/SSCI.2015.79.

[60] X. Chang, W. Li, and A. Y. Zomaya, "A lightweight short-term photovoltaic power prediction for edge computing," *IEEE Transactions on Green Communications and Networking*, vol. 4, pp. 946–955, 4 Dec. 2020, ISSN: 24732400. DOI: 10.1109/TGCN.2020.2996234.