MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Evaluation of different techniques for solar irradiance nowcasting"

verfasst von / submitted by

Pascal Gfäller BA, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna, 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 921
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Informatik UG2002
Betreut von / Supervisor:	Univ.-Prof. Dipl.-Inform.Univ. Dr. Claudia Plant

Abstract

Photovoltaic power production is increasingly becoming one of the central pillars in the shift to an electrical future based on renewable power sources. The use of solar irradiance has great potential, as daily solar power provides magnitudes more power than required. A downside of solar irradiance as a power source is its volatility. The extent of solar irradiance at the surface of the earth is influenced by a variety of atmospheric phenomena, most commonly clouds. Short-term forecasts, also called nowcasts, are therefore relevant to maintain a stable equilibrium in electrical grids. To achieve nowcasts on a grid scope, forecasting of solar irradiance over large areas from satellite data can substitute forecasting of power output for individual sites. As a contribution to this field, a variety of spatiotemporal neural networks are considered in this thesis to provide nowcasts for a target area extending over the borders of Austria. The models were evaluated with respect to their different architectural patterns and designs, displaying varying benefits in accuracy and in their robustness to missing data. Based on these results, key aspects from the best performing models were implemented into a new combination model, IrradPhyDNet. The base performance on a dataset containing only valid ground-truth frames can although be misleading regarding their forecasting performance in real-world applications. To provide stable nowcasting in the context of flawed data, an evaluation of inherent robustness to missing data and mitigation strategies was performed. The models' robustness was therein improved with a novel technique: Timestep-dropout, the training of neural networks on data with probabilistically removed timesteps. This technique provided significant improvements to the intrinsic robustness to missing irradiance data in the input tensors. Through multiple evaluations against baseline methods, the neural networks could showcase more accurate results, with clean as well as with flawed data. Within these comparisons, the newly introduced combination model performed better than the respective source networks in multiple aspects, especially when combined with the training technique timestep-dropout.

Kurzfassung

Die Stromerzeugung aus Photovoltaik wird immer mehr zu einer der zentralen Säulen im Wechsel hin zu einer elektrischen Zukunft auf Basis erneuerbarer Energiequellen. Die Nutzung von solarer Energie hat großes Potenzial, da die tägliche Einstrahlung um Größenordnungen mehr Strompotenzial bereitstellt als benötigt wird. Ein Nachteil von Sonneneinstrahlung als Energiequelle ist ihre Unbeständigkeit, da die reale Menge an Energie auf der Erdoberfläche durch eine Vielzahl atmosphärischer Phänomene, primär Wolken, beeinflusst wird. Kurzfristige Prognosen, auch Nowcasts genannt, sind daher relevant, um ein stabiles Gleichgewicht in Stromnetzen aufrechtzuerhalten. Um Nowcasts im Ausmaß von Stromnetzen zu erzielen, kann die Vorhersage der Sonneneinstrahlung über große Gebiete aus Satellitendaten die Vorhersage der Stromproduktion für einzelne Standorte ersetzen. Als Beitrag zu diesem Gebiet werden in dieser Arbeit verschiedene spatiotemporale neuronale Netze betrachtet, um Nowcasts für ein Zielgebiet zu liefern, das sich über die Grenzen Österreichs erstreckt. Die Modelle wurden hinsichtlich ihrer unterschiedlichen architektonischen Muster und Designs bewertet, die Vorteile in Bezug auf Genauigkeit und Robustheit gegenüber fehlenden Daten aufwiesen. Basierend auf diesen Ergebnissen wurden Schlüsselaspekte der besten Modelle in ein neues Kombinationsmodell, IrradPhyDNet, implementiert. Da die Basisleistung auf einem Datensatz, der nur gültige Ground-Truth-Frames enthält, hinsichtlich ihrer Vorhersageleistung in realen Anwendungen irreführend sein kann, wurde die inhärente Robustheit gegenüber fehlenden Daten und Kompensationstechniken weiter evaluiert. Um die Robustheit der Modelle zu verbessern, wird hier eine neuartige Technik angewendet: Das Training von Modellen auf Daten mit probabilistisch entfernten Zeitschritten. Die hier als Timestep-Dropout bezeichnete Technik lieferte signifikante Verbesserungen der intrinsischen Robustheit der neuronalen Netze. Durch mehrere Auswertungen im Vergleich zu Baselinemethoden konnten die neuronalen Netze genauere Ergebnisse liefern, sowohl aus vollständigen als auch aus fehlerbehafteten Daten. Darin schnitt IrradPhyDNet in mehreren Aspekten dieser Auswertungen besser ab als die Quellnetzwerke, insbesondere in Kombination mit der Trainingstechnik Timestep-Dropout.

Contents

List of Tables

List of Figures

1. Introduction

Solar irradiance forecasting has historically not received the same attention as other meteorological weather prediction. This has recently started to change with the increasing focus on the use of solar irradiance as a source of energy. In a shift towards renewable electrical energy production, away from fossil fuels with all their climatological and environmental detriments, solar power is growing to be one of the essential solutions. Solar irradiance can be harnessed in multiple ways, as a source of heat or in direct conversion to electrical energy, via Photovoltaic (PV) panels. PV power production is highly correlated to solar irradiance [1, 2], as it is its extrinsic source of energy. To derive PV power output estimations for the near future, which is increasingly relevant with wider adoption to maintain a supply and consumption equilibrium in electrical grids, solar irradiance forecasting can be applied at either local level or over wide areas. Solar irradiance at the surface level is not as steady as other means of generating electrical power. It fluctuates not only over day-night cycles and seasonally based on the respective geographic solar angle, but is also influenced by a variety of atmospheric dynamics. Among these, some of the most influential are clouds and aerosols [3], the movement of clouds can be considered the "the primary uncertainty in current surface solar global horizontal irradiance (GHI) forecasts" [4]. While aerosols and water vapour are more relevant in direct normal irradiance (DNI) forecasting, they also remain of interest for GHI and PV power forecasting [5]. These influencing meteorological factors can be introduced separately into a model, in this study they will be incorporated via the use of a solar irradiance data product provided by the Satellite Application Facility on Land Surface Analysis (LSA-SAF).

Spatiotemporal forecasting methods provide the possibility of centralized forecasts for wide areas, while there also exist a variety of techniques that derive forecasts at a local level from stationary measurement sites. The general benefit of spatiotemporal techniques that utilize satellite derived data lies in the ability to provide forecasts without the necessity to maintain local solar irradiance measurement sites and incorporate a wider spatial context.

The principal goal of this thesis is the evaluation of spatiotemporal neural networks' performances in a specific solar irradiance forecasting configuration and thereby also to ascertain if certain architectural model characteristics are beneficial or detrimental to performance. Some of the models and techniques used here have originally been applied to other tasks like general next frame prediction and radar echo precipitation forecasting, comparative data is available in these domains, but no satellite derived solar irradiance nowcasting dataset has been applied to them.

Based on an initial comparison of architectures and their performance, a new spatiotemporal neural network, IrradPhyDNet, is introduced which achieved mostly higher or very similar accuracy on the solar irradiance dataset than the competitors. The architecture of IrradPhyDNet is composed of cells and patterns adapted from the other models in the comparison. The design was derived experimentally and is intended to incorporate features that allow for more robust but also accurate forecasting. To evaluate the model not only against the other neural networks, a specific machine learning method created for the task of solar irradiance nowcasting, an implementation of an Analog Search algorithm, was selected to, in addition to ground-based measurements and Persistence, show the extent of detail and accuracy in the forecasts.

The temporal predictive range of interest is nowcasting, which is often defined as being below or up to six hours of lead time. For this study the lead time was limited to three hours, to support a higher temporal resolution of 15 minutes. Since nowcasting is time-sensitive, but can deteriorate significantly when gaps in the real-time or near-real-time input data occur, the models were also tested on unsteady or flawed input data. Gaps in the data as well as minor faults can have a negative impact on models' predictive performance and have to be compensated for in a practical application. Here, considered as an alternative to input data interpolation, increasing model robustness to missing timesteps via training on imperfect data is attempted. The technique is applied to all models to estimate potential improvements that could be achieved with this technique in comparison to training on only clean data. The use of this training routine, referenced as timestep-dropout, provides a significant improvement on the models' robustness to missing temporal data, while also contributing to some models' general performance.

The achievements of this thesis include:

- A comparison of three established neural networks, with an optimized variant and a new neural network design against Persistence baseline and a specific purpose machine-learning method in solar irradiance nowcasting.
- The introduction of a new spatiotemporal neural network design, based on the evaluation of the selected neural networks.
- A study on robustness to missing data, extended by applying a training technique to improve the robustness of spatiotemporal neural networks to missing timesteps.

2. Related Work[1]

Forecasting of solar irradiance can be separated into two main methodological branches, physics- or statistics-based. In contrast to physics-based methods (i.e. numerical models) statistics-based methods involve data-driven methods that rely comparatively less on explicit in-depth understanding and integration of meteorological and climatological domain knowledge. As a special subtype of statistical as well as machine-learning methods, and the focus of this comparative study, deep learning models are discussed separately, due to their increasing relevance as accurate predictors of future solar irradiation. Learning from satellite data or images shares a common data-centric basis as video analysis and next frame prediction. By also focusing attention on established methods and techniques which were able to resolve similar predictive problems, be it from the domain of general next-frame prediction or the recently more prolific development of spatiotemporal deep learning models for precipitation forecasting, solar irradiance forecasting tasks have and could still further benefit.

2.1. Traditional solar irradiance forecasting methods

2.1.1. Statistical and Machine Learning models

Within the wider field of statistical and machine-learning approaches, the perspective is generally data-centric and thereby provides applicability of a variety of standard methods for specific weather forecasting tasks. Most regressor techniques can be used to forecast site-based solar irradiance, this includes moving-average and auto-regressive methods [6, 7], tree and random-forest regressors [8] and also extends to other linear and non-linear regressors. But as solar irradiance timeseries exhibit strong non-linearity [9], the usefulness of linear methods on raw irradiance data is limited. In a comparison of Autoregressive Integrated Moving Average (ARIMA) to neural network approaches, Hoyos-Gomez et al. [10] found benefits with shorter time-series, while ARIMA was still performing worse than deep learning methods in most cases, with similar results found in [11]. ARIMA also showed to be outperformed by a sophisticated naïve reference approach and Kalman-filtered Numerical Weather Prediction (NWP) models in stationary intra-day solar irradiance forecasting [12], while being comparatively simple and much more easy to compute. These techniques, on their own, are now more commonly only included as baselines for comparison. For the prediction from spatial and spatiotemporal data, specifically designed methods as well as more complex pipeline approaches, that employ

[1]Parts of this chapter are based on a literature survey that was written by me within the Master Seminar (053049, 2022S) at University of Vienna.

feature engineering and aggregation, are more relevant.

In the domain of more strictly statistics based or non-deep machine learning solar irradiance forecasting methods in a spatiotemporal setting, Optical Flow was used by a variety of approaches for the motion estimation of clouds from satellite images [13, 14] or sky images [15]. While the derivation of cloud movements with Optical Flow methods is a viable approach, it has in the recent past fallen short of results achievable with deep learning models (e.g.[4]).[2]For single site forecasts from spatiotemporal data, so to maintain the spatial flexibility and scope of the input area, other techniques have been introduced. One such approach, introduced by Ayet and Tandeo, is an Analog Search method [17].It is composed of a machine learning pipeline to generate solar irradiance nowcasts for a spatial site from an ensemble of analogs found via aggregate features derived from spatiotemporal satellite data[3] This analog based technique differs in the range of lead time to other related approaches: A prior analog ensemble method for solar power prediction was introduced by Alessandrini et al.[18] with lead times of up to 72h, which was also utilized in a multi-model approach [19]. Cervone et al. [20] later used an artificial neural network (ANN) and an analog ensemble on NWP forecasts to predict PV power output up to 72h ahead.

2.1.2. Numerical Weather Prediction

The prominent techniques and methods used in meteorological institutions are mostly centred on the use of complex NWP models. These models are based on the mathematical approximation and simulation of atmospheric compositions and dynamics. NWPs are computationally intensive, requiring more sophisticated computing clusters and computing time than neural networks, especially when comparing inference. Additionally, they are by requirement heavily reliant on deep domain knowledge, inbuilt as parametrizations into the models. In comparison to recent spatiotemporal neural networks, currently used NWPs often also show worse performance in nowcasting and short-term forecasting, but can be considered competitive to superior above a time horizon of 6 hours [3]. In direct comparison, neural networks of limited complexity can surpass NWPs in the nowcasting range, see for example Knol et al.'s comparison between a spatiotemporal neural network and the Weather and Research Forecast (WRF) NWP model [21]. The inherent latency, or "computational bottleneck" [22], also redacts from their potential use in the domain of nowcasting.

NWP methods should nevertheless not be considered obsolete by the limited success of neural networks in specific subdomains of weather prediction. In general, multi-feature forecasting deep learning methods are at this stage not a viable option, and through a primarily data-driven approach, only some models provide a degree of the physical

[2]Optical Flow is here not included as a direct reference method but as the conceptual basis for pseudoflow, a technique introduced in PFST-LSTM [16] (see Chapter 4.1.2).

[3]An implementation of this Analog Search method will serve as a comparative baseline for the accuracy of the neural networks in terms of single spatial site forecasting (see Chapter 6.3.1 and Supplement A.1.3).

consistency of NWPs [23]. Further, the incorporation of various feature products from NWP forecasts can thereby also be valuable, for example as DL model's input features for solar irradiance nowcasting [24, 25] and solar power output forecasting [26, 27].

2.2. Solar irradiance forecasting with neural networks

2.2.1. From Convolutional and Temporal to Convolutional-Temporal networks

For data-driven solar irradiance forecasts from time-series, a variety of deep learning architectures and design patterns are applicable: While linear cells processing each individual entry separately are an option, a more promising solution lies in Recurrent Neural Networks (RNN). The recurrence aspect of networks and the underlying recurrent cells lies in hidden memory states which keep track of relevant aspects in previous timesteps. Thereby, predictions are not only based on a single input timestep but on previous timesteps, or more directly the memory states of previous timesteps. Long Short-Term Memory (LSTM) and Gated Reccurrent Unit (GRU) cells, advanced developments based of the simple RNN, have refined recurrence with the capability to forget and update memory states. LSTM networks are for example used effectively in stationary forecast that rely on local historical timeseries data [28]. Abdel-Nasser et al. used LSTM models as part of the ensemble model Reliable Deep Forecasting Model (RDFM) [29], using the Choquet integral as the aggregator, it showed higher accuracy than ARIMA, non-ensemble LSTM networks and Persistence. Another variant to LSTM, the Bi-directional long short-term memory (BiLSTM), was also applied to forecasting solar irradiance on stationary temporal data [30, 31], and brings the benefit of learning memory states in both directions, forwards into the future and backwards into past timesteps. BiLSTM cells may therefore be able to learn additional features in general timeseries [32], but also have drawbacks in terms of training time and data requirement and don't seem to be used in spatiotemporal architectures discussed below.

Spatial data, such as satellite derived irradiance maps, sky images or other variants of spatial aggregate data, is more effectively learnable with Convolutional Neural Networks (CNN). Convolutional cells pass kernel filters over the input dimensions, thereby learning to aggregate relevant feature patterns within a subspace of the domain. CNN cell variants are primarily defined by the dimensionality of their filters. 1D-CNN cells have been used for solar power forecasting from timeseries [33],while 2D convolutional cells are primarily used for the aggregation of spatial information, for example to extract cloud shape features for site-based solar irradiance forecasting [34]. Besides the use of 2D-CNN cells for spatially distributed data, they have also been used for solar forecasting from temporal and other non-spatial data formatted as a 2D-grid. Timeseries data restructured into a matrix containing rows of days and columns of hours for local forecasting [35] or with the intent to enhance the use of inherent correlations within a grid composed of measurement sites and measured meteorological parameters [36]. While 3D-CNN cells are capable of learning directly from spatiotemporally distributed data, fully-convolutional networks are seldomly utilized and regularly perform worse than other models in comparisons [37].

The combination of both convolutional layers and LSTM layers for deriving forecasts from temporally distributed features was explored by Ghimire et al.[38] and showed merit over the use of only convolutional or recurrent layers. A 1D-CNN-LSTM was also applied by Wang et al. on a discrete wavelet decomposition of stationary solar irradiance data [11] with similar improvements over the separate convolutional and LSTM networks as well as over ARIMA.
These types of neural networks, consisting of convolutional and recurrent cells, are also applicable to spatiotemporal forecasting: Spatial features in a timeseries can be input as individual images into 2D-convolutional cells to learn spatial aggregate features for recurrent cells, that in turn process the output as a sequence. In this manner, a kind of hybrid network can be established, which separates the learning over the relevant dimensions to the respective adequate layers. This method of separate learning of spatial and temporal information from spatiotemporal data was also utilized for solar irradiance prediction from sky videos [39]. Similarly in the context of single timestep ahead prediction, the recurrent layer or sub-network may also be replaced with a fully-connected layer or sub-network, to output a vector or single scalar [40] of solar irradiance predictions.

In contrast to hybrid sequential learning of spatial and temporal feature dimensions, there is a comparatively novel alternative approach of learning the spatial feature elements along the timeseries. This approach showed to be beneficial over the hybrid separate learning approach [41] or the use of 3D-CNNs for a fully-convolutional network [37]. There are also counter examples [42, 43], where the hybrid model Eclipse with a dataset of sky images could achieve higher accuracy within very short-term forecasting than approaches with more complex and sophisticated cell designs. Thus, indicating that the specifics of the input data and predictive goal as well as the architectural design of a neural network beyond the use of specific cells has to be considered.

2.2.2. ConvLSTM and other advances in spatiotemporal prediction

Overcoming the limitations of hybrid spatial and temporal learning neural networks, Shi et al. combined the separate functional cell designs of 2D-convolution and recurrent LSTM in the ConvLSTM cell [44] and thereby initiated a recent advent of new spatiotemporal neural networks. The corresponding first use in the ConvLSTM network has a ConvLSTM-based Encoder-Decoder design, and was able to gain a substantial increase in accuracy over a fully-connected LSTM network and performed better than multiple versions of a then state-of-the-art Optical Flow algorithm [44]. The core idea behind the ConvLSTM lies in the application of 2D-convolutional filters in a recurrent LSTM cell, to apply spatial learning within temporal learning, thereby maintaining a spatiotemporal 3D-tensor within and between the layers of the ConvLSTM network. This idea was the basis for a host of developments in spatiotemporal deep learning which have recently also been increasingly adapted in solar irradiance nowcasting.

Spatiotemporal predictive neural networks have been used more and to greater effect for the forecasting of weather phenomena, consequentially leading to research in new architectures and layer designs. While solar irradiance forecasting has received some attention, the more prominent meteorological field of research is precipitation forecasting. Many spatiotemporal network designs focused on this field of meteorological forecasting primarily [16, 22, 37, 44, 45, 46] or commonly used radar echo maps as a benchmark dataset. There's however no indication that these models should be limited to such datasets, as the structure of the input data remains spatiotemporally distributed in principle. Thus, most models, potentially aside from some specialized techniques, could deliver similar predictive quality for the forecast of solar irradiance.

Correspondingly the use of convolutional-recurrent cells in solar irradiance forecasting has also increased, using a variety of data sources: A multilayer ConvLSTM networks has been applied to solar irradiance nowcasting from satellite derived data over the Korean peninsula [47], with improvements over two conventional machine-learning approaches. A model utilizing Convolutional Gated Recurrent Units (ConvGRU), first introduced in [48], was able to surpass Optical Flow utilizing data from the Meteosat Second Generation Cloud Physical Properties (MSG-CPP), for the short-term forecasting of solar irradiance over the area of the Netherlands [21]. A further successor to ConvLSTM, the Eidetic LSTM[4] [50] , was adapted by Yu et al. [51] to forecast cloud amount with images from the infrared channel of GEO-KOMPSAT-2A. The cloud forecasts improved the predictive accuracy for local PV power generation in relation to a Convolutional Self-Attention LSTM trained without cloud amount in 5-hour ahead forecasts. IrradianceNet[5] by Nielsen et al. [4] introduced a redesign of the original ConvLSTM's encoder-decoder network structure [44] into an autoencoder architecture for the nowcasting of cloud albedo for solar irradiance prediction over a large area of Europe.

In contrast, stationary data sources involving the use of pyranometry remain a comparatively impractical option for large area forecasting. But due to the indirect nature of other sources, measurements from pyranometers are valuable for validation [4, 47]. Another alternative locally gathered spatiotemporal data can be found in sky imagery, which corresponds to the inverse perspective of satellites, using an often multispectral, wide-angle camera or sensor pointed from a local site above. Le Guen and Thome [52] introduced a new physical disentangling spatiotemporal neural network (see also chapter 4.1.3) and applied it to local sky image nowcasting. Within a limited lead time of 5 minutes, their approach has been achieving higher accuracy than the ConvLSTM network [44] and a more recent adaptation to spatiotemporal LSTM, the Memory-In-Memory network [53].
A different noteworthy approach to the spatiotemporal use of stationary data was introduced with a spatiotemporal Graph Neural Network (st-GNN) [54] that instead of learning

[4]A spatiotemporal LSTM with improved memory mechanism, based on PredRNN [49].

[5]See also chapter 4.1.1.

from equidistant maps, takes individual stationary measurement sites as a connected spatial feature-graph. The use of multiple, relatively close ground-based solar irradiance measuring sites, or similarly solar power plants, is also possible with other approaches. By aggregating the individual sites into a 2D-grid by relative proximity using a ConvLSTM Encoder-Decoder network for nowcasting to the same grid [55] or by interpolating from the individual sites' measurements [56]).

Since there are still continuous improvements and no universal solution, there is no ultimate but a variety of models that can be considered state-of-the-art for specific tasks. While solar irradiance forecasting has a past and present in physics-based and statistical forecasting models, the currently most active development is in the field of deep learning. The continued data-oriented reuse and adaptation of patterns from other domains, like next-frame prediction, has enabled the creation of numerous, more competitive solar irradiance nowcasting models.

3. Datasets

For solar irradiance nowcasting a variety of sources of irradiance data is applicable, with satellite derived imagery or grid-data being a prime choice for wide area nowcasting. Meteosat and ESA maintain multiple satellites in geostationary orbit, from which large area sensoric data is harnessed for a variety of tasks. In this study a irradiance data product based on Meteosat Second Generation (MSG) satellite measurements is used and combined with associated temporal and topographic features.

3.1. Downward Surface Shortwave Flux (LSA-207)

The MSG Downwelling Surface Short-wave radiation Fluxes – Total and Diffuse (MDSSFTD, LSA-207) is a processed data product, which was provided by the Satellite Application Facility on Land Surface Analysis (LSA-SAF), part of the distributed European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Application Ground Segment, ©2020 EUMETSAT. The relevant feature of this product is the total Downwelling Surface Short-wave radiative Flux (DSSF) measured in W/m^2 of global irradiation. The wavelengths considered in MDSSFTD are in the range of 0.3 μm to 4.0 μm per time and surface unit [57] and thereby encompasses the wavelength range of GHI, which is defined as the "geometric sum of the direct and diffuse horizontal (short-wave) components" [5].

The spatial resolution of the dataset is 0.05 by 0.05 degree, corresponding to approximately 3.1 by 3.1 km at the 0-degree sub-satellite point, covering the whole 0-degree MSG disk. Individual map frames are provided at a frequency or temporal resolution of 15 minutes, [57]. This is the current minimum inter frame time for the MSG satellite SEVIRI (Spinning Enhanced Visible and InfraRed Imager), upon whose short-wave channels the MDSSFTD product is largely based. The temporal resolution is limited by the capture time of a specific line in the image along the latitudinal extent, thereby being a temporally slightly mismatched aggregate [57]. To provide more accurate irradiance maps, separate methods for clear-sky and cloudy conditions are used to estimate DSSF. Additional features used in these methods involve ozone content and water vapour data from the European Centre for Medium-Range Weather Forecast (ECMWF), solar and view azimuth and angle as well as digital elevation from the LSA-SAF systems, the Daily Land Surface Albedo (MDAL LSA-SAF), and a land/sea mask and cloud mask (from NWC-SAF). [57]

3.1.1. Pre-processing of satellite data

The complete dataset was cropped to cover the target area of Austria as well as some additional spatial context, extending over the border of Austria. The exact coordinates used for the boundaries of the dataset are shown in Table 3.1, which produced a cropped grid of 256 horizontal by 64 vertical cells.

North	49.20
South	46.05
East	19.49
West	6.69

Table 3.1.: Coordinate bounds of the target area

The bounds were also expanded to this degree to include additional Baseline Surface Radiation Network (BSRN) stations in Payerne, Switzerland and Budapest, Hungary, to complement data from the site at Sonnblick in Austria. These sites are included for ground-based validation (see chapter 3.1.2) as well as reference locations for a comparison of the models' performances in terms of single-site forecasting (see chapter 6.3). A sample irradiance map of the target area is displayed in Figure 3.1, to illustrate the covered area.

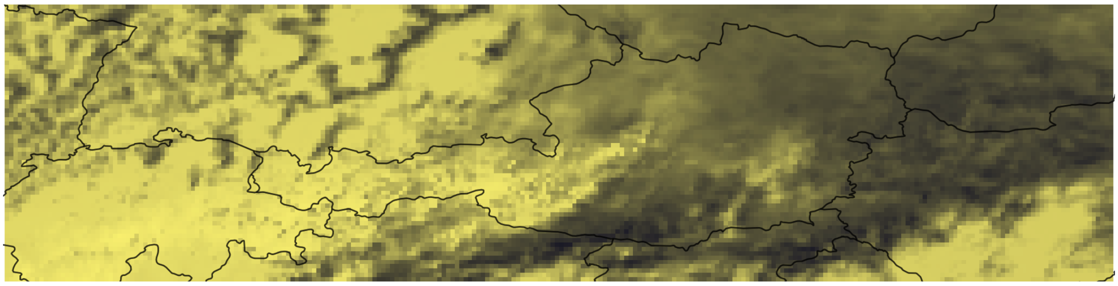


Figure 3.1.: Sample from the cropped irradiance dataset (11:00 23.05.2021)

The temporal extent of the included dataset is 2016 to 2021, providing a sufficient range of training data from the years 2016 to 2019, reserving 2020 for validation data and 2021 being kept aside as test data. The temporal resolution of the dataset is provided at 15 minutes over 24-hour days, and not altered to maintain the maximum possible temporal resolution, while limiting the lead times to achieve a temporal high-resolution configuration. The technical maximum of DSSF, is 1500 W/m^2 [57], a value way above the typically observed range of values in Central Europe. To avoid a compression of the results into a much shorter normalized numerical range, the dataset was analysed regarding the maximum value and upper quantiles for a more practical upper bound. The absolute maximum value over the whole temporal range of the dataset within the geographical crop is 1121.9, with a 99.9% quantile of 970.4. To avoid the mishandling of future outliers in the input data, which might exceed the observed maximum, a small buffer to the actual observed maximum of approximately 2% was introduced by shifting the upper normalization bound to 1150 W/m^2.

The LSA-207 dataset as loaded from the LSA-SAF servers showed some amount of missing timesteps and flaws in the data. For the whole time-range 0.29% of timesteps where missing and 1.36% were complete NaN(Not a Number)-arrays after unpacking and cropping, leading to a total of 1.65% of faulty timesteps. This highlights the possibility of gaps in the data as an issue to consider[1].

Individual missing timesteps and blocks of consecutive missing timesteps were interpolated along the temporal axis to maintain a continuous gapless dataset. Temporal interpolation led to a sufficient degree of realistic patterns in the timestep gaps, with seasonal spatial average day and night irradiance that resemble the raw uninterpolated data within the same time of day in their absolute value ranges and relative changes. Beneficial to the evaluation of the models is the factor of the distribution of flawed timesteps over the whole dataset: Analysis of the raw dataset showed a concentration of flaws within primarily the first year, 2016 with over 7% of timesteps missing or flawed, the last two years, which were selected for validation and testing, had only minor amounts of consecutive missing values, leading to an ideal base for more realistic individual training evaluation and termination and giving more validity to the final per-model evaluation results. As the effort is primarily comparative and the same dataset is shared over all models, minor deviations should have no effect on the relative results.

To maintain acceptable training times over all models, in this comparison a temporal stride of 11 timesteps was applied on the training dataset, thereby limiting the dataset to 12749 individual X/y-pairs of 12 timesteps each, while maintaining the full range of years with varying start and end time of day per sample. For the test dataset, to provide a more diverse range of samples, the stride was reduced to 5 timesteps, leading to 7004 individual sample-pairs.

3.1.2. Comparison to ground station data

More direct measurements of solar irradiance, specifically GHI, can be derived from ground-based stations. As these are not viably providable for an area of even a smaller country such as Austria in a similar spatial resolution as the MSG data, they are mainly of interest for validation and comparisons on specific sites within the target area. This approach is especially meaningful in the context of the satellite's indirect sensoric measurements, which can lead to certain deviations from the directly measured ground-station data. This potential error is enhanced by the averaging effect the square-kilometre large grid cell areas have, in contrast to a single point of reference within such a grid cell. Ground stations can provide a localized direct measure for solar irradiance, having thereby benefits of

[1]Which will be discussed further below in the context of robustness as a quality criterium.

accuracy as well as the potential for a much higher temporal resolution. Adversely, a grid of ground stations covering a section of an area a satellite can cover, does not exist. For validation as well as comparisons, the BSRN provides useful professional meteorological data from weather stations across the globe, which, although they are denser in Europe, still have a relatively high distance from one another. In LSA-SAF's internal product validation, LSA-207 was also compared to BSRN data. Table 3.2 shows the average Mean Bias Error and relative Mean Bias Error for different quality criteria ranges.

	MBE (DSSF<200W/m^2) [W/m^2]	rMBE (DSSF>=200W/m^2) [%]	MBE (fd<0.5) [-]	rMBE (fd>=0.5) [%]
Clear-sky case	8.6	0.78%	0.06	-22.20%
Cloudy-sky case	-6.6	2.78%	0.03	-15.80%
All-sky case (clear + cloud)	3.6	0.25%	-0.04	-17.70%
Accuracy Requirements	***20W/m^2***	***10%***	***0.1***	***20%***

Table 3.2.: MDSSFTD Validation Results Summary against BSRN in situ measurements, from [57], ©2020 EUMETSAT, Reprinted, with permission.

LSA-SAF's validation results were derived from a comparison against four BSRN stations' measurements over 8 months. These achieved quality criteria will here be supported by an additional comparison to BSRN sites within the target area and time, to provide context for the evaluation in Chapter 6.3. Within this study's target area are three BSRN stations: Payerne in Switzerland, Sonnblick in Austria and Budapest in Hungary. Details on the locations and local topography of the sites are outlined in Table 3.3.

BSRN station	Coordinates (lat-lon)	elevation	Topography
Budapest [58]	47.4291 - 19.1822	139.1m	flat, rural
Payerne [59]	46.8123 - 6.9422	491.0m	hilly, rural
Sonnblick [60]	47.054 - 12.9577	3108.9m	mountain top, rural

Table 3.3.: BSRN stations

For all three sites data of the test year 2021 was gathered and temporally averaged to 15 min resolution. Missing values were not considered and skipped, as well as values during night-time between 19:00 and 6:00. From the LSA-207 dataset the respective grid-areas for the coordinates of the BSRN sites were sampled for comparison. In Figure 3.2 the co-occurrence of measurements from satellite to ground stations is displayed. DSSF of LSA-SAF and GHI from the BSRN sites show an adequately high correlation, that further validates the use of the DSSF. The correlation is however not equal over all sites, which is presumably based to a high degree on the specific geographical characteristics of the weather station sites.

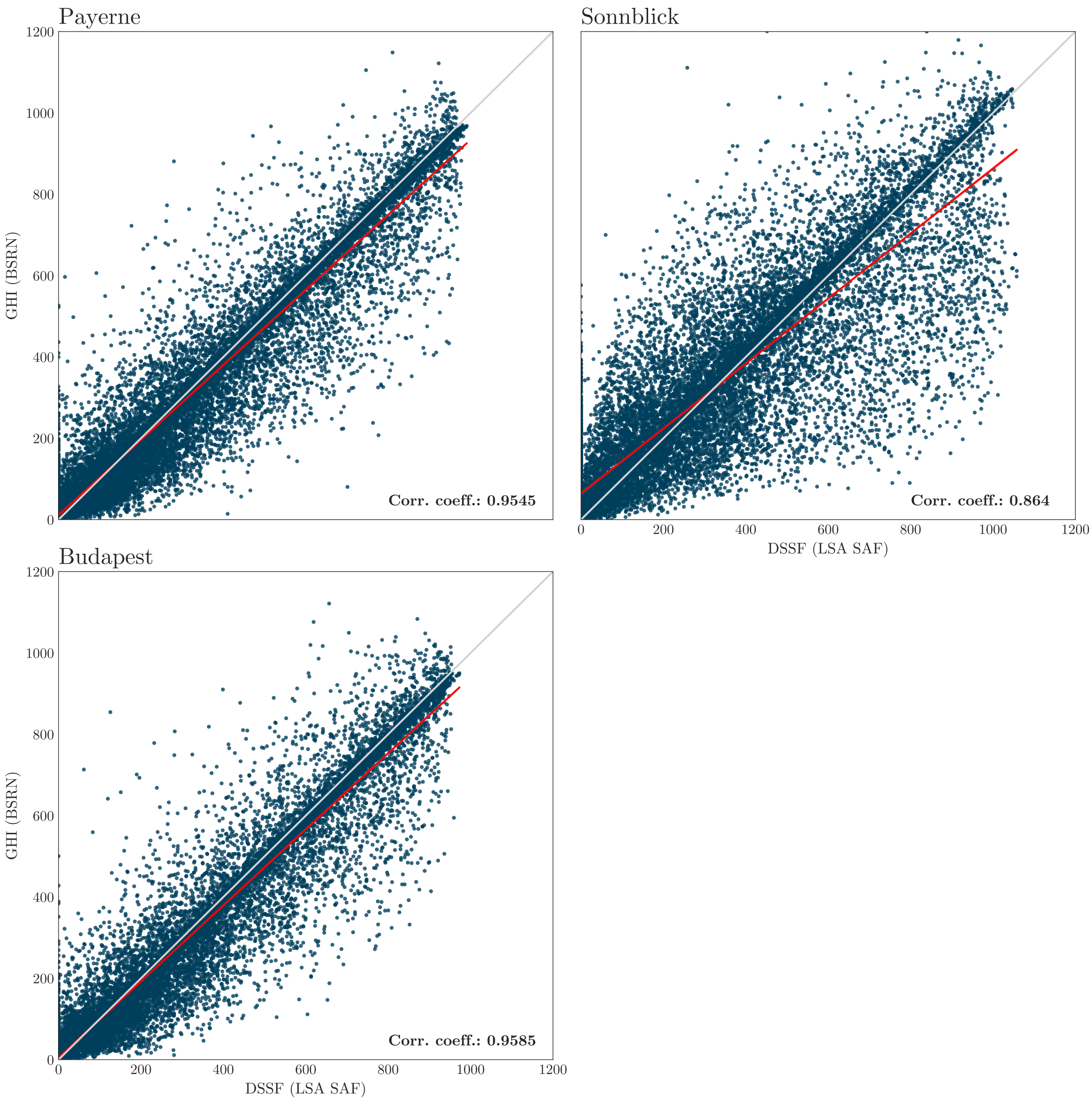


Figure 3.2.: BSRN-GHI - LSA-207-DSSF comparison across 3 BSRN stations 2021 (Pearson correlation)

Sonnblick shows a comparatively larger deviation in relation to the other two sites, which most likely is determined by Sonnblick's location atop Hoher Sonnblick, at an altitude of 3108.9 metres. Payerne's and Budapest's weather stations are in contrast situated at the respective local ground levels in vicinity to the town of Payerne and city of Budaptest in less extreme environments. Sonnblick's data is therefore expected to show a more unique perspective that is harder to match by the multiple square kilometres encompassing single grid cell MSG can attribute to the location and its surroundings.

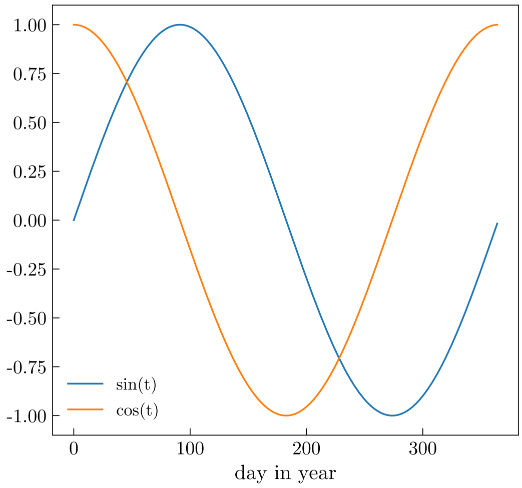
3.2. Temporal and spatial background information

3.2.1. Temporal data

To generalize seasonal and hourly differences of a limited dataset, additional context features may prove beneficial. The solar irradiance dynamics from the satellite image features, inherently containing cyclical changes, were enhanced by the inclusion of the timestamps at which they were taken. This is assumed to allow the models the linking of cyclically occurring patterns to relevant time-of-day and day-of-year features. In some other approaches the removal of seasonality in training is considered viable pre-processing, mostly for 1D-temporal data. Here seasonality will not be removed and later added to the predictions but learned by the models through the day-of-year features to implicitly understand the effects of seasonality. Temporal data is here considered especially relevant to achieve more accurate forecasts in the early hours of daylight, as the DSSF solar irradiance data contains no indicator of nightly cloud movements. Through temporal data, a relatively accurate approximation of seasonal early-day solar irradiance should be learnable.

There are multiple options to encode date and time features extracted from a timestamp. Expressed as minutes in a day and days in a year encoded as integer values, they could simply be normalised to floating point values, but this would ascertain a separate linearity within inherently cyclical time features. Encoding for example days in year into a normalized continuous feature variable would ascertain an ordering with an upper bound that is disconnected by the whole range of values from the next day, if it happens to be the first day of a new year. An approach to remove this effect is to dissolve numerical time features into separate categorical features with one-hot-encoding. Given 366 possible values for days in a year, their linear numerical increase could be removed by using a sparse representation of 366 buckets containing one true Boolean and 365 bucket marked as false. This wouldn't mislead the model through a continuity with an upper and lower bound, but as the representation is sparse with a scope of 366 values, or 1440 in the case of minutes per day, it is fairly inefficient and requires the model to relearn cyclical information that could be introduced to it explicitly in the input feature.
To achieve a more direct representation of the cyclical nature of the time features, a sine and a cosine transformation is used to convert an integer value into two variables expressing the cyclical nature and avoid distortion into linear values or hiding their cyclical nature by categorization. Two variables per one input time variable are required, as the usage of only cosine or sine functions would lead to a false equivalence of values mapped to the same height at the ascent and descent of the respective cyclical function curves.[2]

[2]https://towardsdatascience.com/cyclical-features-encoding-its-about-time-ce23581845ca



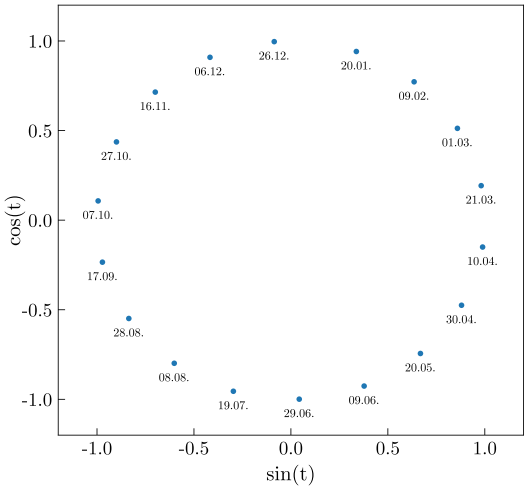


Figure 3.3.: Cosine and Sine curves for day in year feature mapping and sample mapping of dates to cosine and sine features.

Some approaches as for example MetNet2 [22] contain additional lead time conditioning, an inclusion of the timestep to be forecasted as an input feature. This can be a relevant inclusion if auto-regressive or single timestep forecasts are considered. As the approach here is focused on single-shot multi-step nowcasting, no additional temporal information is required.

3.2.2. Spatial data

Additional static spatial data has the potential of delivering relevant context that influences dynamics to models for temporal interactions in the spatial domain. This could improve the achievable forecasting accuracy, especially in the context of the Alpine regions of central Europe, that cover a large part of the target area. Mountain ranges impact solar irradiance, mostly by cloud forcing, through their influence on cloud dynamics [61]. Therefore, the inclusion of topographic and orographic data may prove beneficial as a secondary influential factor. This approach was here undertaken experimentally, while not being uncommon as parts of solar irradiance forecasting feature-sets [4] and precipitation forecasting [22, 45], which shares a common interest in cloud formation, dynamics and dissipation. Beyond the elevation map no other topographic and orographic features were included, as the irradiance datasets resolution limits the applicability of aspect or slope maps, as they would lose most of their relevance when averaged. Elevation maps from the Copernicus Land Monitoring Service's EU-DEM digital surface model[3] are provided in a 25mx25m resolution that far exceeds the requirements of this study, the elevation maps were resampled to 2,5x2,5km averages and then aligned by closest match to the irradiation data. The rescaling of the original data allowed the use of appropriate averages

[3]https://land.copernicus.eu/imagery-in-situ/eu-dem

for an area as well as a relatively close matches to the irradiance data grid coordinates, or respectively a distance of the centres of the individually assigned coordinate matches below 0.025 degrees.

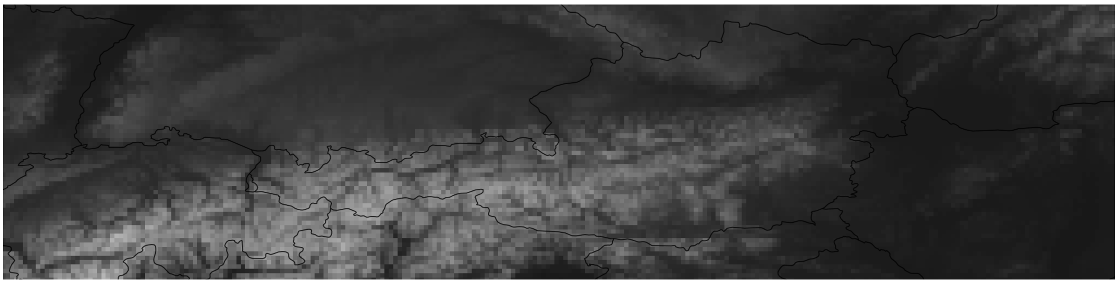


Figure 3.4.: Projection of elevation map to the target area and resolution

The chosen elevation map and pre-processing lead to a spatial feature with a temporally varying correlation to the irradiance data. Within the days of a year, a general trend with strong variation for individual days can be seen in Figure 3.5, while through the hours of a day a clearer connection is perceivable.

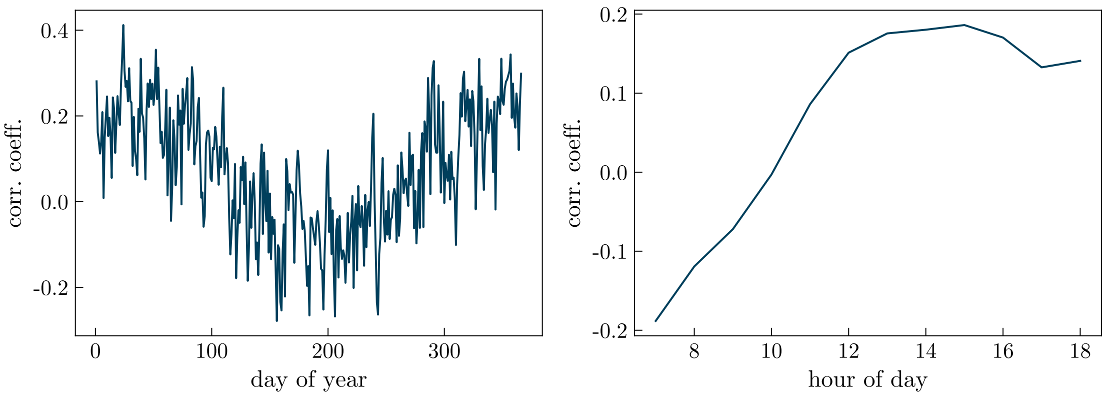


Figure 3.5.: Spearman correlation between elevation map and irradiance data, aggregate over irradiance dataset 2016-2021 (2.6% of results were not significant)

In both cases the relatively low average correlation coefficients cannot lead to the assumption of the elevation map, on its own, being a sufficient predictive feature, but in combination with the temporal features described above, a neural network should be able to derive meaningful patterns of information from this data.[4]

[4]A proof of relevancy was derived via a dataset ablation study (see supplement A.1.1), showing a beneficial effect of the inclusion of the elevation map, especially in combination with the temporal features.

4. Models considered in the comparison

The first group of models introduced in the comparison is a selection from the wider field of spatiotemporal predictive neural networks. All but one of the neural networks were not originally used for solar irradiance nowcasting from satellite derived data but have been used to perform predictive tasks on similar datasets, like radar echo maps and sky images. As the basic structure of predictive spatiotemporal neural networks is determined by the shape of the input and output data, some agnosticism to the specific selection of dataset can be assumed. Additionally, to this very general data-derived definition, the internal layout of cells or layers in spatiotemporal neural networks varies and thereby also the ability to learn and predict certain characteristics and patterns. Some models have specific components that involve approximations of other well-known physics or statistics based predictive methods. This is contrasted against the ConvLSTM based IrradianceNet, which doesn't include any specific adapted technique but follows a more generalist design.

In addition to these already published neural networks, a novel spatiotemporal neural network design, IrradPhydNet, will be introduced below. IrradPhyDNet is based on design elements found in the aforementioned neural network architectures and their respective cell designs and is the result of an experimental approach to recombining parts of the architectures estimated to have a beneficial impact on the neural networks' predictive performance.

All models were, where necessary, adjusted to the task applied to them. Parameter increases to facilitate the more complex data structure were attempted but lead to limited effects. For one model, PhyDNet, an alternative configuration implemented from information provided by the authors led to a significant increase in base performance, Changes to PhyDNet based on supplemental information led to a sufficiently different architecture, to include it as a separate model.

4.1. Published models

4.1.1. IrradianceNet - spatiotemporal autoencoder

IrradianceNet is a novel autoencoder variation [4] to the original encoder-decoder design of the 2015 ConvLSTM Network by Shi et al. [44]. The basic ConvLSTM cell design used in the autoencoder is kept the same as in the original introduction, utilizing a 2D convolution as a spatial learning operator within a recurrent LSTM-cell. The network structure is composed of a three-layer deep encoder and a three-layer deep decoder section. To achieve the necessary compression of the input feature-space, the encoder utilizes convolutional downscaling via a two-step wide stride to compress the input data along the spatial axes into informative latent space representations. Following, the inverse operations are applied on the compressed representation by the decoder-section of the network, which is spatially upsampled via transposed convolutional upscaling. In the compression and decompression stages, the channel size is inversely increased and decreased incrementally along the encoder and decoder layers.

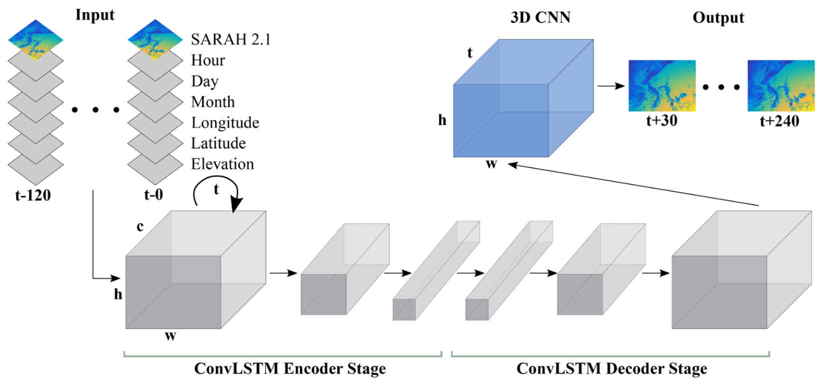


Figure 4.1.: Architecture of IrradianceNet, from [4], ©2021 Nielsen, et al., Reprinted, with permission.

The final 3-dimensional convolutional layer, presented in figure 4.1, is described in the paper as a means to convert the output from deterministic to probabilistic or limit it temporally to a single lead time[4]. For this thesis, the deterministic configuration was chosen, reducing the latent space to a single deterministic prediction along the spatial and temporal dimensions. While probabilistic predictions have merit in themselves, their accuracy often lags behind deterministic models, as Nielsen et al. also found [4].

IrradianceNet was introduced with the same goal of solar irradiance nowcasting but used a different approach in terms of dataset and input feature space: Nielsen et al. derived effective cloud albedo from the SARAH 2.1 historical dataset and included temporal information, longitude and latitude and an elevation map as input features in their spatiotemporal tensors to achieve cloud cover forecasts. Actual irradiance forecasts are then calculated in post-processing with clear sky irradiance data. The intended spatial scope was a grid of 512 by 512 while the typical input of the model is 128 by 128, for each spatial feature map, 8 timesteps that cover 4 hours in half hour increments, are input. To provide large-area forecasts, the original study of IrradianceNet used patch-based cropping to limit the input frames spatially and reduce concurrent memory requirements. This method can here be avoided due to the smaller area under consideration. One benefit of this decision lies in the avoidance of internal edges in the final predictive frame, that would arise if the full frame would be separated into patches which then were individually applied as input to the neural network[1]. Another consequence of quartering the maps

[1] Nielsen et al. discussed and mitigated this issue by linearly interpolating the edge areas between frames[4].

and applying frames in different geographical regions out of a larger frame to the same neural network would be longer training times through the quadrupled training dataset. Moreover this technique introduces the requirement of the model to learn relevant feature differences in geographical origin of the different samples, which was avoided for one out of lack of necessity, but also as other measures were taken to reduce overall training time and thereby facilitate the comparative effort.

Source code and modifications

From among the different trained variants in [4], IrradianceNet is here used as a deterministic single-shot multi-step ahead forecasting model, with symmetrical input and output length. As all other models, IrradianceNet was configured to handle non-square spatial data and for symmetrical single-shot multi-step prediction which can be derived directly from the last layer of the autoencoder. Within this comparison IrradianceNet is otherwise kept the same as published to the authors' GitHub repository [2]. As the model didn't saturate the available memory completely, parameter optimizations were attempted but showed no benefit over the base configuration.

4.1.2. PFST-LSTM – spatial memory and pseudo flow

Pseudo flow spatiotemporal LSTM (PFST-LSTM)[16] is a representative of more advanced developments in cell designs that followed research on spatiotemporal predictive networks after the introduction of ConvLSTM-cells by Shi et al.[44]: The PFST-LSTM cell shares the same notion of ConvLSTM, but includes two later additions to the original concept: Spatiotemporal memory, as introduced by Wang et al. [49], an additional memory state mapping the spatiotemporal state, and the newly introduced pseudoflow alignment module, a cell level module for the incorporation of spatial displacement between timesteps[16].
The typical implementations of spatiotemporal neural networks have no explicit elements for the capturing of changing spatial alignments over time as they are location-invariant and thereby match the same grid position over consecutive timesteps without having specific methods to calculate alignment of particles or elements in movement [16].
A previous attempt at solving issues with positional alignment of moving objects in a grid or frame is TrajGru [37], which is focused on capturing the motion but "does not have a special design for preserving spatial appearance details" [16] and thereby falls short in terms of maintained detail leading to comparatively blurrier results.

The idea behind the pseudoflow learning introduced by Luo et al., is related to and inspired by optical flow, a method for the estimation of intensity-invariant pixel-shift over consecutive frames, via the computation of optical flow fields. Optical flow as a methodological framework is applicable to the cloud motion estimation in solar irradiance nowcasting [62], and was also used as an input to general next frame prediction [63].

[2]ConvLSTM_small from GitHub - https://github.com/holmdk/IrradianceNet

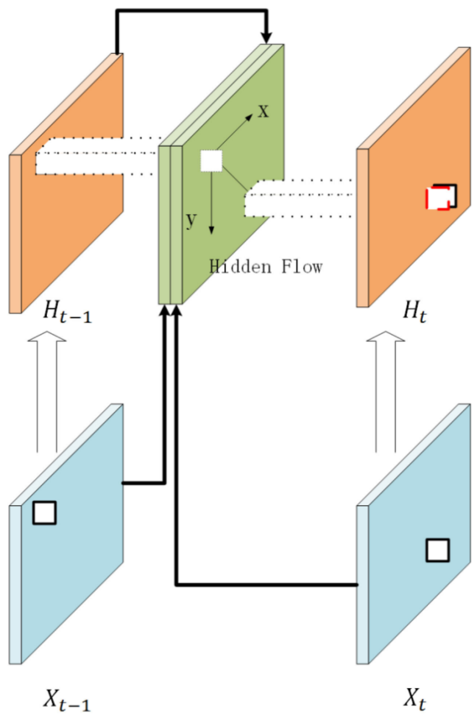


Figure 4.2.: Notion of Pseudoflow in PFST-LSTM, from [16], ©2021 Luo et al., Reprinted, with permission.

PFST-LSTM in contrast doesn't use optical flow but incorporates a pseudo-flow generation operator, composed of multiple single-layer-convolutional layers and a bilinear sampling warp function to approximate a displacement mapping or respectively movements within motion fields along the temporal axis. PFST-LSTM doesn't suffer from intensity invariance as optical flow does [16] and should therefore also be able to capture differing fields of irradiance intensity in dynamic overcast weather.

In figure4.3 the cell structure and flow of input, memory and hidden states through the PFST-LSTM cell is illustrated. PFST-LSTM uses previously established elements from ConvLSTM[44] (in white), the spatiotemporal memory from PredRNN[49] (in yellow) and places the pseudo-flow displacement generator (in blue) at the beginning to align memory and hidden states based on the spatial changes between the current and previous input. The PFST-LSTM cells are structured in a three layer deep recurrent encoder-decoder for sequence-to-sequence prediction, similar to the original ConvLSTM network design[44] and even more closely resembling PredRNN, introducing the spatiotemporal memory flow between the last layer at timestep t-1 and the first layer of timestep t[49].

The research interest in the original paper focused on radar echo maps (as was the

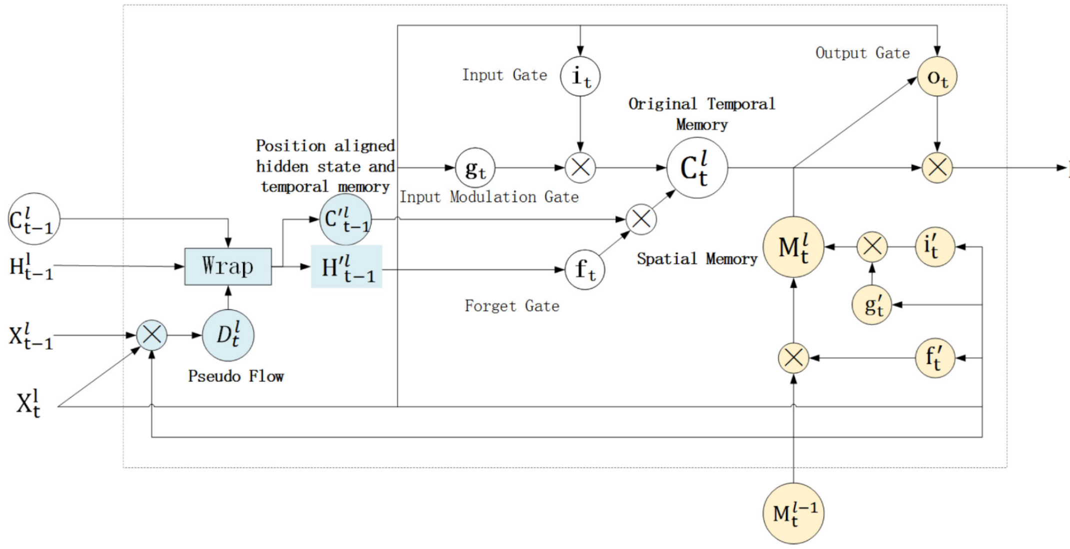


Figure 4.3.: PFST-LSTM cell design, from [16], ©2021 Luo et al., Reprinted, with permission.

original ConvLSTM networks primary purpose [44]), which differ from satellite data in terms of frequency of capture (radar echo maps, being more than double as frequently) and in terms of spatial extent and resolution. Radar echo is measuring particles in a much smaller area. A partial similarity to solar irradiance satellite data can be assumed in these particles to one of the negative influencing factor as clouds and other atmospheric partial opacities. This minor commonality may the basis for certain benefits when using PFST-LSTM on the solar irradiance dataset. But as the satellite images are typically much denser in information, as no constant background intensity is present most of the time, the benefits of this type of network probably won't show it's full effect.

Source code and modifications

The source code of PFST-LSTM is provided on GitHub [3]. The configuration provided was originally used on the precipitation forecasting task of the CIKM ANALYTICUP 2017 competition and found to be very lightweight, i.e. using less learnable parameters than the other models. While modifications to the models parameter count initially led to improvements on a smaller pre-test dataset, the difference on the final dataset was comparatively minor (<1% reduction of RMSE). Additionally, when confronting both variants with missing timesteps in the input data the CIKM configuration was slightly more robust and maintained lower error metric scores. Based on these findings the version

[3] https://github.com/luochuyao/PFST-LSTM

included in the final comparison was chosen to be the CIKM configuration, which also corresponds to the version outlined in the original paper.

4.1.3. PhyDNet – disentangling physical dynamics

A potentially relevant approach to improving the nowcasting performance of neural networks lies in the inclusion of neural differential equations as a means to approximate dynamics in satellite images over time.
While to the author's knowledge there's no differential equations known which can serve as a general approximator and predictor for solar irradiance, in other fields such as precipitation nowcasting, general differential equations where successfully utilized [64]. Out of the potentially useful approaches the most promising variant of employing the beneficial effects of partial differential equations was found in PhyDNet [65], a hybrid neural network including a residual ConvLSTM sub-network in addition to physics-informed PDE-based sub-network. The approach is based on the assumed beneficial effect of disentangling the learning of major dynamic movements and the changes in pixel-level details in spatiotemporal predictive tasks. The aforementioned lack of adequate analytical solutions to changes in solar irradiance over time in the form of partial differential equations is resolved in PhyDNet through the exclusion of predefined PDEs. Instead the model includes PhyCell, a layer architecture that's able to approximate potentially useful PDEs during the training process.
PhyDNet belongs in part to the class of physics-constrained and PDE learning neural networks but is not limited to appropriate physically determined features, through the introduction of its two-branch disentangling architecture and a cell design that allows for the approximation of a variety of PDEs. The spatiotemporal predictive network was originally intended for next frame prediction in generalist videos and was adapted by the authors to perform local solar irradiance nowcasting from fisheye lens equipped cameras in a later study [52], which serves as proof of viability within the context of this exploratory comparison.

The disentangling two-branches within the architecture differ by their cell designs and their respective purpose: One branch is dedicated to learning the physical dynamics in the input data and the other to capturing residual information. The residual components are thereby assumed to be part of the data that cannot be adequately learned by linear PDE and are delegated to conventional data-driven learning, via recurrent convolution using three layers of ConvLSTM cells. Residual components are mainly perceivable as pixel-level details, that are not primarily relevant to the latent physical dynamics, but remain a necessary part of accurate future frame prediction. The disentangled physical component is passed in parallel to a single layer of PhyCell, a novel physical cell functionally based upon the sum of a physical predictor and a correction term, concerning the interactions between latent state and the embedded input data. As in generalist video prediction "physical laws do not apply in the input pixel space", learning a semantic latent space disentangled from residual factors in which physical laws are applicable is the intended goal [65], which in terms of solar irradiance satellite maps may be more present but physical learning could also suffer under the lack of consistent background intensity. To derive the next frame, the partial results of the respective branches are added and a final decoder provides the output from the physical and residual components.

The physical component PhyCell is able to model a wide range of linear PDEs which correspond to a multitude of physical models, among others the heat equation, the wave equations and the advection-diffusion equations [65]. The process of using constrained convolutions for the learning of partial derivatives is reliant on discretizing the PDEs through the Euler numerical scheme, a commonality shared with other approaches like PDE-Net [66].

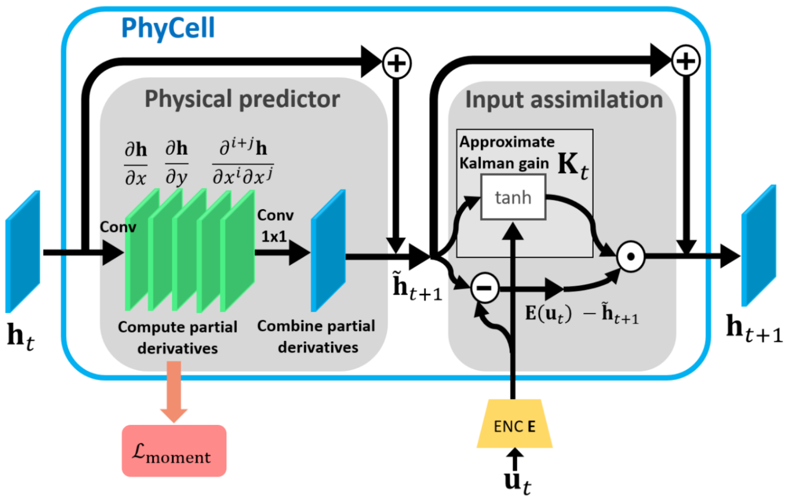


Figure 4.4.: PhyCell design and dynamics, from [65], ©2020 IEEE, Reprinted, with permission.

The implementation of PhyCell is based on convolutions, applying the filters as approximations of the individual partial derivatives. The cells are constrained by moment loss, a kernel filter that enforces the learning of the specified range of PDEs through the comparison of the moment matrix of the filter weights to a target moment matrix [65]. The target moment matrix is for this purpose the Kronecker matrix corresponding to the kernel size, which is applied as a loss term via the Frobenius Norm of the difference[67]. The correction term is gated by an approximation of Kalman gain, which if set to 0, omits the current input from being accounted in the calculation of the next timestep, thereby providing a "Prediction Mode" based only on the past physical hidden state[65]. This is a special feature of PhyDNet, that should allow for a better performance in the context of unreliable data, specifically missing data[4].
PhyDNet is designed as an auto-regressive model generating one frame into the future

[4]Prediction Mode was although not contained in the updated PhyDNet-dual source code, as a parameter that seemingly enabled it in the earlier version led to errors. It is therefore not considered in the

at a time and iteratively creating a predictive timeseries, encoding the input frames to derive new frames in a separate decoding iteration.

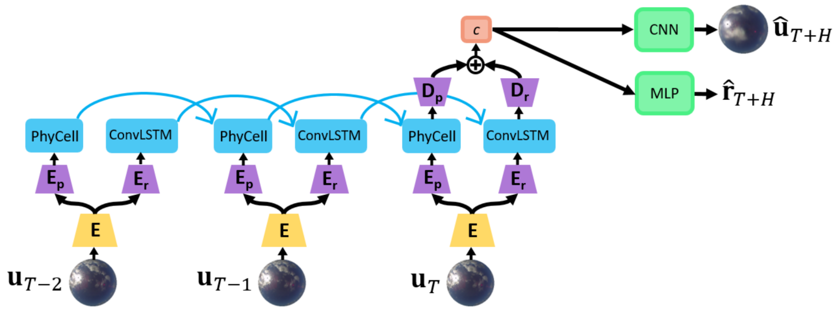


Figure 4.5.: PhyDNet-dual architecture, from [52], ©2020 IEEE, Reprinted, with permission.

PhyDNet is trained and tested in two variants, that exhibit differences in their internal architecture and parameter count. The first model variant PhyDNet-dual corresponds to changes in the published code for the more recent publication adapting PhyDNet for a solar irradiance forecast study with sky images [52], while the second was edited to incorporate architectural aspects found in the supplement to the original paper [67].

Source code and modifications

The source code of PhyDNet is provided publicly on GitHub[5]. The current version at the time of this comparison has received a commit changing the configuration from a single encoder and decoder to additional separate encoders and decoders for the convolutional and differential layers, which was introduced in [52]. Figure 4.5 is an illustration of this newer PhyDNet-dual architecture[6] from [52], its significant difference to the previous version of PhyDNet being the separate convolutional encoders E_p and E_r and decoders D_p and D_r for the respective physical and residual components of the model. Thereby separate image embeddings are learned for the physical and residual branch, which provided improvements in accuracy when compared to the original single encoder/decoder setup [52]. This change was introduced to provide separate embeddings and should be beneficial to the network's parallel streams of cells. This version corresponds to the first version of PhyDNet that was trained and tested, and will further be referenced as PhyDNet-dual.

Improvements to the initial performance of PhyDNet-dual were later achieved with the Sea Surface Temperature (SST) configuration from the original paper, which was imple-

comparison, as the accuracy benefits of the newer architecture, as presented in [52], were more of interest.

[5]https://github.com/vincent-leguen/PhyDNet

[6]The CNN and MLP (Multilayer Perceptron) output layers displayed in 4.5 (upper right) are only relevant to the study with fisheye images undertaken in [52], here irradiance is predicted via the MLP and the next frame via the CNN output layers.

mented based on the information provided in the supplemental material[67]. Out of the selection PhyDNet was originally benchmarked with, the SST dataset resembles the solar irradiance dataset closest. It can be reasonably assumed that this configuration should at least be among the most competitive of the different benchmark configurations. The most significant change for this variant is the fitting of the encoders with skip-connections to the decoder layers, with a relative increase in learnable parameters. The new separate encoders and decoders were kept for this variant, thereby a combination of one of the original configurations with the newer PhyDNet-dual redesign was created. All further mentions of this variant of PhyDNet will be in the form of PhyDNet-SST, with PhyDNet referencing both of the variants.

Another general change in the training configuration was an adjustment to the teacher-forcing-ratio[7], as the original training regime set the maximum number of epochs to 2001 with teacher-forcing being possible in the first 333 epochs. Differences in the dataset's spatial and temporal magnitude as well as the applied configuration, would have led to a potential maximum training time of up to two months for PhyDNet-SST, an extent of GPU-time that could not be assigned to one singular network in this comparison. Based upon time limitations but also a noticeable trend of convergence at less than 100 epochs, the teacher-forcing ratio was reduced to being active only within the first 20 epochs. A negative aspect of these modifications, in combination with the training routine used here, where instabilities, like exploding loss, after the Teacher Forcing epochs ended when training PhyDNet-SST. These instabilities were circumvented by reloading the last best checkpoint while maintaining the current optimizer state, thereby continuing training with a reduced learning rate. Overall PhyDNet-SST was harder to train than the other models, with the configuration outlined in Chapter 5.1 and required a higher number of training runs to get results that could be assumed close to the best achievable with this variant.

[7]Teacher forcing ratio is a technique that applies the target frame within the auto-regressive decoding stage, instead of the previously encoded input frames. The change of frames can occur per timestep to predict with a probability that is steadily decreasing with the increasing epoch count while training. The method's goal is to increase stability while training, leading to faster convergence in RNNs. Teacher forcing is not applied during inference.

4.2. IrradPhyDNet - own development

The idea behind the development of IrradPhyDNet is based on the different qualities observed on initial tests of PhyDNet-dual and IrradianceNet. PhyDNet-dual provided less accurate results, but showed a tendency to predict irradiance patterns under conditions of missing timesteps more accurately than other models [8], while IrradianceNet provided better forecasts when applying only clean data. The basis of these different benefits were assumed to lie in the autoencoder-ConvLSTM architecture, and the two-branch architecture of PhyDNet with the PDE-learning PhyCell. By combining these central elements, it was intended to derive relatively robust and accurate forecasting results.

Early attempts to achieve the benefits of both neural networks focused on creating a neural network ensemble, which was attempted with a variety of spatiotemporal aggregators, ranging from averaging and voting regressors to neural network designs that were considered to extract beneficial aspects of the respective forecasted frames. This approach was of course inherently limited and could not improve over the respective best performances of the individual models, but served as a baseline for continued development and led to the reconsideration of combining patterns and cells into a new model. This change would allow for the forecasting from the combined latent space representations instead of final forecasts reduced to a single feature. Enabling a final decoder to learn the nowcasts from the larger feature-space of separate disentangled branches was assumed to provide significant advantages over the ensemble approach.

The final version, named IrradPhyDNet, is the result of experimental attempts to combine architectural aspects from two models in the comparison, IrradianceNet[4] and PhyDNet[65], that seemed to contribute to the predictive performance in the context of clean and flawed data. These aspects include the ConvLSTM-autoencoder architecture of IrradianceNet (see Chapter 4.1.1) for more accurate forecasts on clean data and the physical disentangling concept, including the PhyCell design, from PhyDNet (see Chapter 4.1.3) for flawed data.

Leading to this final version, a variety of combinations of sub-networks were experimentally implemented and evaluated, with an initial consideration of an iterative single-step forecasting approach as used by PhyDNet, with adaptations to IrradianceNet's autoencoder. This did however provide worse results than transforming the PhyCell into a multi-step single-shot cell and using the existing code base of IrradianceNet. By focusing on IrradianceNet from which the autoencoder and base code structure was adapted, a parallel sub-network composed of a single layer of PhyCell and respective encoder and decoder convolutional layers was added, to create the disentangling architecture in relation

[8] This was later mitigated with the alternative configuration PhyDNet-SST as described above by adapting the additional information provided in the supplemental material[67]. These changes lead to a significant improvement in its base forecasting performance, voiding the initial assumptions about the limited use of PhyDNet architectures, which was focused primarily on the superior results when confronted with flawed data.

to the design of PhyDNet.

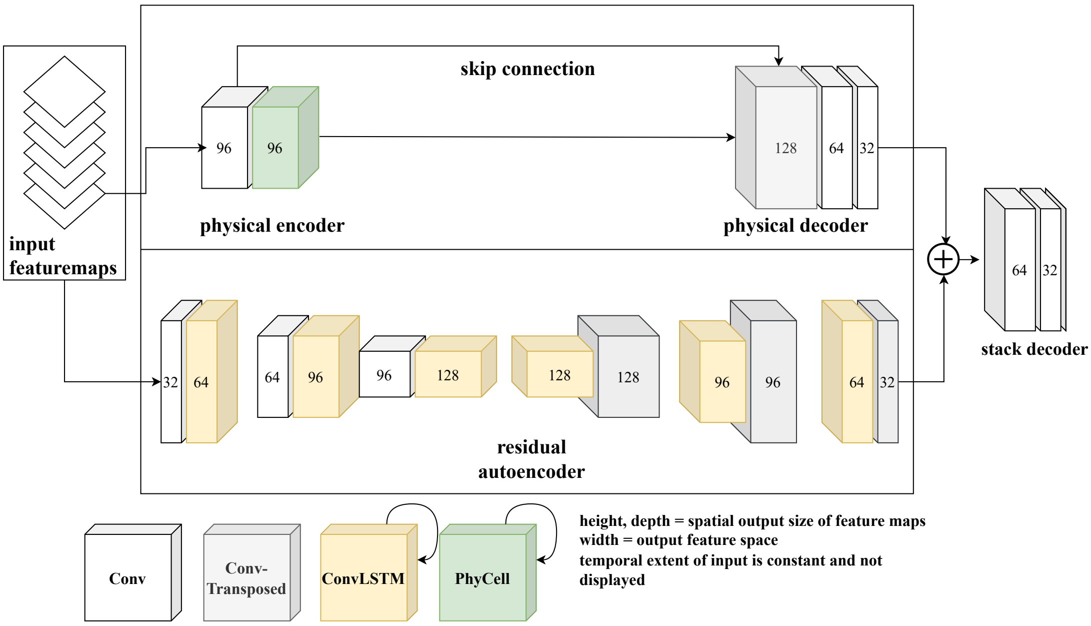


Figure 4.6.: Architecture of IrradPhyDNet

The disentangling architecture in its adapted form of IrradPhyDNet is presented in Figure 4.6: The upper section thereby correspond to the physical sub-network, with a modified version of the PhyCell from [65], and the lower section contains the sub-network taken from [4], which is intended to perform residual learning complementing the physical sub-network. The autoencoder structure is kept the same as in IrradianceNet [4], with a three layer deep downsampling encoder reducing the spatial extent to a quarter of the original size and a three layer deep upsampling decoder, to reconstruct the frames to the original spatial extent. To adjust to the new requirements for this sub-network, the filter kernel size of the ConvLSTM cells was reduced from 5x5 to 3x3, as the intent for this branch was not to capture large movements but to focus on finer details. This change also provided a significant reduction in learnable parameters, thereby facilitating the integration of the second branch in terms of required memory. The physical branch or sub-network based on the physical disentanglement and PhyCell concept of PhyDNet [65], is composed of a single encoder layer reducing the spatial size to half for the PhyCell to learn discretized PDE-based representations, using the same number of PhyCell features (49) and 7x7 kernels, as PhyDNet. The derived physical feature space is separately decoded[9] within the upper branch by multiple layers of convolutional cells, beginning with a transposed convolutional layer to restore the forecasts to the original input size. Two central adaptations were implemented in the physical branch, that provided respective

[9]Separate branch-wise encoders and decoders were also used by Le Guen and Thome in [52].

improvements to the forecast accuracy:

- PhyCell modification: To provide compatibility between the different forward predicting modes of the cells used in the models of origin, PhyCell's internal design was altered from its original single-step forwarding methodology to a self-iterating cell design, thereby being able to perform multistep single-shot prediction. This also incorporates a deviation from the original training and testing regime, where PhyCells, as well as the other components of the network are first iteratively encoding the true input frames, before iteratively decoding the predictions.

- Physical Skip-Connection: As the decoder following the PhyCell includes an up-sampling step to restore the original spatial extent of the input maps, a skip connection was introduced from the encoder to the first layer of the decoder. The outputs of the PhyCell and its upstream down-sampling encoder are concatenated, with the intent to facilitate the upsampling process and freeing the PhyCell from some of the additional requirements of learning an adequate mapping for upscaling. Thereby allowing the PhyCell to focus on dynamic components of the input, while sufficient context for up-scaling should be maintained over the skip connection.

To further lead to the PhyCell's effective usage as a physical predictor, the sub-network was only provided with irradiance data. This avoids any unnecessary learning of the encoder layer to ignore static inputs or even hinder learning of the PhyCell. In contrast to PhyDNet, this is increasingly important due to the reduced number of encoder layers and the omission of a common first encoding layer, that should otherwise be able to disentangle the relevant features. Static input was here instead delegated entirely to the ConvLSTM-sub-network at the separate input layer. To lead to the correct learning of the PDEs latent space by the PhyCell's convolutional filters, training was constrained with moment loss as used in PhyDNet. The resulting latent outputs from the individual sub-networks are then stacked channel-wise and processed in a final convolutional multilayer decoder to derive the spatiotemporal forecasts. Tests on alternative version of IrradPhyDNet with a corresponding symmetrical general Encoder and Decoder did hinder forecast accuracy, assumedly by having a negative impact on PhyCell's learning, as the moment loss decrease was reduced in these variants.

The goal of this endeavour was to combine beneficial traits of both network architectures, while not assuming a perfect combination was achievable. The experimentally derived combination solution IrradPhyDNet is one successful example of this effort, while other approaches were less effective and led to a deterioration in performance. In relation to the optimized PhyDNet-SST, IrradPhyDNet's architecture may not provide a significant improvement in base performance. The autoencoder architecture, that is assumed to be a more potent residual predictor than the three-layer ConvLSTM used in PhyDNet could hinder a delicate balance necessary to achieve effective distentangled physically constrained learning.

5. Methodology

5.1. Training configuration and hyperparameter tuning

The training configurations of the four different models in the comparison, were kept as close as viable to the descriptions in the publications or respective secondary sources given by the authors. This includes certain techniques, as for example teacher forcing and momentum loss in PhyDNet. Some limitations were although introduced to keep the overall training time to a sensible maximum, while also striving to provide the least detriment and high potential benefits, as well as a fair comparison.

Training configuration	
Input size	256 x 64 x 12 x 6
Output size	256 x 64 x 12 x 1
Batch size	4, 16 (PhyDNet-dual, PFST-LSTM)
Optimizer	torch.optim.Adam
Initial learning rate	0.001, 0004 (PFST-LSTM, IrradPhyDNet)
Learning rate schedulers	ReduceLROnPlateau(patience=5, factor=0.5)
Loss function	Mean Absolute Error (MAE)
Additional losses	Momentum loss (PhyDNet, IrradPhyDNet)
Special training techniques	Teacher Forcing (PhyDNet)
Regularizers	EarlyStopping(patience=15, restoreBest=True)

Table 5.1.: Training configuration

Differences in the original problem statements concerning the spatial and temporal range of inputs and outputs where not considered and set to the same range of 256 by 64 spatially by 12 input to 12 output timesteps. This, among other choices has to be considered as potentially detrimental to the average predictive performance in contrast to the achieved performances on the original problem definitions and data configurations.

5.1.1. Learning rate and early stopping

The initial learning rate (LR) was set to values defined in the corresponding papers, being 0.0004 for PFST-LSTM and initially 0.002 for IrradianceNet, while using the default LR of 0.001 for the two other models. IrradianceNet didn't perform well with this choice and showed exploding loss on the final dataset, but training stability could be recovered by reducing the LR to the same default value of 0.001. IrradPhyDNet also suffered from unstable training, which also could be greatly reduced with a reduction of the LR to 0.0004. The initial learning rate was not kept constant over the training epochs to mitigate the issue of getting stuck in local minima. Validation loss was fed into a Reduce Learning Rate on Plateau scheduler, which halved the learning rate after 5 epochs of patience if no new optimum could be achieved.

By introducing early stopping with a high enough patience period, a relatively optimal solution can be expected to be reached. A patience period of 15 epochs was selected, that showed to be ideal in terms of training time reduction and achievable performance in initial tests. Early stopping triggers after a new optimum on the validation data was achieved and followingly not surpassed, if for 15 epochs no new optimum was achieved the weights are reset to the previous optimum and training ends. The maximum number of epochs was set to 125, as this was a sensible individual maximum training time of about 5 days, that appeared necessary in the context of the variety of models (and secondary training configurations; see chapter 5.2). In actuality, all models stopped earlier or in one single case showed clear signs of overfitting, i.e. diverging training and validation losses and degrading validation Mean Squared Error (MSE), terminating four epochs before early stopping would have taken effect[1].

5.1.2. Batch size

Batch-sizes were chosen as the largest power of 2 possible with the utilized hardware and the chosen dataset. The default batch-size of four[2] corresponds to the batch-size used for IrradianceNet, IrradPhyDNet and PhyDNet-SST. While PhyDNet-dual and PFST-LSTM are able to use a larger batch-size of 16. Where some models had parameters based on a statically predefined batch-size, these were changed to a dynamical read out of the current batch-size or fixed to the new batch-size. Automatic mixed precision (AMP) was used to facilitate the training process on the more memory-intensive models and assumed to have no significant detrimental effect on the quality of the output[3]. As this method is only practically viable on graphical processing units (GPUs) with CUDA-support from the last three generations at this point in time, training was done on a Nvidia 2080-Ti with 12GB of VRAM.

5.1.3. Loss function

All models were initially intended to be trained with MSE loss, as it seems to be most commonly used loss function in spatiotemporal neural networks. While training was

[1] In this case, as would have occurred if terminated by early stopping, the last best checkpoint was used.
[2] The batch size of four is achieved with Automatic Mixed Precision (AMP) enabled, a batch size of three would otherwise be the approximate equivalent on a GPU with 12GB of VRAM without applying mixed precision.
[3] https://pytorch.org/blog/accelerating-training-on-nvidia-gpus-with-pytorch-automatic-mixed-precision/

without issues in the early epochs, the MSE loss sometimes deteriorated to NaN[4] after a medium to high number of epochs. As this issue wasn't fixable via LR reductions or a variety of suggested fixes, the loss function was instead chosen to be MAE. The loss function is kept the same over all models to keep its effects equal and avoid obvious differences in the comparison metrics based on optimization for a specific metric through the loss function. As this study is undertaken as a comparative effort of different neural network designs with limited resources, no separate focus was placed on the loss functions used, on this topic see for example Tran and Song's study on the effect of loss functions on precipitation forecasting accuracy [46]. Concerning the choice of loss function taken here: Even while Tran and Song considered MAE a "poor choice" in the context of combined loss functions, by still showing a 2.3% better performance than only MSE as a loss function, MAE was considered a viable choice of loss function and in the context of their findings concerning standalone loss function use an even more ideal choice.

5.2. Robustness to missing data

A model's performance under ideal conditions is valuable and relevant to understand its forecasting capabilities in an optimized, error-free end-to-end pipeline. In a controlled experimental setting, data is often considered clean by default and pre-processed to this respective goal where necessary. In contrast data sources in a real forecasting application may vary in reliability and accuracy over their lifetime. Missing data can not only occur in raw data but also in a processed product, like the one used in this comparison. Issues during measurements, data aggregation and transmission can introduce flaws and gaps in the data, as observed in chapter 3.1.1. Even when these events don't generally occur to a worrying extent, a forecasting method's ability to cope with an input with missing values can be considered vital in these exact instances.

An alternative to the typical delegation of flawed data to imputation techniques, lies in a model's robustness to certain kind of data flaws. Approaches to improving a models robustness, encompass techniques like the use of specialized loss functions [55], spatial redundancy via proximity of ground stations [56], modified input layers utilizing probabilistic modelling of input values for partially missing data [68] or recurrent [69] and convolutional imputation layers [70]. However, these approaches are mainly focused on uni- or multivariate timeseries and would require specific adaptations to potentially achieve similar results on spatiotemporal data.

[4]Pytorch's Automatic Mixed Precision may lead to this behaviour when MSE loss is used, at the time of writing there was an open issue on pytorch's GitHub repository (https://github.com/pytorch/pytorch/issues/40497). Automatic Mixed Precision was although a strongly beneficial factor concerning memory usage as well as training time, being an essential tool in this comparative effort's interest on time and scope.

5.2.1. Simulating missing data

The different neural network architectures, as well as different training techniques may lead to higher levels of robustness to missing timesteps. To derive comparative and reproducible estimates of these capabilities within the range of compared neural networks, a probabilistic replacement of timesteps was implemented.

$$\forall T \in X : f(T) = \begin{cases} 0_{T_{m*n}} & \text{if } p < c \\ T & \text{else} \end{cases} \tag{5.1}$$

X : sample of 12 timesteps T

T : spatial timestep of shape $m * n$

c : constant threshold, $c \in \{0.1, 0.25, 0.5, 0.75\}$

p : pseudo-random number drawn from $U([0.0, 1.0])$

The missing timesteps are randomly replaced with the application of the per-timestep function f(T), thereby providing a pseudorandom replacement of individual timestep-grids, with a per-run fixed constant corresponding to the expected probability of a value(-grid) being replaced. The seed of the pseudorandom number generator is fixed to avoid comparative issues derived from inconsistent missing value placement and imputation, thereby avoiding the complication of varying forecast difficulty. As the simulated missing timesteps take the form of zero-filled arrays in the same shape as the input, they are a minimalist form of invalid value replacement, corresponding to the missing irradiance values for nightly measurements. The value replacement is not applied on the whole feature-set, only the irradiance feature DSSF. The static spatial and temporal features are kept intact, as they would be easily insertable or reproducible.

5.2.2. Timestep-dropout - Training on imperfect data

The different neural networks showcased varying inherent robustness to missing timesteps in pre-tests. This inherent model robustness was intended to be increased via training on imperfect data, based on the assumption of enabling the models to learn to differentiate beneficial data from flawed data. This assumption goes somewhat against the general methodology of training on a perfect and clean dataset to derive more accurate models and may also show corresponding flaws when confronted with clean input data in certain problem statements. But it can be esimtated, that a model trained on a flawed dataset, if confronted with a similarly flawed data source, should derive higher accuracy. While no preliminary assumptions about the results derived with clean data in contrast to a model trained only on clean data can be made in a general sense, the technique may also serve as a regularizer in temporal and spatiotemporal networks, in essence functionally similar to intra-layer or more accurately input-layer dropout. In lack of a clearly associated term, this method will further be called timestep-dropout, meaning here a probability-based dropout applied before the input layer. This technique is thereby not equivalent to the usual meaning of the term dropout as being an intra-model conditional removal of

values. The process of dropping out individual timesteps uses the approach for simulating missing timesteps described in 5.2.1. The intensity of timestep-dropout is governed by the hyperparameter p, corresponding to the probability threshold. P is fixed during training and contrasted with a pseudo-random value derived for each timestep in the input features temporal range, thereby providing a fixed probability with varying extent and position of dropped timestep features. The value for p that was empirically determined optimal was 0.3, or a 30% chance of a timestep being replaced (see supplement A.1.2).
This technique was adapted from [65], where training and testing on a variety of ratios of missing data was employed to test robustness. As training on the same amount of missing data, that is expected for prediction, is not feasible for real-world data sources, the implementation here instead uses a fixed amount of missing values during training, analogous to dropout layers. While this appears to have so far not been referenced as a general technique to improve model robustness or provide regularization, there are a variety of related methods.

Some earlier work with similar elements and approaches exists, incorporated into convolutional and recurrent neural networks with the intent to serve as a regularizer: In [71] a type of timestep-dropout, called "timestep jitter", was included in the decoder layer of a convolutional speech representation learning network, thereby applying the dropout in latent space instead of the input data. Dropblock [72] applied masks to drop out contiguous regions of feature maps on ResNet-50 for an image classification task. Another similar technique introduced in the context of training RNNs is Zoneout [73], which instead of dropping or replacing timesteps or values with zeros, takes the previous timestep's value and applies it again. Zoneout has predecessors in different approaches of recurrent dropout [74, 75, 76], which are similarly applied as regularization techniques with the intent to improve general training performance. All techniques so far mentioned are applied at the intra-layer level, with the intent to regularize the respective models. More direct similarity can be attested to Input Dropout introduced in [77], which applied channel-wise dropout to images with additional modalities in the first layer, to improve among other tasks 3D-object tracking in the context of occlusions. With Input Dropout, the stochastically selected channels are dropped as groups of actual features to be present during testing and the additional modalities. Timestep-dropout is, in contrast, more directly inspired by actual data phenomena and needs no introduction of additional modalities. Since a model trained with timestep-dropout is not requiring any further external pre-processing or other preparatory work to facilitate robust prediction, similar to regularization, it's effects are contained in the learned weights. The timestep-dropout trained model can be used standalone. The concept was not created with the intention of having a potential regularizing effect but strives to condition the model during training to expect and handle flawed input data better than the same model trained without this technique.

5.2.3. Data imputation versus robustness

Robust models or integrated mitigation techniques are not a predetermined necessity when confronted with data sources that occasionally show missing or faulty values. These flaws can be substituted via imputation of the missing or faulty data in pre-processing. In this study, the dataset initially included a limited, but under real-world forecasting conditions in individual instances potentially problematic set of missing and erroneous values. Data imputation techniques in the domain of solar time-series have been introduced in a wide range of more sophisticated techniques from the MICE algorithm [78], a Markov Chain Monte Carlo method, ML-based imputation of solar irradiance via other meteorological features [7], to Generative Adversarial Networks trained for data imputation [79].

To simulate a variety of less and more extreme data flaws, a range of values was replaced with zero-filled grids as described above but will also be contrasted against the same flaws alleviated by temporal data interpolation. The imputation method used was linear interpolation over an array of input timesteps, along the time dimension. While there are more sophisticated methods of imputing missing values, linear interpolation has been found to provide very adequate results on an hourly solar power timeseries with less than 20% of values missing [80], and similar results found with adequate results for up to 50% missing values in [81] on 5-minute interval solar data. Linear interpolation is enhanced with extrapolation to fill missing values that are not surrounded by real values with meaningful information. Since linear extrapolation introduces the risk of overextending the range above or below the real minimum and maximum, all input data samples were clipped to the expected range from 0.0 to 1150.0.

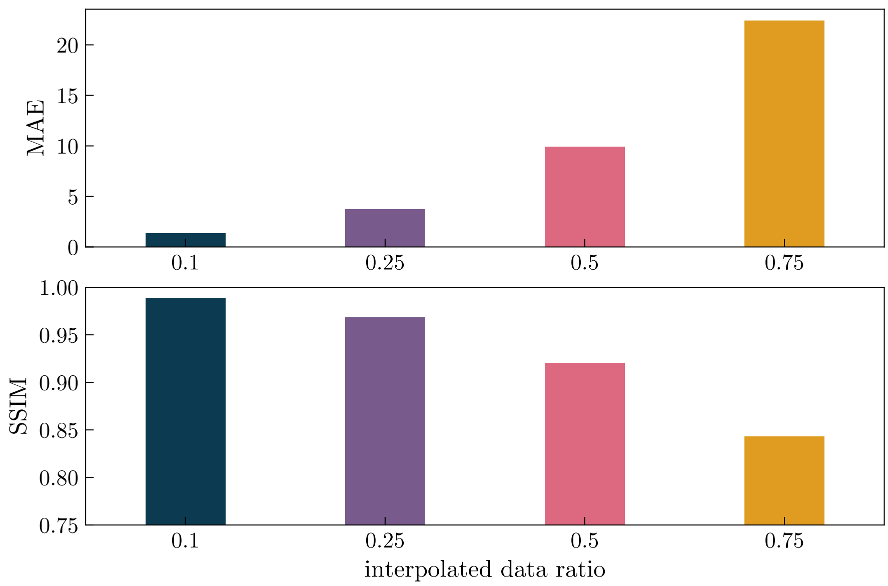


Figure 5.1.: MAE and SSIM for interpolated test input-data

Figure 5.1 illustrates the achieved imputation quality derived from applying linear interpolation on the four simulated test-datasets with missing timesteps. The linear interpolation shows a similar degradation of results with the increasing number of missing values to interpolate as observed in the literature referenced above. While the spatiotemporal dataset used here can be assumed to being generally more error prone to impute. Models trained with timestep-dropout will also be tested on interpolated data. In addition to a validation in all configurations, it was assumed, that exposure to data flaws could lead these trained models to place less importance in individual timesteps, that may contain suboptimal information through the interpolation.

5.3. Evaluation metrics

To give a wide comparative perspective on the forecasting results achieved with the different models and techniques, a selection of quantitative metrics and relative measures was used.

Mean Absolute Error (MAE)

The absolute distance between forecasted values and the corresponding ground truth values is measured by applying MAE. The metric takes an aggregate average that doesn't discriminate errors by their extent and is thereby robust to outliers.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{5.2}$$

As MAE was used as the training loss, the neural networks' scores in this metric are assumed to be more optimized. To provide a more complete perspective on the achieved accuracies, the other metrics described below are inlcuded.

Root Mean Squared Error (RMSE)

To penalize larger deviations more than smaller deviations between prediction and ground truth, Mean Squared Error (MSE) applies squaring on the individual error distances. By squaring, the calculated mean error does however no longer correspond to the original scale of the data. To re-establish the ability to provide more meaningful and intuitive comparisons with the value range of the irradiance data, like MAE, RMSE is derived by applying a square root on the aggregate MSE. RMSE will be included to highlight the occurrence of outlying errors in the forecasted values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{5.3}$$

Mean Bias Error (MBE)

The metrics described above are based on the absolute and squared distances, thereby ignoring the positive or negative sign of the error. Yet, in meteorological prediction, as well as other related tasks such as PV power prediction, the existence of a general positive or negative bias in the forecast is a relevant attribute to test for. In the context of PV or solar irradiance forecasting a predicted surplus or overproduction has a different impact than the possibility of underproduction of PV power. By taking the aggregate of the signed differences, an existing average bias can be revealed, but other flaws may remain hidden, as complementary positive and negative distances cancel each other.

$$MBE = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i - y_i \tag{5.4}$$

Structural Similarity Index Measure (SSIM)

In contrast to the above metrics, which are based on direct grid value comparisons in spatiotemporal tasks and don't cover concepts of consistency or structure within the spatial grid areas, SSIM [82] will be applied as a perceptual image quality metric.
With SSIM, blurriness and thereby a loss of local detail can be estimated, which is a common problem in spatiotemporal predictive models and strived to be reduced or overcome [16, 83, 84]. Blurriness is equivalent to loss of detail in individual areas, it is here included as a measure of overall spatial quality of the results.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{5.5}$$

μ : sample mean (signal luminance)

σ^2 : sample variance (signal contrast)

σ_{xy} : discrete covariance between samples

C : constants to avoid instabilities with values close to 0

Forecast Skill (FS)

To contextualise forecasting results into a frame of reference which is not only determined by the absolute difference to the ground truth, but relative to the complexity and difficulty in a specific forecasting task, Forecast Skill can be used. Forecast skill derives a relative improvement of a model or method over a chosen baseline method, based on their respective results derived via an error metric. The metric calculated on both, the contending model and the baseline method, is not predetermined and can be selected on a case-by-case basis depending on the usefulness for a specific problem.

$$FS = 1 - \frac{d(\hat{y}, y)}{d(\hat{y}_{pers}, y)} \tag{5.6}$$

$d \in$ metric

The most common choice of baseline method are variants of the Persistence method. Persistence, in its naïve variant, assumes that the value of the next k timesteps will be the same as the current value. The accuracy of this method depends on the amount of timesteps and temporal resolution, and can be competitive in very-short-term forecasting, especially more sophisticated versions of Persistence [85]. The general formulation of naïve Persistence is:

$$\hat{y}(t + h) = y(t) \tag{5.7}$$

t : time of observation

h : temporal offset

6. Results

In this study a variety of spatiotemporal models were comparatively included to showcase achievable results derived from clean data, the robustness to missing timesteps and their localized accuracy on a selection of sites. These evaluations will also highlight the effects of the training technique timestep-dropoout. The single-site and single-area evaluations include a separate method, the Analog Search, to estimate the models accuracy differences in relation to a more advanced baseline.

All models have been trained multiple times to arrive at a relative certainty that representative results can be given. The comparisons below showcase the models with the respective best performance achieved on clean input data (as in Chapter 6.1). To avoid a dilution of difference in the metrics by nightly data, all forecasted values correspond to the temporal interval from 6:00 to 19:00 (CET). This temporal filtering of the forecasted values also contributes to the avoidance of exaggeratedly positive appearing results.

6.1. Nowcasting performance of spatiotemporal deep learning models

In Figure 6.1 all models, trained on either the regular training routine or with timestep-dropout, are shown in terms of SSIM and RMSE. This initial evaluation on the dataset without missing values shows the margin of difference between the individual models, which all achieved an overall respectable performance. The newly introduced Irrad-PhydNet shows slight improvements in RMSE and SSIM over the other models, in its timestep-dropout trained variant, while it's regular trained variant is performing very similarly to IrradianceNet. PhyDNet-SST improves significantly over PhyDNet-dual and is in both trained variants very accurate within the scope of this comparison. When evaluating the effect of the different training regiments, it is noticeable that timestep-dropout has a beneficial effect on IrradPhyDNet, PFST-LSTM and PhyDNet variants, with consistent gains in their achievable RMSE, while the effect on IrradianceNet is slightly negative. Effects in terms of SSIM across the models are different, showing minor detrimental effects for all models except IrradPhyDNet. What can be assumed from these results, is that timestep-dropout has some regularizing effect that alters the achievable predictive accuracy, but directionality and extent are not consistent for different neural networks.

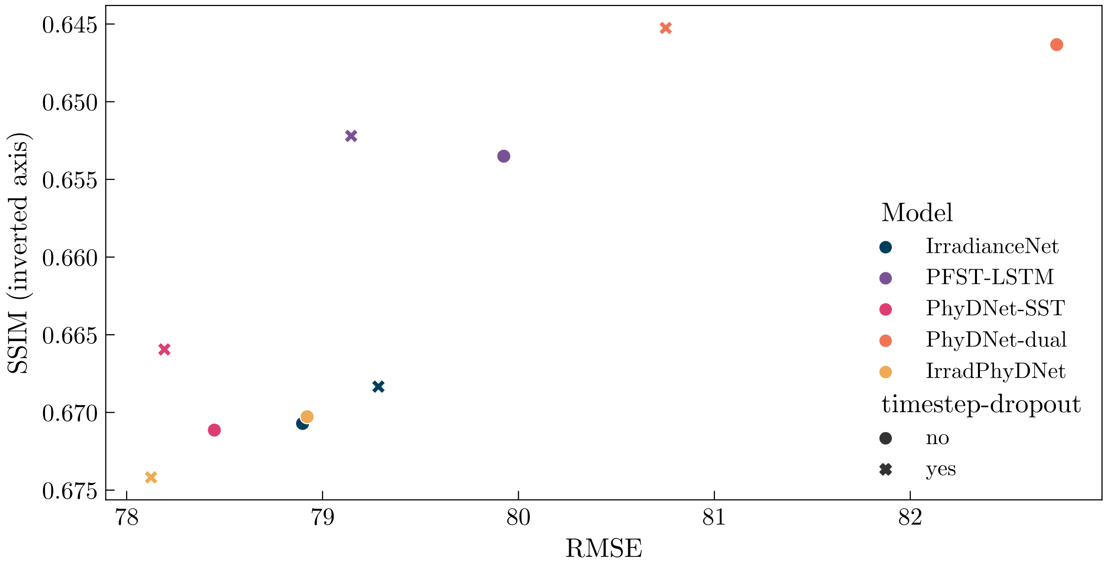


Figure 6.1.: Mean SSIM and RMSE - all models with and without timestep-dropout

To further illustrate the mean achieved results, Table 6.1 shows individual metrics for the models, in their regular training and with timestep-dropout applied during training. Two noteworthy differences can be observed in the more significant reduction in MAE for PFST-LSTM and PhyDNet-dual as well as a changes in MBE to lower deviations for all models in the comparison.

Model	RMSE ↓	MAE ↓	SSIM ↑	MBE =	FS-RMSE/MAE ↑
IrradianceNet	78.90	44.36	0.671	-1.18	(0.598, 0.687)
with timestep-dropout	79.29	44.56	0.668	**-0.59**	(0.596, 0.686)
PFST-LSTM	79.93	44.73	0.654	-4.25	(0.593, 0.685)
with timestep-dropout	79.15	44.41	0.652	-3.19	(0.597, 0.687)
PhyDNet-SST	78.45	**43.55**	0.671	-3.35	(0.601, **0.693**)
with timestep-dropout	78.19	43.84	0.666	-1.83	(**0.602**, 0.691)
PhyDNet-dual	<u>82.75</u>	<u>47.38</u>	0.646	<u>-4.32</u>	(<u>0.579</u>, <u>0.666</u>)
with timestep-dropout	80.75	46.10	<u>0.645</u>	-3.00	(0.589, 0.675)
IrradPhyDNet	78.92	44.15	0.67	-1.60	(0.598, 0.689)
with timestep-dropout	**78.12**	43.68	**0.674**	-0.68	(**0.602**, 0.692)

Table 6.1.: Mean metric results on clean input data - all models with and without timestep-dropout

The selection of spatiotemporal neural networks has shown to be competitive among each other, with only PhyDNet-dual performing a little worse, thereby validating the initial selection of models. All models achieve an approximate Forecast Skill over Persistence of close to 70% in terms of MAE and approximately 60% in RMSE. Based on the individual results, the different architectures are in a relative proximity that leaves any of them as a viable choice.

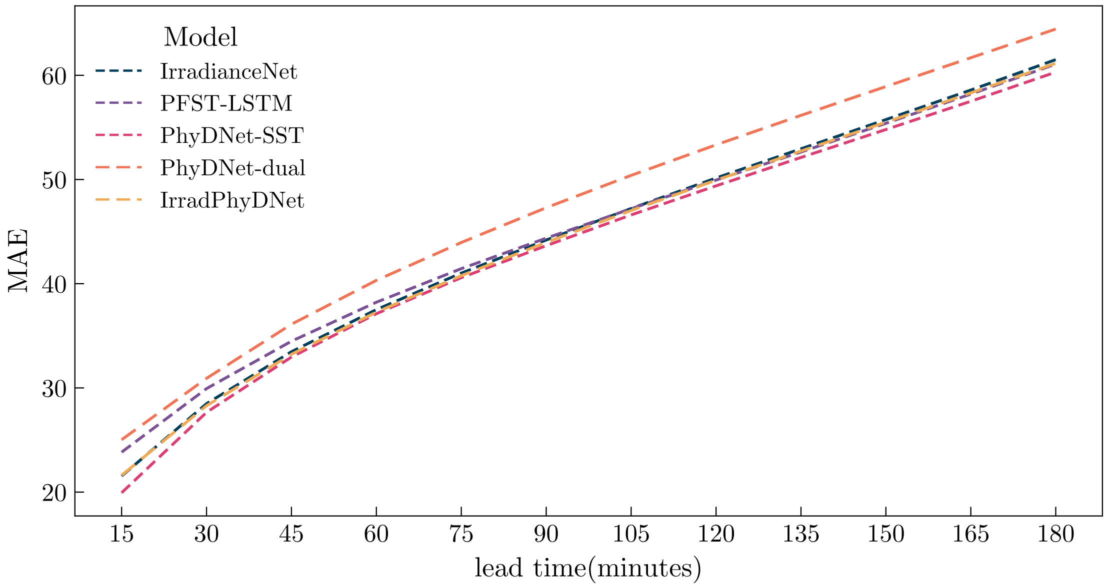


Figure 6.2.: Mean MAE over lead time - all models without timestep-dropout

In Figure 6.2 all models are compared in their achieved MAE over the range of lead times. Here, IrradianceNet, PhyDNet-SST and IrradPhyDNet perform relatively similarly, with PhyDNet-SST achieving the lowest error over most of the lead times while PhyDNet-Dual performs worst over the whole range and PFST-LSTM improves noticeably, from similar to PhyDNet-dual in early forecasts, toward the later lead times. Through a comparison with the results achieved with timestep-dropout training, shown via the relative forecast skill to the regularly trained models in Figure 6.3, specific effects of timestep-dropout training become apparent. While IrradianceNet has overall no benefit from being trained with timestep-dropout, IrradPhyDNet shows a general positive effect, which increases with later lead times. The same development can be seen for the PhyDNet variants and PFST-LSTM, while for the former there is also a significant degradation of early lead time performance, for the latter the same effect is visible but less pronounced. Overall timestep-dropout training seems to lead to models favouring later lead times, over the earlier lead times.

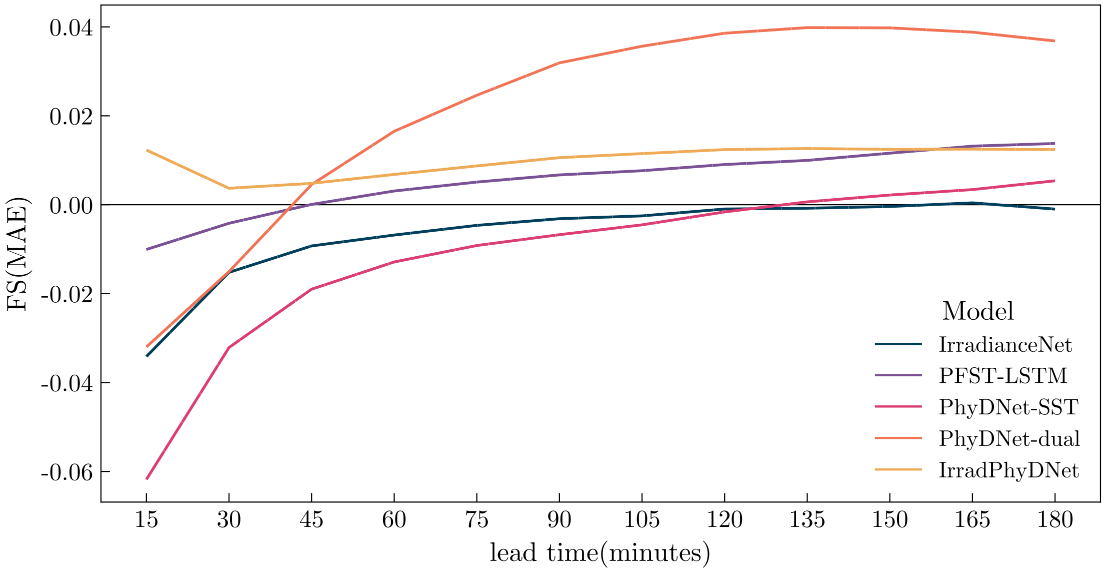


Figure 6.3.: Forecast skill (MAE) over lead time - models trained with timestep-dropout in relation to models trained without

To illustrate qualitative aspects, two nowcasting samples are showcased in Figure 6.4 with the ground truth data for the respective timeframes and metrics to contextualize the frames within the larger context of the overall forecasting performance. The forecasts were inferred with IrradPhyDNet and temporally cropped to 30-minute slices from the quarter-hourly output frames. Some propensity to estimate cloud movement and dissolution can be seen in the sample output. Temporal changes in overall irradiance across the target area seem to be forecasted accurately in the samples displayed, likely significantly supported by the inclusion of the timestamp-derived features. Similar to the quantitative degradation of results towards later lead times outlined above, there is an observable blurriness to later forecasted frames. The prediction of opacities, most likely clouds, drifting into frame from the edges of the frame is one of the expected weaker aspects of the predicted irradiance maps, which would be mitigatable with further out-of-frame context or respectively cropping the forecasted frames.

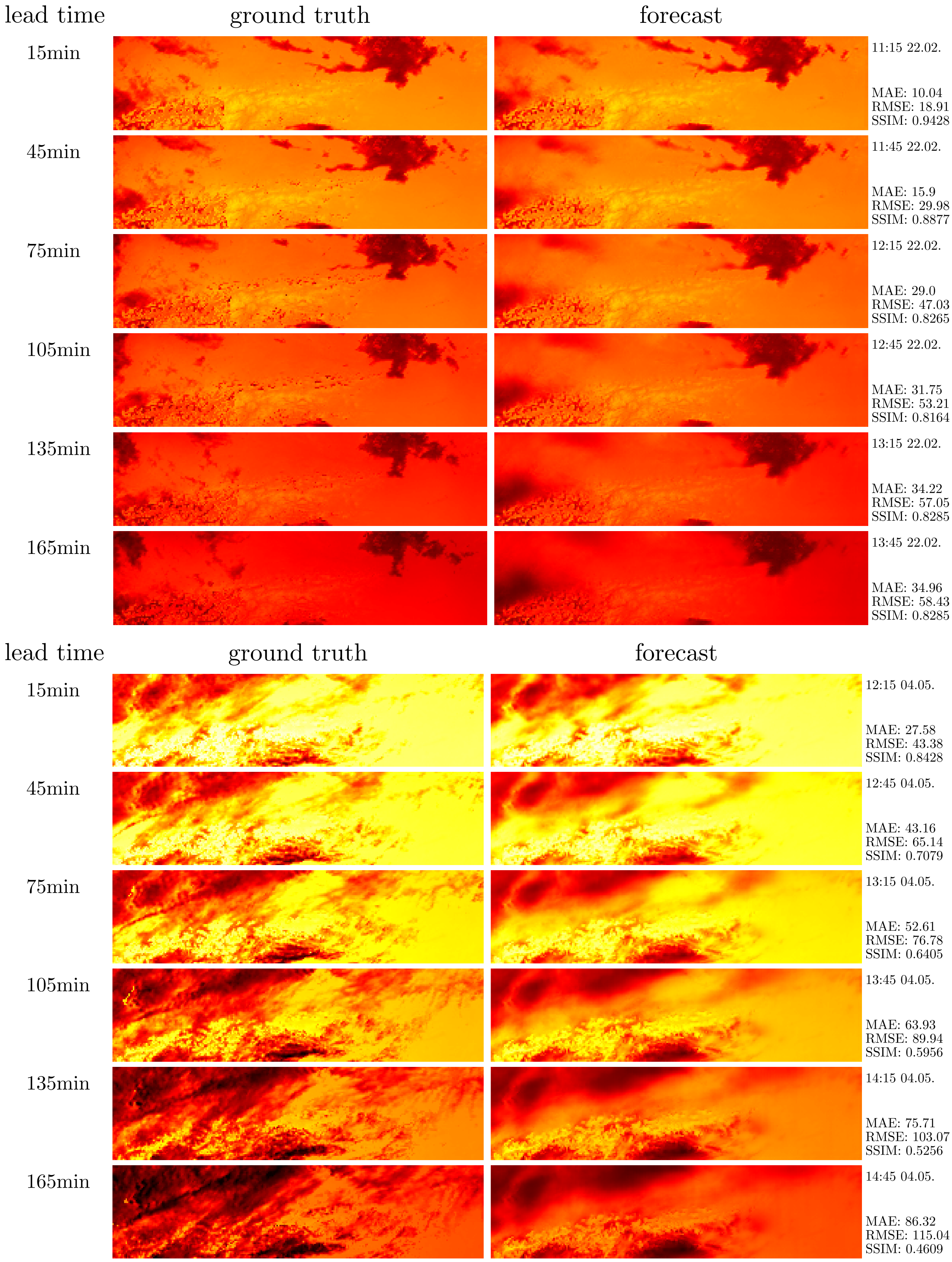


Figure 6.4.: Forecast samples from IrradPhyDNet, half-hourly slices

6.2. Robustness evaluation

As discussed above, the occurrence of missing or invalid values is relevant to consider in real nowcasting applications. To facilitate such an evaluation, the method outlined in Chapter 5.2.1 was applied on the input timesteps of the test dataset, with multiple sets of test results with different probabilities of removing irradiance data from the timesteps.

The motivation behind the introduction of timestep-dropout, has proven very valid, as can be seen in Figure 6.5.

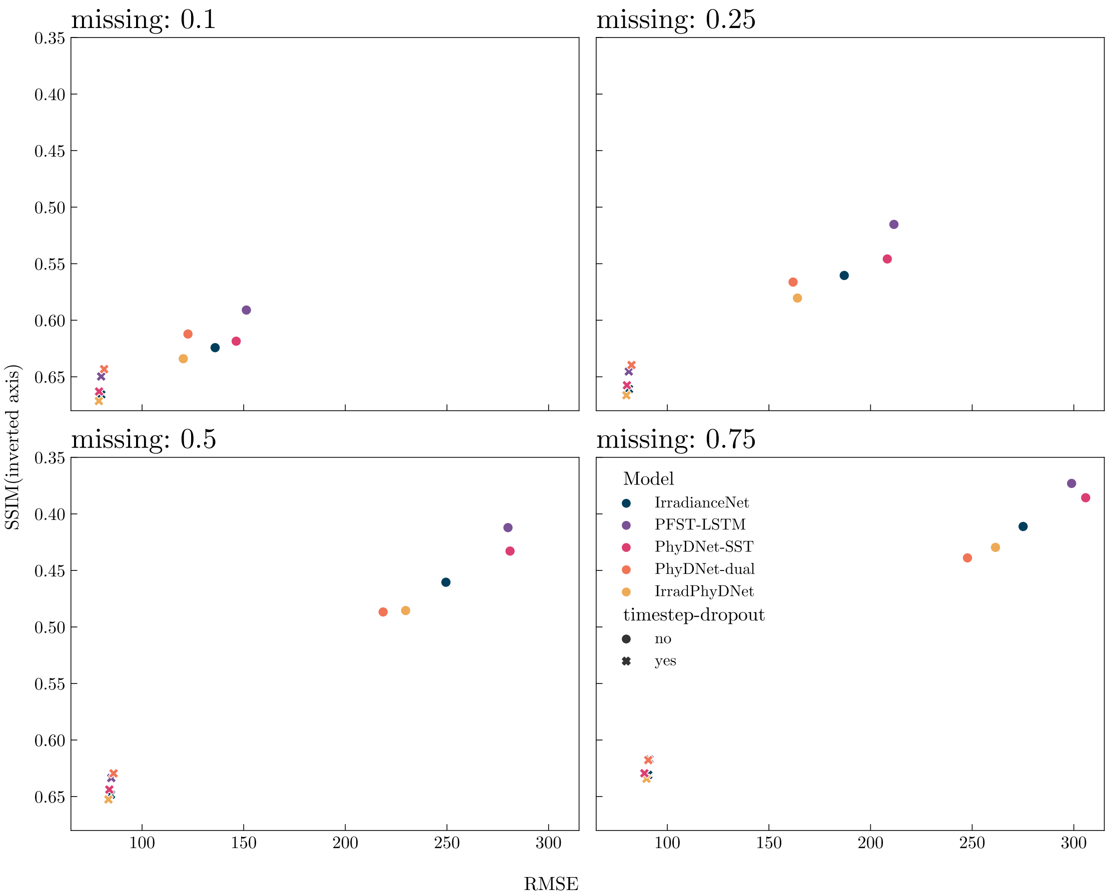


Figure 6.5.: Mean SSIM and RMSE over range of missing values (10% - 75%) - all models with and without timestep-dropout

While the models trained without timestep-dropout perform increasingly worse, with the loss of additional irradiance frames, the timestep-dropout training regiment leads to a significantly lower negative impact. When not considering timestep-dropout, the performance in terms of RMSE and SSIM degrades steadily with increasing amount of missing timesteps. Among the models trained regularly, IrradPhyDNet and PhyDNet-dual show the least degradation or respectively the highest accuracy on missing data. PFST-LSTM, that also was not designed with attempts on or claims to robustness, and PhyDNet-SST perform overall worse in this comparison of intrinsic robustness. Concerning PhyDNet-SST this is again a notable difference to PhyDNet-dual, as was inversely found in the previous comparison.

Model	data	RMSE ↓				SSIM ↑			
		0.1	0.25	0.5	0.75	0.1	0.25	0.5	0.75
IrradianceNet	M	135.94	187.05	249.49	275.06	0.624	0.560	0.461	0.411
	I	81.23	85.46	99.23	116.80	0.664	0.652	0.620	0.582
with timestep-dropout	M	79.93	81.14	84.70	90.63	0.666	0.661	0.648	0.631
	I	80.60	83.43	94.23	109.66	0.664	0.655	0.627	0.592
PFST-LSTM	M	151.32	211.49	280.03	298.91	0.591	0.515	0.412	0.373
	I	82.80	87.68	104.28	127.28	0.648	0.637	0.608	0.573
with timestep-dropout	M	79.86	81.07	84.87	91.29	0.650	0.645	0.634	0.617
	I	81.04	84.92	98.63	117.78	0.648	0.640	0.616	0.587
PhyDNet-SST	M	146.31	208.23	281.04	305.84	0.619	0.546	0.433	0.386
	I	81.52	86.95	101.85	119.47	0.663	0.648	0.612	0.579
with timestep-dropout	M	78.84	80.13	83.89	**88.73**	0.663	0.658	0.644	0.629
	I	80.42	84.72	97.59	112.83	0.660	0.648	0.619	0.591
PhyDNet-dual	M	122.58	161.88	218.60	247.68	0.612	0.566	0.487	0.439
	I	84.51	87.44	97.02	109.42	0.642	0.633	0.610	0.585
with timestep-dropout	M	81.33	82.39	85.98	90.66	0.643	0.640	0.630	0.618
	I	82.30	85.12	94.13	105.36	0.642	0.634	0.613	0.590
IrradPhyDNet	M	120.28	164.03	229.68	261.46	0.634	0.580	0.486	0.430
	I	81.30	85.70	99.53	117.21	0.664	0.651	0.617	0.578
with timestep-dropout	M	**78.74**	**79.87**	**83.47**	89.75	**0.671**	**0.666**	**0.653**	**0.634**
	I	79.75	83.19	95.34	111.10	0.669	0.659	0.629	0.593

Table 6.2.: Results from missing (M) and interpolated (I) data - all models with and without timestep-dropout

Over most of the range of probabilistically removed timesteps IrradPhyDNet trained with timestep-dropout performed best in terms of pixelwise error and on the whole range in structural similarity, with PhyDNet-SST achieving a lower RMSE on 75% missing timesteps. The effect of timestep-dropout is also consistent in this broader comparison, as also all other models achieved their best performance when trained with timestep-dropout. The interpolation of missing data in the input showed to be viable and provided regularly trained models with adequate results up to a range of 25% of timesteps missing, while degrading significantly beyond this margin. The application of interpolation on the missing input values had an equalizing effect on the models, removing most of the differences found when accounting for their intrinsic robustness to missing timesteps. This clarifies that there is no strict necessity for a different training regime, if only very occasional lesser gaps, that could be effectively interpolated, are expected. For models trained with timestep-dropout, there is a second noteworthy difference in their tendency to achieve higher accuracy when not confronted with the interpolated data, while timestep-dropout also showed improvements on the interpolated data.

In extreme cases, complete temporal gaps in the measurements or respective processed products may occur, that leave no values to be interpolated and extrapolated from in the input timeframe. In Figure 6.6, the whole range of previously shown missing percentages is prolonged by the case of 100% irradiance data missing. This leads to a necessity to extract relevant information for the forecasts from the remaining elevation and temporal features. Estimates can only be provided based of the time of day and day of year, aided by the topographical elevation feature. The included results for Persistence deviate in this aspect from the other models in Figure 6.6, as Persistence received the required single solar irradiance map it requires and thereby is kept constant. While most of the models are not capable of improving upon the Persistence derived result when no irradiance input data is provided, IrradPhyDNet and PhyDNet-dual, trained with timestep-dropout, are capable of providing slightly better forecasts even in this case. Showing a potential benefit of the disentangling architecture, if conditioned with timestep-dropout, to be able to rely not only on dynamic features and utilize the static features more effectively, in cases where the prior is not provided.

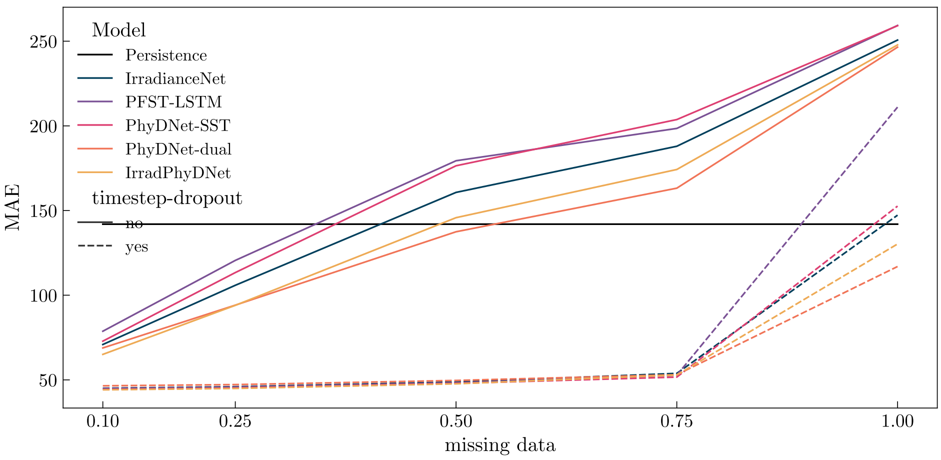


Figure 6.6.: Mean MAE for range of missing values (10% - 100%)- all models with and without timestep-dropout

Figure 6.6 also shows a higher robustness in terms of MAE for both, IrradPhyDNet and PhyDNet-dual, over most of the range of ratios. While PhyDNet-dual displayed a higher error than IrradianceNet in Table 6.2 in terms of RMSE, this is mostly different concerning absolute errors, showing the architecture to be in this context more robust overall, but having issues with outliers. Taking a closer look at the results for only the timestep-dropout trained models (Figure 6.7), shows similar results with MAE as observed above with RMSE, specifically for IrradPhyDNet and PhyDNet-SST, while again PhyDNet-dual was seemingly less able to gain more robustness from timestep-dropout, likely due to the limitations of the smaller architecture.

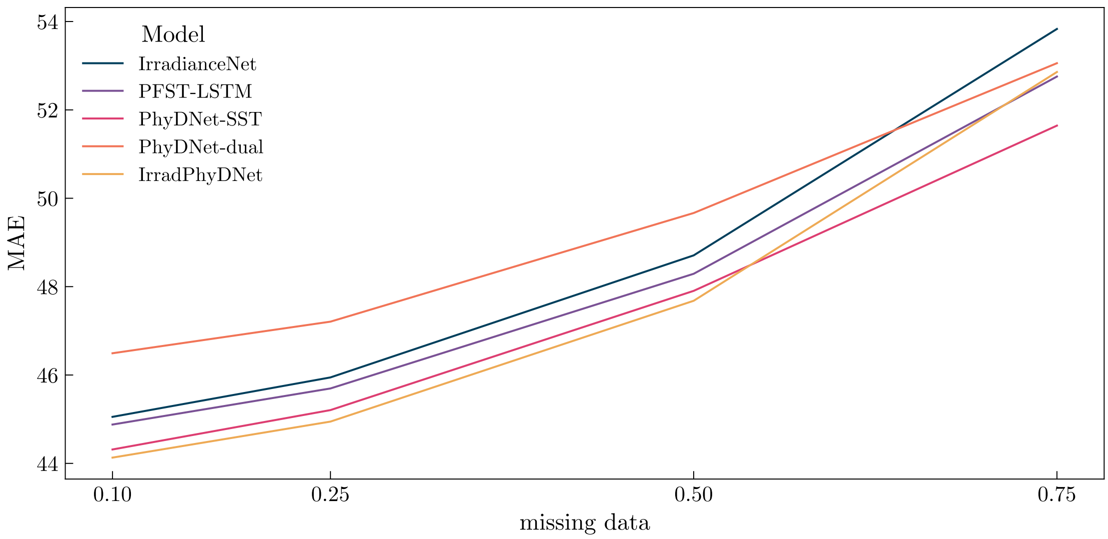


Figure 6.7.: Mean RMSE for range of missing values (10% - 75%)- all models with timestep-dropout

Below, in Figure 6.8, are again qualitative samples showcasing the output frames of IrradPhyDNet trained with timestep-dropout and regularly trained. In this case the results are forecasted from input data with 50% probability of a frame being removed. For comparison not only the ground truth but the same outputs without missing input data are included. As can be seen, the forecasted frames from the timestep-dropout trained model do lose some detail but the overall structure is maintained, giving still usable frames. Contrary, the regularly trained IrradPhyDNet, while being among the most intrinsically robust models, degrades significantly and can in the first instance maintain some patterns observable in the ground truth for some of the forecasted timesteps, while in the second the outputs degrade to a barely similar irradiance map.

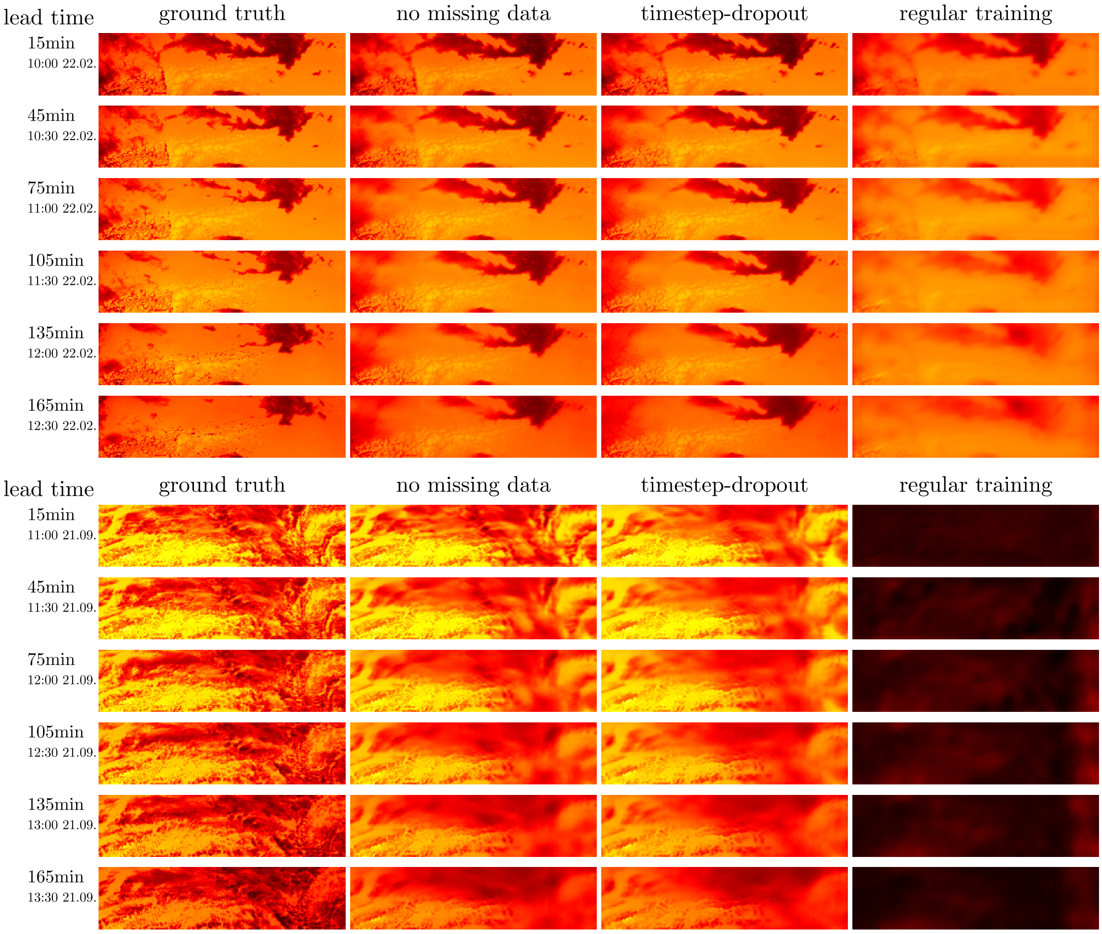


Figure 6.8.: Forecast samples from IrradPhyDNet, 50% missing data half-hourly slices

6.3. Performance of spatiotemporal models in local forecasting[1]

As solar irradiance forecasts are important for PV power output prediction and PV power plants can be relatively isolated, forecast quality for individual local stations or cells within the target area should also be considered in the context of large scale nowcasting. To evaluate the models' performance on this aspect, a comparison to GHI measurements from BSRN sites within the target area and a secondary comparison with the LSA-207 grid cells around the BSRN sites, was undertaken.

To compare the models to an established technique for local area prediction from spatiotemporal satellite data, an Analog Search technique (for details see supplement A.1.3), implemented based on the paper by Ayet and Tandeo [17], was used here as an advanced baseline method. The technique was originally designed for 6 hour-ahead, hourly forecasts for single grid areas from a larger spatiotemporal map.

6.3.1. Site forecasting evaluation on BSRN sites

The aforementioned BSRN stations were used to derive validation results for the spatiotemporal forecasts on a site-oriented setting and contrasted against the Analog Search's nowcasting capabilities as an advanced baseline. The results can be seen in Table 6.3.

	MAE ↓			RMSE ↓		
Model	Budapest	Payerne	Sonnblick	Budapest	Payerne	Sonnblick
Persistence(GHI)	148.43	147.59	159.55	202.49	206.92	213.38
Analog Search	100.15	90.45	123.83	151.87	133.32	181.49
IrradianceNet	**61.00**	**65.88**	91.94	**106.77**	**106.05**	141.62
PFST-LSTM	62.13	66.12	**87.48**	108.65	106.80	**135.64**
PhyDNet-SST	61.41	67.23	90.21	107.71	108.55	140.81
PhyDNet-dual	64.68	68.24	89.91	111.20	108.58	140.31
IrradPhyDNet	61.43	67.10	89.35	107.09	108.65	139.20

Table 6.3.: Results on BSRN site GHI

While all models' site nowcasts improve over the Analog Search's performance, in this setting, there's a notable shift in the relative performance of the four neural networks. PFST-LSTM shows the best performance on Sonnblick, while IrradianceNet performs best on Payerne and Budapest. Overall, as the results show an improvement over both baseline methods, they appear viable for site forecasting, but as they also use no local data, the performances for individual sites seem to tend to favour the different spatiotemporal

[1]For this comparison a more limited dataset of 2000 samples has been used, which have been randomly chosen by starting date and time within average daylight hours. For brevity a further comparison with timestep-dropout training is omitted here, the full tables are supplied in the Appendix A.1.4.

models relatively arbitrarily. When compared over the temporal range of lead times (see Figure 6.9), the Persistence method utilizing the sites GHI data shows to be superior to the other methods within the first 15 minutes in Payerne and, to all but PhyDNet-SST, even 45 minutes in Sonnblick, while the Analog Search is at the earliest more effective at 45 minute predictions and takes 90 minutes to exceed Persistence on Sonnblick.

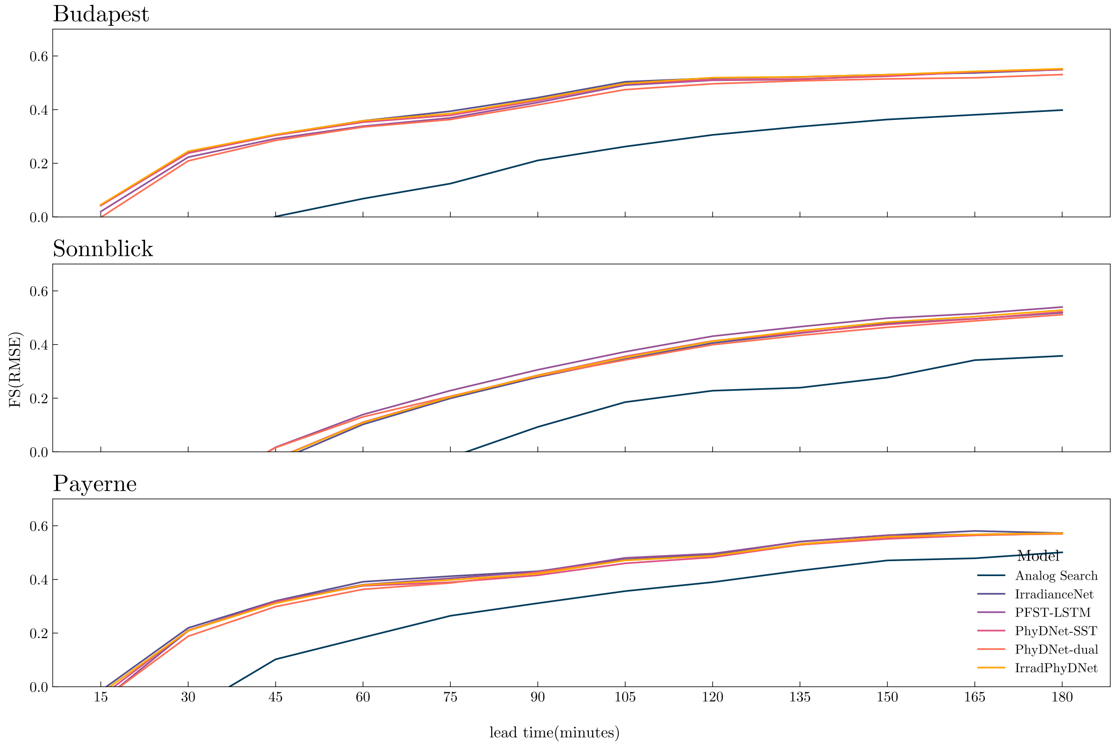


Figure 6.9.: Mean Forecast Skill over BSRN sites - all models without timestep-dropout

The site of Sonnblick is however relatively extreme in its location and Budapest and Payerne are both close to the edge of the target area, leading to a potential lack of spatial context information to achieve higher accuracy.

6.3.2. Local area forecasting evaluation

As the above results are based on the BSRN station GHI data, they are restricted to a more limited spatial perspective compared to the source dataset's resolution and scope. To enhance this perspective with local area forecasting results, a further study on the source dataset was undertaken. To provide a more meaningful sample for the evaluation of local area forecasting performance Hohe Warte, or respectively the north of Vienna, Austria[2], was chosen as an additional location. As there is no equivalent GHI dataset

[2]Hohe Warte: Latitude - 48.243, Longitude - 16.355, Altitude – 196m

available, the data for this site exclusively, and for the results below as well for all other sites, corresponds to the closest grid cell in the DSSF data. This approach is beneficial to the models' performance as it keeps forecasts within the same spatial domain, that is more appropriate for approximate solar irradiance results over multiple square kilometre cells than a specific site. The selection thereby contains two target cells in mountainous terrain, Payerne and Sonnblick, and two target cells in hilly to flat terrain, Hohe Warte and Budapest.

The Forecast Skill relative to the naïve Persistence method along the lead time, in Figure 6.10, shows again a relatively large margin between the neural networks and the Analog Search. Especially within the first 45 minutes, where the Analog Search is not able to surpass Persistence reliably, while the spatiotemporal neural networks are able to improve upon Persistence's results over the whole range of lead times. These results are thereby significantly more in favour of the neural networks, than the site-based results shown in Chapter 6.3.1.

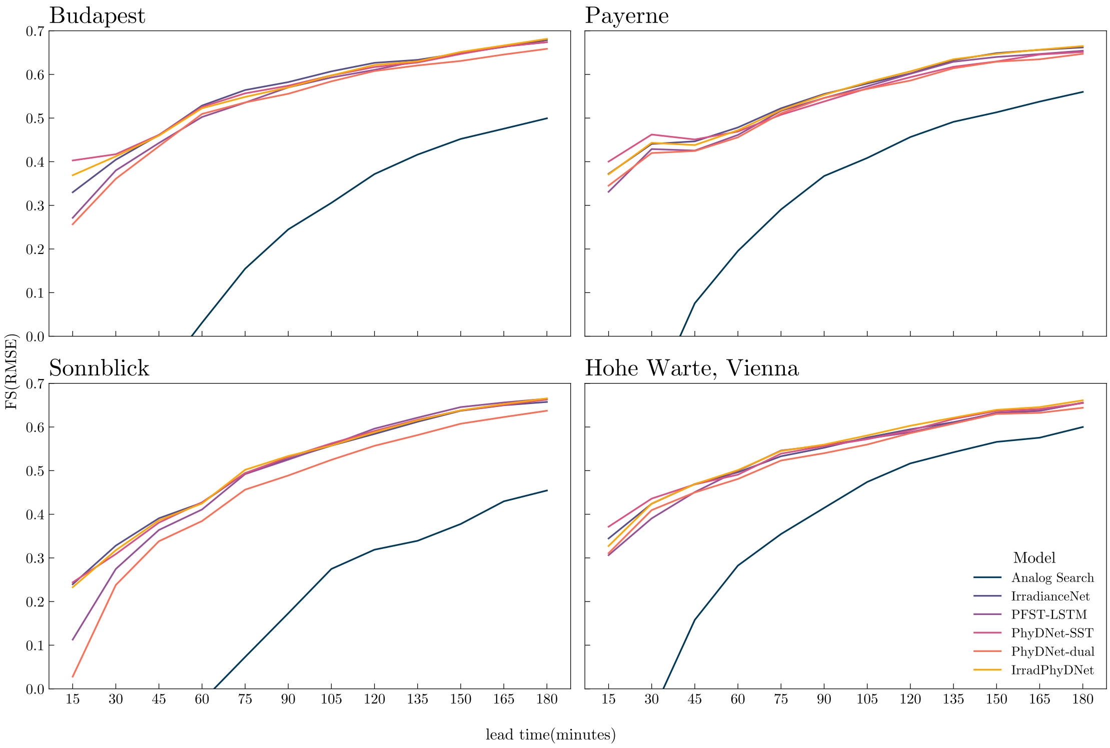


Figure 6.10.: Mean Forecast Skill over closest grid cells to sites - all models without timestep-dropout

Model	MAE ↓				RMSE ↓			
	Budapest	Payerne	Sonnblick	Hohe Warte	Budapest	Payerne	Sonnblick	Hohe Warte
Persistence	<u>144.06</u>	<u>151.10</u>	<u>160.09</u>	<u>142.7</u>	<u>194.32</u>	<u>208.19</u>	<u>214.74</u>	<u>195.12</u>
Analog Search	86.57	81.87	109.74	70.95	132.42	121.23	159.73	104.66
IrradianceNet	38.85	**46.43**	56.12	43.11	**74.26**	**82.64**	90.14	79.46
PFST-LSTM	39.30	47.08	57.05	42.83	76.20	84.36	90.29	79.79
PhyDNet-SST	**38.29**	46.64	**54.97**	**42.15**	74.90	84.89	89.68	78.99
PhyDNet-dual	42.30	49.59	62.05	45.19	78.34	86.15	97.70	81.27
IrradPhyDNet	38.55	46.50	55.32	42.84	74.71	83.85	**89.24**	**78.95**

Table 6.4.: Results on closest grid cells

Compared to the site-based GHI error metrics in the previous chapter, all models perform relatively better, with the spatiotemporal neural networks again achieving lower errors to ground truth than the Analog Search. The results in terms of MAE and RMSE for the neural networks are relatively close to results derived for the whole spatial map (see Chapter 6.1). Differences can be observed between the selected grid areas, notably Budapest's surroundings provided the best results for all neural networks. For the area around Sonnblick, PhyDNet-SST and IrradPhyDNet were more competitive, achieving respectively lowest RMSE and MAE. Similarly on Hohe Warte the models achieved the lowest absolute error and had the lowest issues with outliers as can be assumed from RMSE. The lower errors for PFST-LSTM when evaluated on BSRN site GHI data for Sonnblick (Table 6.3), are not consistent, here on the local area forecasts PFST-LSTM shows the second highest errors compared to the other neural networks.

The control results from Hohe Warte show relatively good results on all models, especially the Analog Search baseline, it seems to benefit the most from a more central and geographically less extreme site. The spatiotemporal neural networks are somewhat more robust to these edge cases and achieve consistently higher results than Persistence. Nonetheless the difference in accurate predictions in relation to the neural networks, highlights a benefit of their predictive capability over the whole frame. The models can equally not estimate developments beyond the horizon of the spatial grid, but suffer less from these edge cases. To contextualize the results in defence of the Analog Search method, it should be noted that while its performance is overall worse than the spatiotemporal neural networks considered here, it was not intended for the temporal resolution and more limited scope used here and it is the most flexible approach: While the neural networks would likely require retraining to be expected to perform well on a different target area, the Analog Search can dynamically accept a dataset from a different region with a different spatial extent.

7. Discussion and potential for future work

7.1. Discussion

The goal of this thesis is a comparative effort of established spatiotemporal neural networks, that were assumed viable for solar irradiance nowcasting from a data-driven perspective, with an additional focus on achievable robustness to missing data. The comparison of the models led to the development of a new neural network, based on architectural patterns observed, that is performing better or comparative to the other models. The achieved results of this effort were presented above and allow for the assessment that these goals could be met.

The different spatiotemporal neural networks all achieved relatively similar performance in solar irradiance nowcasting on clean data, with PhyDNet-dual being slightly less accurate in regular nowcasting from clean data. PhyDNet-SST on the contrary achieved some of the best results, while also having the most issues during training. While predictive performance deviates from the original publications on other tasks, the data-driven perspective on the selection of neural networks proofs valid. The large area forecasting of solar irradiance has shown minor performative differences between the models with and without timestep-dropout. Here the new model, IrradPhyDNet, in combination with timestep-dropout training, as well as PhyDNet-SST were performing overall best by a slight margin. In the context of clean input data, timestep-dropout has shown to provide some training regularization, which provided better forecast results for most models and specifically increased the models' abilities to maintain higher accuracy in the later lead times. The effect could be found in multiple error metrics, most clearly RMSE, which could be reduced for all models but IrradianceNet, with the application of timestep-dropout. Timestep-dropout appears to have some benefit as a regularizer, while this effect may also be achievable, and more consistently, with other regularization techniques, it is here only considered as a secondary aspect, as the primary intention was to achieve more robust models via this training regime.

As spatiotemporal forecasts over larger areas are not only in themselves relevant but as predictions for single sites within the spatial grid, a comparison within four local grid cells and three BSRN sites was undertaken. Relative to Persistence as well as the implementation of the Analog Search method, the neural networks showed, with the exception of early lead times in extreme locations such as Sonnblick, significantly better performance. This is also notable as the models where not in any way optimized or receiving any post-processing to increase predictive accuracy for the chosen local grid cells. This highlights some of the potential of centralized large grid forecasts from satellite derived data to predict the overall solar irradiance derived electrical power within the geographical region corresponding to an electric grid.

In terms of forecasting from input data with missing timesteps, all models showed significant increases in predictive error, if no mitigation strategy was applied. Interpolating missing values has shown to be quite adequate in restoring relevant dynamics in the input feature-maps and leading to much improved results. These were however still exceeded by results derived with timestep-dropout trained models. The increase in robustness to all models comes relatively close to a full compensation of accuracy loss due to missing values and only decreases in effect at and above 50% missing timesteps, which should be sufficient for many real-world applications. The effect seems relatively agnostic to the specifics of the model architectures and achieves similar improvements relative to the base robustness on all compared models. The exact influence this method has on the learned parameters of the models is so far unclear but based on the effects in terms of robustness, it can be assumed that it may impact the temporal memory of the recurrent cells in that it changes the memory gating in its selection and exclusion of certain timesteps from being propagated. Thereby the zero-array inputs, but also the interpolated data, may have less influence on memory states and allow the model to focus on more useful timesteps and valid information contained in the other features. Another aspect can be assumed to lie in the more effective use of the other features, when a model experienced gaps in the irradiance feature during training. Overall, it has been found that all models either need pre-processed data or the effects of trainable robustness, via timestep-dropout, to achieve adequate results under missing timesteps in the input data. Intrinsic robustness of model architectures was not found to be sufficient in this comparison.

The here newly introduced model IrradPhyDNet, has achieved to be robust to missing timesteps while maintaining accuracy in forecasts from clean data. When compared with all other models, notably the optimized PhyDNet-SST, it is not superior in every case, but, especially with timestep-dropout, achieves a relatively good balance between best- and worst-case forecasting, as defined in this study.

The results presented above do although most likely not correspond to the best achievable results from the models and datasets used and should not be considered as such. Through the wide scope of the comparison and focus on relatively large spatiotemporal models, a limit had to be set on the amount of training data to reduce epoch-wise and overall training time. Overfitting was one of the more pronounced issues that was attempted to be compensated, using a larger part of the dataset could have reduced issues and further increased forecast accuracy. The specified training configurations may as well not have been ideal for all neural networks but were chosen with the same interest of limiting overall training time to an acceptable amount, while striving to achieve representative results within this scope.

7.2. Future work

As a comparative effort with limited hardware resources, including training on a new dataset as well as separate research on robustness resulting in a modified training routine, not all relevant strains of spatiotemporal predictive neural networks could be included. Notable mentions that could merit further study are the MetNet networks [22, 45] and the generative adversarial forecaster Deep Generative Model of Radar (DGMR) [86], which were intended for larger scale precipitation forecasting. Some models and techniques were experimented with but left out of the final comparison, primarily due to a lack of progress within a sensible timeframe. The use of adversarial loss to improve forecasting performance and the use of PredRNN [87], led to forecasts found to be inadequate and are likely the result of an insufficient adaptation to this specific forecasting task. Further attempts at including these approaches may be interesting but could not be considered here.

Further studies on improvements derived by including an autoencoder architecture in the disentangling IrradPhyDNet would be necessary to estimate it's true benefits. While the presented version of IrradPhyDNet showed mostly better performance than the other models, there is likely room for further improvements to the design. This includes changes to the existing architecture, as well as further attempts at incorporating other cell and module designs. Experiments on including the PFST-LSTM's pseudoflow spatiotemporal-memory cells showed a lack of similar benefits in combination with the physically disentangling architecture, but should not be discarded completely as the experimental recombination attempt here could not reasonably cover all possible architectural patterns. In all cases, further in-depth studies could show additional benefits or shortcomings, as the wider scale of this study limited the potential in this regard, a follow-up could have merit.

Similarly, to estimate the exact effects and generalizability of timestep-dropout and its robustness enhancing and regularizing effects, further research is necessary. As this study was limited on a specific dataset and a single target area and range of lead times, no final conclusive statements to the general effects in spatiotemporal and recurrent neural networks can be given here. But the so far achieved results would merit further broadened as well as in-depth studies on the effects of this training technique in spatiotemporal as well as recurrent neural networks, especially if trainable robustness is of interest.

8. Conclusion

The need for more accurate and available solar irradiance nowcasts, induced by a shift to more extensive PV power production, could be resolved with new deep learning models. Given the shape of the data being a spatiotemporal 3D-tensor, a datacentric view leads to the assumption that neural networks introduced to solve other tasks, like next-frame prediction, should provide viable predictive performance. Forecasting results can however differ based on specific architectural designs.

In this thesis, a comparison of different spatiotemporal neural networks for the task of solar irradiance nowcasting over a wide area, encompassing Austria and parts of neighbouring countries, was performed. The dataset for these wide area forecasts, a solar irradiance data product, derived primarily from satellite imagery, was combined with cyclically encoded timestamps and an elevation map as secondary features. In the range of nowcasting, the for this dataset highest possible temporal resolution of 15 minutes, in 12 timesteps summing up to three hours, was symmetrically applied as the temporal input and output range. The selection of models was based on a common denominator of including a relatively recent type of recurrent convolutional cell, the ConvLSTM, or a derivative improvement to this cell design. Neural Networks utilizing ConvLSTM and similar cells designs have been established to be the current state-of-the-art in spatiotemporal prediction. The evaluation of the models on the dataset, without missing values, lead to the conclusion that a variety of different approaches was similarly viable for solar irradiance nowcasting. Cell designs and characteristics from the best performers in multiple regards were experimentally recombined into IrradPhyDNet, a new spatiotemporal neural network design, that surpassed the competition by a slight margin in terms of mean forecast accuracy with an alternative training technique. These results were complemented with forecasting of single sites or respective grid cell areas within the spatial domain, where a non-deep machine-learning approach could be outperformed in all aspects by the neural networks.

The second field of interest was robustness to missing temporally distributed data, which is derived from practically occurring problems, as was also found in the dataset used for this study. The models showed varying degrees of intrinsic resilience, but overall, the decay in terms of forecast error corresponded to and increased drastically with the extent of missing data. This decay was greatly reduced with a technique applied in this thesis: Timestep-dropout, a method of training recurrent models on missing data, that also showed some partial benefits as a regularizer for predicting from clean data. More importantly, this training technique led to a great reduction of the performance decay from missing timesteps, making forecasts on unstable data sources more viable.

This was contrasted with missing timesteps ad-hoc pre-processed via linear interpolation. Forecasts with interpolated data were more accurate from all models, but less so than results from timestep-dropout trained models with unaltered flawed input data. Overall the use of spatiotemporal neural network designs seemed effective for solar irradiance nowcasting. Models can, with certain configuration and pattern changes, potentially be further improved. While missing data can be an issue in practical applications, with the right incorporation of mitigation strategies it is less challenging and still lead to viable nowcasting results.

Bibliography

[1] R. Ahmed et al. 'A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization'. In: *Renewable and Sustainable Energy Reviews* 124 (1st May 2020), p. 109792. ISSN: 1364-0321. DOI: 10.1016/j.rser.2020.109792. URL: https://www.sciencedirect.com/science/article/pii/S1364032120300885.

[2] Luyao Liu et al. 'Prediction of short-term PV power output and uncertainty analysis'. In: *Applied Energy* 228 (15th Oct. 2018), pp. 700–711. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2018.06.112. URL: https://www.sciencedirect.com/science/article/pii/S0306261918309826.

[3] Francis M. Lopes et al. 'Short-term forecasts of GHI and DNI for solar energy systems operation: assessment of the ECMWF integrated forecasting system in southern Portugal'. In: *Solar Energy* 170 (1st Aug. 2018), pp. 14–30. ISSN: 0038-092X. DOI: 10.1016/j.solener.2018.05.039. URL: https://www.sciencedirect.com/science/article/pii/S0038092X18304729.

[4] Andreas H. Nielsen, Alexandros Iosifidis and Henrik Karstoft. 'IrradianceNet: Spatiotemporal deep learning model for satellite-derived solar irradiance short-term forecasting'. In: *Solar Energy* 228 (1st Nov. 2021), pp. 659–669. ISSN: 0038-092X. DOI: 10.1016/j.solener.2021.09.073. URL: https://www.sciencedirect.com/science/article/pii/S0038092X21008306.

[5] Manajit Sengupta et al. *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: Third Edition*. NREL/TP-5D00-77635. National Renewable Energy Lab. (NREL), Golden, CO (United States), 1st Apr. 2021. DOI: 10.2172/1778700. URL: https://www.osti.gov/biblio/1778700.

[6] Dazhi Yang, Panida Jirutitijaroen and Wilfred M. Walsh. 'Hourly solar irradiance time series forecasting using cloud cover index'. In: *Solar Energy*. Solar Resources 86.12 (1st Dec. 2012), pp. 3531–3543. ISSN: 0038-092X. DOI: 10.1016/j.solener.2012.07.029. URL: https://www.sciencedirect.com/science/article/pii/S0038092X12003039.

[7] Vichaya Layanun, Supachai Suksamosorn and Jitkomut Songsiri. 'Missing-data imputation for solar irradiance forecasting in Thailand'. In: *2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*. 2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE). Sept. 2017, pp. 1234–1239. DOI: 10.23919/SICE.2017.8105472.

[8] Junho Lee et al. 'Reliable solar irradiance prediction using ensemble learning-based models: A comparative study'. In: *Energy Conversion and Management* 208 (15th Mar. 2020), p. 112582. ISSN: 0196-8904. DOI: 10.1016/j.enconman.2020.112582. URL: https://www.sciencedirect.com/science/article/pii/S0196890420301199.

[9] Ata Madanchi et al. 'Strong short-term non-linearity of solar irradiance fluctuations'. In: *Solar Energy* 144 (1st Mar. 2017), pp. 1–9. ISSN: 0038-092X. DOI: 10.1016/j.solener.2017.01.008. URL: https://www.sciencedirect.com/science/article/pii/S0038092X17300178.

[10] Laura S. Hoyos-Gómez, Jose F. Ruiz-Muñoz and Belizza J. Ruiz-Mendoza. 'Short-term forecasting of global solar irradiance with incomplete data'. 12th June 2021. arXiv: 2106.06868 [cs, stat]. URL: http://arxiv.org/abs/2106.06868.

[11] Fei Wang et al. 'Wavelet Decomposition and Convolutional LSTM Networks Based Improved Deep Learning Model for Solar Irradiance Forecasting'. In: *Applied Sciences* 8.8 (8 Aug. 2018), p. 1286. ISSN: 2076-3417. DOI: 10.3390/app8081286. URL: https://www.mdpi.com/2076-3417/8/8/1286.

[12] Gang Zhang et al. 'Solar forecasting with hourly updated numerical weather prediction'. In: *Renewable and Sustainable Energy Reviews* 154 (1st Feb. 2022), p. 111768. ISSN: 1364-0321. DOI: 10.1016/j.rser.2021.111768. URL: https://www.sciencedirect.com/science/article/pii/S1364032121010364.

[13] Zhenzhou Peng et al. 'Solar irradiance forecast system based on geostationary satellite'. In: *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. 2013 IEEE International Conference on Smart Grid Communications (SmartGridComm). Oct. 2013, pp. 708–713. DOI: 10.1109/SmartGridComm.2013.6688042.

[14] Yueyuan Li, Xuemei Chen and Ming Yang. 'Optical Flow Based Solar Irradiance Forecasting in Satellite Images'. In: *2019 IEEE International Conference on Real-time Computing and Robotics (RCAR)*. 2019 IEEE International Conference on Real-time Computing and Robotics (RCAR). Aug. 2019, pp. 442–447. DOI: 10.1109/RCAR47638.2019.9043950.

[15] Soumya Tiwari, Reza Sabzehgar and Mohammad Rasouli. 'Short Term Solar Irradiance Forecast based on Image Processing and Cloud Motion Detection'. In: *2019 IEEE Texas Power and Energy Conference (TPEC)*. 2019 IEEE Texas Power and Energy Conference (TPEC). Feb. 2019, pp. 1–6. DOI: 10.1109/TPEC.2019.8662134.

[16] Chuyao Luo, Xutao Li and Yunming Ye. 'PFST-LSTM: A SpatioTemporal LSTM Model With Pseudoflow Prediction for Precipitation Nowcasting'. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), pp. 843–857. ISSN: 2151-1535. DOI: 10.1109/JSTARS.2020.3040648.

[17] A. Ayet and P. Tandeo. 'Nowcasting solar irradiance using an analog method and geostationary satellite images'. In: *Solar Energy* 164 (1st Apr. 2018), pp. 301–315. ISSN: 0038-092X. DOI: 10.1016/j.solener.2018.02.068. URL: https://www.sciencedirect.com/science/article/pii/S0038092X18301993.

[18] S. Alessandrini et al. 'An analog ensemble for short-term probabilistic solar power forecast'. In: *Applied Energy* 157 (1st Nov. 2015), pp. 95–110. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2015.08.011. URL: https://www.sciencedirect.com/science/article/pii/S0306261915009368.

[19] Elena Collino and Dario Ronzio. 'Exploitation of a New Short-Term Multimodel Photovoltaic Power Forecasting Method in the Very Short-Term Horizon to Derive a Multi-Time Scale Forecasting System'. In: *Energies* 14.3 (3 Jan. 2021), p. 789. ISSN: 1996-1073. DOI: 10.3390/en14030789. URL: https://www.mdpi.com/1996-1073/14/3/789.

[20] Guido Cervone et al. 'Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble'. In: *Renewable Energy* 108 (1st Aug. 2017), pp. 274–286. ISSN: 0960-1481. DOI: 10.1016/j.renene.2017.02.052. URL: https://www.sciencedirect.com/science/article/pii/S0960148117301386.

[21] Dennis Knol et al. 'Deep Learning for Solar Irradiance Nowcasting: A Comparison of a Recurrent Neural Network and Two Traditional Methods'. In: *Computational Science – ICCS 2021*. Ed. by Maciej Paszynski et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021, pp. 309–322. ISBN: 978-3-030-77977-1. DOI: 10.1007/978-3-030-77977-1_24.

[22] Lasse Espeholt et al. 'Skillful Twelve Hour Precipitation Forecasts using Large Context Neural Networks'. 14th Nov. 2021. arXiv: 2111.07470 [physics]. URL: http://arxiv.org/abs/2111.07470.

[23] M. G. Schultz et al. 'Can deep learning beat numerical weather prediction?' In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379.2194 (5th Apr. 2021), p. 20200097. DOI: 10.1098/rsta.2020.0097. URL: https://royalsocietypublishing.org/doi/full/10.1098/rsta.2020.0097.

[24] Jesus Lago et al. 'Short-term forecasting of solar irradiance without local telemetry: A generalized model using satellite data'. In: *Solar Energy* 173 (1st Oct. 2018), pp. 566–577. ISSN: 0038-092X. DOI: 10.1016/j.solener.2018.07.050. URL: https://www.sciencedirect.com/science/article/pii/S0038092X18307138.

[25] Steven D. Miller et al. 'Short-term solar irradiance forecasting via satellite/model coupling'. In: *Solar Energy*. Advances in Solar Resource Assessment and Forecasting 168 (1st July 2018), pp. 102–117. ISSN: 0038-092X. DOI: 10.1016/j.solener.2017.11.049. URL: https://www.sciencedirect.com/science/article/pii/S0038092X17310435.

[26] Nicolas Holland et al. 'Solar and PV forecasting for large PV power plants using numerical weather models, satellite data and ground measurements'. In: *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*. 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC). June 2019, pp. 1609–1614. DOI: 10.1109/PVSC40753.2019.8980496.

[27] Johan Mathe et al. 'PVNet: A LRCN Architecture for Spatio-Temporal Photovoltaic PowerForecasting from Numerical Weather Prediction'. 20th Mar. 2020. arXiv: 1902.01453 [cs]. URL: http://arxiv.org/abs/1902.01453.

[28] Xiangyun Qing and Yugang Niu. 'Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM'. In: *Energy* 148 (Apr. 2018), pp. 461–468. ISSN: 03605442. DOI: 10.1016/j.energy.2018.01.177. URL: https://linkinghub.elsevier.com/retrieve/pii/S0360544218302056.

[29] Mohamed Abdel-Nasser, Karar Mahmoud and Matti Lehtonen. 'Reliable Solar Irradiance Forecasting Approach Based on Choquet Integral and Deep LSTMs'. In: *IEEE Transactions on Industrial Informatics* 17.3 (Mar. 2021), pp. 1873–1881. ISSN: 1941-0050. DOI: 10.1109/TII.2020.2996235.

[30] Cong Li et al. 'Hourly solar irradiance prediction using deep BiLSTM network'. In: *Earth Science Informatics* 14.1 (1st Mar. 2021), pp. 299–309. ISSN: 1865-0481. DOI: 10.1007/s12145-020-00511-3. URL: https://doi.org/10.1007/s12145-020-00511-3.

[31] Tian Peng et al. 'An integrated framework of Bi-directional long-short term memory (BiLSTM) based on sine cosine algorithm for hourly solar radiation forecasting'. In: *Energy* 221 (15th Apr. 2021), p. 119887. ISSN: 0360-5442. DOI: 10.1016/j.energy.2021.119887. URL: https://www.sciencedirect.com/science/article/pii/S0360544221001365.

[32] Sima Siami-Namini, Neda Tavakoli and Akbar Siami Namin. 'The Performance of LSTM and BiLSTM in Forecasting Time Series'. In: *2019 IEEE International Conference on Big Data (Big Data)*. 2019 IEEE International Conference on Big Data (Big Data). Dec. 2019, pp. 3285–3292. DOI: 10.1109/BigData47090.2019.9005997.

[33] Chiou-Jye Huang and Ping-Huan Kuo. 'Multiple-Input Deep Convolutional Neural Network Model for Short-Term Photovoltaic Power Forecasting'. In: *IEEE Access* 7 (2019), pp. 74822–74834. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2921238.

[34] Emilio Pérez et al. 'A deep learning model for intra-day forecasting of solar irradiance using satellite-based estimations in the vicinity of a PV power plant'. In: *Solar Energy* 218 (1st Apr. 2021), pp. 652–660. ISSN: 0038-092X. DOI: 10.1016/j.solener.2021.02.033. URL: https://www.sciencedirect.com/science/article/pii/S0038092X21001353.

[35] Haixiang Zang et al. 'Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning'. In: *International Journal of Electrical Power & Energy Systems* 118 (1st June 2020), p. 105790. ISSN: 0142-0615. DOI: 10.1016/j.ijepes.2019.105790. URL: https://www.sciencedirect.com/science/article/pii/S0142061519307409.

[36] Haixiang Zang et al. 'Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations'. In: *Renewable Energy* 160 (1st Nov. 2020), pp. 26–41. ISSN: 0960-1481. DOI: 10.1016/j.renene.2020.05.150. URL: https://www.sciencedirect.com/science/article/pii/S0960148120308557.

[37] Xingjian Shi et al. 'Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model'. 5th Oct. 2017. arXiv: 1706.03458 [cs]. URL: http://arxiv.org/abs/1706.03458.

[38] Sujan Ghimire et al. 'Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms'. In: *Applied Energy* 253 (1st Nov. 2019), p. 113541. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2019.113541. URL: https://www.sciencedirect.com/science/article/pii/S0306261919312152.

[39] Talha Ahmad Siddiqui, Samarth Bharadwaj and Shivkumar Kalyanaraman. 'A Deep Learning Approach to Solar-Irradiance Forecasting in Sky-Videos'. In: *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa Village, HI, USA: IEEE, Jan. 2019, pp. 2166–2174. ISBN: 978-1-72811-975-5. DOI: 10.1109/WACV.2019.00234. URL: https://ieeexplore.ieee.org/document/8659184/.

[40] Zhiyuan Si et al. 'Hybrid Solar Forecasting Method Using Satellite Visible Images and Modified Convolutional Neural Networks'. In: *IEEE Transactions on Industry Applications* 57.1 (Jan. 2021), pp. 5–16. ISSN: 1939-9367. DOI: 10.1109/TIA.2020.3028558.

[41] Ying-Yi Hong, John Joel F. Martinez and Arnel C. Fajardo. 'Day-Ahead Solar Irradiation Forecasting Utilizing Gramian Angular Field and Convolutional Long Short-Term Memory'. In: *IEEE Access* 8 (2020), pp. 18741–18753. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2967900.

[42] Quentin Paletta, Guillaume Arbod and Joan Lasenby. 'Benchmarking of deep learning irradiance forecasting models from sky images – An in-depth analysis'. In: *Solar Energy* 224 (1st Aug. 2021), pp. 855–867. ISSN: 0038-092X. DOI: 10.1016/j.solener.2021.05.056. URL: https://www.sciencedirect.com/science/article/pii/S0038092X21004266.

[43] Quentin Paletta et al. 'ECLIPSE: Envisioning CLoud Induced Perturbations in Solar Energy'. In: *Applied Energy* 326 (15th Nov. 2022), p. 119924. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2022.119924. URL: https://www.sciencedirect.com/science/article/pii/S0306261922011813.

[44] Xingjian Shi et al. 'Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting'. 19th Sept. 2015. arXiv: 1506.04214 [cs]. URL: http://arxiv.org/abs/1506.04214.

[45] Casper Kaae Sønderby et al. 'MetNet: A Neural Weather Model for Precipitation Forecasting'. 30th Mar. 2020. arXiv: 2003.12140 [physics, stat]. URL: http://arxiv.org/abs/2003.12140.

[46] Quang-Khai Tran and Sa-kwang Song. 'Computer Vision in Precipitation Nowcasting: Applying Image Quality Assessment Metrics for Training Deep Neural Networks'. In: *Atmosphere* 10.5 (2nd May 2019), p. 244. ISSN: 2073-4433. DOI: 10.3390/atmos10050244. URL: https://www.mdpi.com/2073-4433/10/5/244.

[47] Jong-Min Yeom et al. 'Spatial mapping of short-term solar radiation prediction incorporating geostationary satellite images coupled with deep convolutional LSTM networks for South Korea'. In: *Environmental Research Letters* 15.9 (Aug. 2020), p. 094025. ISSN: 1748-9326. DOI: 10.1088/1748-9326/ab9467. URL: https://doi.org/10.1088/1748-9326/ab9467.

[48] Nicolas Ballas et al. *Delving Deeper into Convolutional Networks for Learning Video Representations*. 1st Mar. 2016. DOI: 10.48550/arXiv.1511.06432. arXiv: 1511.06432 [cs]. URL: http://arxiv.org/abs/1511.06432. preprint.

[49] Yunbo Wang et al. 'PredRNN: Recurrent Neural Networks for Predictive Learning using Spatiotemporal LSTMs'. In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper/2017/hash/e5f6ad6ce374177eef023bf5d0c018b6-Abstract.html.

[50] Yunbo Wang et al. 'Eidetic 3D LSTM: A Model for Video Prediction and Beyond'. In: International Conference on Learning Representations. 27th Sept. 2018. URL: https://openreview.net/forum?id=B1lKS2AqtX.

[51] Dukhwan Yu et al. 'Forecasting Photovoltaic Power Generation Using Satellite Images'. In: *Energies* 13.24 (24 Jan. 2020), p. 6603. ISSN: 1996-1073. DOI: 10.3390/en13246603. URL: https://www.mdpi.com/1996-1073/13/24/6603.

[52] Vincent Le Guen and Nicolas Thome. 'A Deep Physical Model for Solar Irradiance Forecasting with Fisheye Images'. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA: IEEE, June 2020, pp. 2685–2688. ISBN: 978-1-72819-360-1. DOI: 10.1109/CVPRW50498.2020.00323. URL: https://ieeexplore.ieee.org/document/9150809/.

[53] Yunbo Wang et al. 'Memory in Memory: A Predictive Neural Network for Learning Higher-Order Non-Stationarity From Spatiotemporal Dynamics'. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019, pp. 9154–9162. URL: https://openaccess.thecvf.com/content_CVPR_2019/

html/Wang_Memory_in_Memory_A_Predictive_Neural_Network_for_Learning_Higher-Order_CVPR_2019_paper.html.

[54] Ahmad Maroof Karimi et al. 'Spatiotemporal Graph Neural Network for Performance Prediction of Photovoltaic Power Systems'. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 35.17 (17 18th May 2021), pp. 15323–15330. ISSN: 2374-3468. DOI: 10.1609/aaai.v35i17.17799. URL: https://ojs.aaai.org/index.php/AAAI/article/view/17799.

[55] Songjian Chai et al. 'A Robust Spatiotemporal Forecasting Framework for Photovoltaic Generation'. In: *IEEE Transactions on Smart Grid* 11.6 (Nov. 2020), pp. 5370–5382. ISSN: 1949-3061. DOI: 10.1109/TSG.2020.3006085.

[56] Ignacio-Iker Prado-Rujas et al. 'A Flexible and Robust Deep Learning-Based System for Solar Irradiance Forecasting'. In: *IEEE Access* 9 (2021), pp. 12348–12361. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3051839.

[57] *LSA_207_MDSSFTD_PUM_v1.0_issue3.pdf*. LSASAF Nextcloud. 2020. URL: https://nextcloud.lsasvcs.ipma.pt/s/ytGYTRFjB9EzwBp.

[58] Dénes Fekete. *Basic and other measurements of radiation at station Budapest (2019-06 et seq)*. PANGAEA, 15th July 2022. DOI: 10.1594/PANGAEA.946379. URL: https://doi.pangaea.de/10.1594/PANGAEA.946379.

[59] Laurent Vuilleumier and Alain Heimo. *Basic and other measurements of radiation at station Payerne (1992-10 et seq)*. PANGAEA, 15th July 2022. DOI: 10.1594/PANGAEA.946384. URL: https://doi.pangaea.de/10.1594/PANGAEA.946384.

[60] Marc Olefs. *Basic measurements of radiation at station Sonnblick (2013-01 et seq)*. PANGAEA, 15th July 2022. DOI: 10.1594/PANGAEA.946389. URL: https://doi.pangaea.de/10.1594/PANGAEA.946389.

[61] Christoph Marty et al. 'Altitude dependence of surface radiation fluxes and cloud forcing in the alps: Results from the alpine surface radiation budget network'. In: *Theoretical and Applied Climatology* 72 (9th Jan. 2002), pp. 137–155. DOI: 10.1007/s007040200019.

[62] Chi Wai Chow, Serge Belongie and Jan Kleissl. 'Cloud motion and stability estimation for intra-hour solar forecasting'. In: *Solar Energy* 115 (May 2015), pp. 645–655. ISSN: 0038092X. DOI: 10.1016/j.solener.2015.03.030. URL: https://linkinghub.elsevier.com/retrieve/pii/S0038092X15001565.

[63] Fitsum A. Reda et al. 'SDC-Net: Video Prediction Using Spatially-Displaced Convolution'. In: *Computer Vision – ECCV 2018*. Ed. by Vittorio Ferrari et al. Vol. 11211. Cham: Springer International Publishing, 2018, pp. 747–763. DOI: 10.1007/978-3-030-01234-2_44. URL: https://link.springer.com/10.1007/978-3-030-01234-2_44.

[64] Soorok Ryu et al. 'Improved rainfall nowcasting using Burgers' equation'. In: *Journal of Hydrology* 581 (1st Feb. 2020), p. 124140. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2019.124140. URL: https://www.sciencedirect.com/science/article/pii/S0022169419308753.

[65] Vincent Le Guen and Nicolas Thome. 'Disentangling Physical Dynamics From Unknown Factors for Unsupervised Video Prediction'. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 2020, pp. 11471–11481. DOI: 10.1109/CVPR42600.2020.01149.

[66] Zichao Long et al. 'PDE-Net: Learning PDEs from Data'. In: *Proceedings of the 35th International Conference on Machine Learning*. International Conference on Machine Learning. PMLR, 3rd July 2018, pp. 3208–3216. URL: https://proceedings.mlr.press/v80/long18a.html.

[67] Vincent Le Guen and Nicolas Thome. *Disentangling Physical Dynamics from Unknown Factors for Unsupervised Video Prediction Supplementary Material*. https://openaccess.thecvf.com/content_CVPR_2020/supplemental/Le_Guen_Disentangling_Physical_Dynamics_CVPR_2020_supplemental.pdf. 2020.

[68] Marek Śmieja et al. 'Processing of missing data by neural networks'. In: *Advances in Neural Information Processing Systems*. Vol. 31. Curran Associates, Inc., 2018. URL: https://proceedings.neurips.cc/paper/2018/hash/411ae1bf081d1674ca6091f8c59a266f-Abstract.html.

[69] Shuo Shan et al. 'A deep-learning based solar irradiance forecast using missing data'. In: *IET Renewable Power Generation* 16.7 (2022), pp. 1462–1473. ISSN: 1752-1424. DOI: 10.1049/rpg2.12408. URL: https://onlinelibrary.wiley.com/doi/abs/10.1049/rpg2.12408.

[70] Wei Liu, Chao Ren and Yan Xu. 'Missing-Data Tolerant Hybrid Learning Method for Solar Power Forecasting'. In: *IEEE Transactions on Sustainable Energy* 13.3 (July 2022), pp. 1843–1852. ISSN: 1949-3037. DOI: 10.1109/TSTE.2022.3173147.

[71] Jan Chorowski et al. 'Unsupervised Speech Representation Learning Using WaveNet Autoencoders'. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 27.12 (Dec. 2019), pp. 2041–2053. ISSN: 2329-9304. DOI: 10.1109/TASLP.2019.2938863.

[72] Golnaz Ghiasi, Tsung-Yi Lin and Quoc V Le. 'DropBlock: A regularization method for convolutional networks'. In: *Advances in Neural Information Processing Systems*. Vol. 31. Curran Associates, Inc., 2018. URL: https://proceedings.neurips.cc/paper/2018/hash/7edcfb2d8f6a659ef4cd1e6c9b6d7079-Abstract.html.

[73] David Krueger et al. *Zoneout: Regularizing RNNs by Randomly Preserving Hidden Activations*. 22nd Sept. 2017. arXiv: 1606.01305 [cs]. URL: http://arxiv.org/abs/1606.01305. preprint.

[74] Yarin Gal and Zoubin Ghahramani. 'A Theoretically Grounded Application of Dropout in Recurrent Neural Networks'. In: *Advances in Neural Information Processing Systems*. Vol. 29. Curran Associates, Inc., 2016. URL: `https://proceedings.neurips.cc/paper/2016/hash/076a0c97d09cf1a0ec3e19c7f2529f2b-Abstract.html`.

[75] Taesup Moon et al. 'RNNDROP: A novel dropout for RNNS in ASR'. In: *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*. 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU). Dec. 2015, pp. 65–70. DOI: `10.1109/ASRU.2015.7404775`.

[76] Stanislau Semeniuta, Aliaksei Severyn and Erhardt Barth. *Recurrent Dropout without Memory Loss*. 5th Aug. 2016. arXiv: `1603.05118 [cs]`. URL: `http://arxiv.org/abs/1603.05118`. preprint.

[77] Sébastien de Blois et al. 'Input Dropout for Spatially Aligned Modalities'. In: *2020 IEEE International Conference on Image Processing (ICIP)*. 2020 IEEE International Conference on Image Processing (ICIP). Oct. 2020, pp. 733–737. DOI: `10.1109/ICIP40778.2020.9190858`.

[78] Concepción Crespo Turrado et al. 'Missing Data Imputation of Solar Radiation Data under Different Atmospheric Conditions'. In: *Sensors* 14.11 (11 Nov. 2014), pp. 20382–20399. ISSN: 1424-8220. DOI: `10.3390/s141120382`. URL: `https://www.mdpi.com/1424-8220/14/11/20382`.

[79] Wenjie Zhang et al. 'SolarGAN: Multivariate Solar Data Imputation Using Generative Adversarial Network'. In: *IEEE Transactions on Sustainable Energy* 12.1 (Jan. 2021), pp. 743–746. ISSN: 1949-3037. DOI: `10.1109/TSTE.2020.3004751`.

[80] Taeyoung Kim, Woong Ko and Jinho Kim. 'Analysis and Impact Evaluation of Missing Data Imputation in Day-ahead PV Generation Forecasting'. In: *Applied Sciences* 9.1 (1 Jan. 2019), p. 204. ISSN: 2076-3417. DOI: `10.3390/app9010204`. URL: `https://www.mdpi.com/2076-3417/9/1/204`.

[81] Noor Bariah Mohamad, Boon-Han Lim and An-Chow Lai. 'Imputation of Missing Values for Solar Irradiance Data under Different Weathers using Univariate Methods'. In: *IOP Conference Series: Earth and Environmental Science* 721.1 (Apr. 2021), p. 012004. ISSN: 1755-1315. DOI: `10.1088/1755-1315/721/1/012004`. URL: `https://dx.doi.org/10.1088/1755-1315/721/1/012004`.

[82] Zhou Wang et al. 'Image quality assessment: from error visibility to structural similarity'. In: *IEEE Transactions on Image Processing* 13.4 (Apr. 2004), pp. 600–612. ISSN: 1941-0042. DOI: `10.1109/TIP.2003.819861`.

[83] Weiqing Wang et al. 'An Ultra-short-term Forecasting Model for High-resolution Solar Irradiance Based on SOM and Deep Learning Algorithm'. In: *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*. 2019 IEEE Sustainable Power and Energy Conference (iSPEC). Nov. 2019, pp. 1090–1095. DOI: `10.1109/iSPEC48194.2019.8974983`.

[84] Yunbo Wang et al. 'PredRNN++: Towards A Resolution of the Deep-in-Time Dilemma in Spatiotemporal Predictive Learning'. 18th Nov. 2018. arXiv: 1804.06300 [cs, stat]. URL: http://arxiv.org/abs/1804.06300.

[85] J. Antonanzas et al. 'Review of photovoltaic power forecasting'. In: *Solar Energy* 136 (15th Oct. 2016), pp. 78–111. ISSN: 0038-092X. DOI: 10.1016/j.solener.2016.06.069. URL: https://www.sciencedirect.com/science/article/pii/S0038092X1630250X.

[86] Suman Ravuri et al. 'Skilful precipitation nowcasting using deep generative models of radar'. In: *Nature* 597.7878 (7878 Sept. 2021), pp. 672–677. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03854-z. URL: https://www.nature.com/articles/s41586-021-03854-z.

[87] Yunbo Wang et al. 'PredRNN: A Recurrent Neural Network for Spatiotemporal Predictive Learning'. 26th Aug. 2021. arXiv: 2103.09504 [cs]. URL: http://arxiv.org/abs/2103.09504.

A. Appendix

A.1. Supplement

A.1.1. Dataset ablation study

To validate the selection of additional temporal and spatial features used in training, IrradianceNet was trained with all possible additional feature configurations. This includes the single solar irradiance feature DSSF, the inclusion of only the spatial feature map of elevation and the use of only the cyclically encoded temporal information derived from the DSSF data.

Included features	RMSE ↓	MAE ↓	SSIM ↑
DSSF	77.43	39.86	0.6889
DSSF + elevation map	77.13	40.02	0.6888
DSSF + date and time	73.66	37.93	0.7043
Full featureset	**72.11**	**37.46**	**0.7078**

Table A.1.: Error metrics for different datasets trained and tested on IrradianceNet. Results were filtered by forecasted hours between 5:00 and 21:00.
[1]

The assumption that benefits in terms of forecast accuracy can be reached by the inclusion of the elevation map, as well as encoded date and time holds. The full dataset as used in the comparison provides the best results, as the metrics in Table A.1 show. The date and time information provides a clearer benefit over the sole inclusion of the elevation map, while still bigger benefits can be achieved by combining both. Figure A.1 explores a different aspect of this in terms of forecast errors by hour of day. As DSSF contains only solar irradiance data, no dynamic nightly information is provided, leading to an increase in error for the first hours of day if no additional temporal information is present to contextualize the lack of irradiance.

[1] The temporal evaluation range was here selected wider than in the main evaluation in Chapter 6, as the accuracy of predictions of irradiance during the early morning and early evening hours was considered relevant to incorporate.

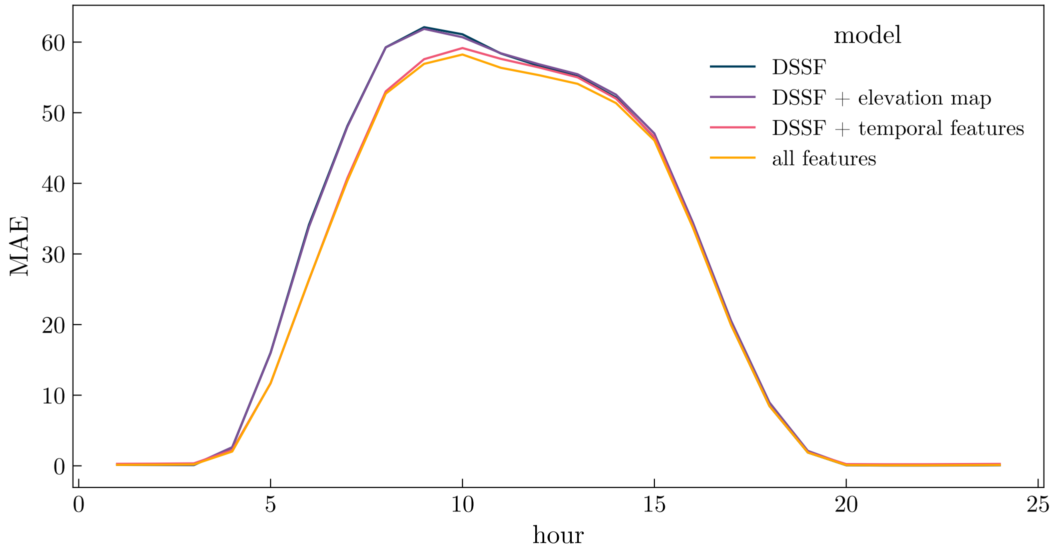


Figure A.1.: MAE over hour of day for different datasets on IrradianceNet.

Missing timesteps and dataset feature space

In the context of missing data, meaning here missing timesteps in the DSSF feature maps, the differences are more pronounced than when tested without missing data.

Dataset features	RMSE ↓	MAE ↓	SSIM ↑
DSSF	<u>150.61</u>	<u>74.89</u>	0.6334
DSSF + elevation map	141.19	70.36	<u>0.631</u>
DSSF + date and time	129.95	62.18	0.6572
Full featureset	**123.23**	**59.64**	**0.6666**

Table A.2.: Error metrics for different datasets trained and tested on IrradianceNet, with 10% missing data applied on testing. Results were filtered by forecast hours between 5:00 and 21:00.

Overall, date and time seem to be relevant to include and have benefits in both contexts, while the elevation map alone leads to improvements in the pixel-wise errors in the presence of missing data, but not the structural similarity. In both instances, with full input data as well as with missing data, the chosen dataset configuration is superior to a sub-selection of features.

A.1.2. Tuning of imperfect data ratio in timestep-dropout

A limited test on a smaller area, of 64 by 64 grid cells, keeping the same temporal forecasting range, was utilized to reveal a potentially optimal choice of ratio from real to missing value-grids in the training loop. Training was also limited to 60 epochs which was mainly in the interest of time but also mostly achieved convergence before or when reaching the epoch limit, through the smaller spatial range. As before, each ratio signifies a random chance for each individual irradiance feature of an input timestep to be set to zero. A range of ratios or probabilities was tested to find a value-candidate for further training on the full-scale dataset. There is no general case to define this optimal value, as other data sources may have a lot higher missing value risk than the selected limited LSA-SAF dataset, and the practical value ideal for training may not correspond to any real value. To determine an optimized use in the context of the present parameters and data, a selection of missing value ratios was defined as [0.1,0.2,0.3,0.4], which were separately applied in training for timestep-dropout.

The trained models were then used to forecast test results for a range of missing data ratios. To monitor changes in general error, with penalizing effect on larger deviations, RMSE was used as a first test metrics and complemented with SSIM to highlight structural deviation changes.

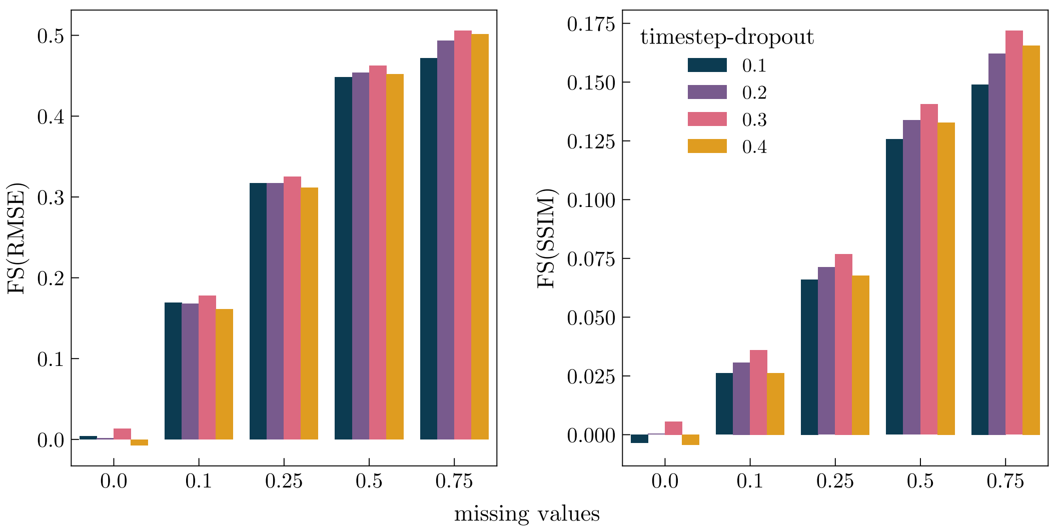


Figure A.2.: Mean Forecast Skill of models trained with timestep-dropout relative to baseline without timestep-dropout

Figure A.2 shows the derived mean Forecast Skills (FS)[2] between the timestep dropout trained models and the baseline models trained regularly. The results show significant improvements to RMSE as well SSIM that seem to correlate with the amount of missing data in the input. While all ratios lead to improvements, there's a clear indication that 0.3 leads to the best results, especially concerning SSIM.

An additional result of the limited test is a hint that timestep-dropout could have a

[2]For SSIM the Forecast skill score calculation was inverted, to adapt to the metrics inverse optimum (1.0) compared to distance metrics.

beneficial effect on the overall predictive performance in some cases. This is visible in the case of zero missing values in the input data when a timestep-dropout of 0.3 was applied, accompanied with a counter indication for too high (0.4) and for SSIM also in the case of too low timestep-dropout ratios. These effect may however not be consistent and dependant on specific fine-tuning of the timestep-dropout ratio.

The above aggregate data is a simplified approach as an alternative to selecting different ratios per model. This might mean that some models may benefit more from this specific ratio. A general ideal selection cannot be achieved, as a repetition of these tests on the full-scale dataset would take multiple months. Within the here applied limitations, there's a relatively clear indication that the ideal value should be below 0.4 and may be above 0.2. The peak ratio of 0.3, across RMSE and SSIM, is selected as an approximation of the ideal value, that could be further narrowed down or individualized to the specific models. As real-world data shows occurrences of consecutive missing values, within the small percentage of missing values, a parameter for the probability of the removal of timesteps following an introduced missing value was considered and tested but lead to too great a variation in the inputs and unstable training.

A.1.3. Analog Search method

Spatiotemporal deep learning models are only one way to forecast solar irradiance in the nowcasting range and should also be compared to other techniques. As solar irradiance forecasts are individually often only of interest for single local PV-energy installations, a technique with spatiotemporal input and single site prediction was selected for additional comparisons. The method selected here is an Analog Search method, which was implemented in a previous project, closely based[3] on the corresponding publication by Ayet and Tandeo [17].

Methods and models focussing on a single geographical site instead of an area could have some advantages over the spatiotemporal models discussed above, as they are able to aggregate more of the spatial context into a single local prediction. Conceptually the analog search method is based on the matching of an observation to past analogous situations, that share a selection of features within a predefined sufficient extent and can therefore provide temporal successors to inform a linear model to predict potential future values for a given observation.

The Analog Search applied here, uses the LSA-SAF dataset in its full geographical extent, with historical data from the years 2016-2020. It derives cloud index and clear sky irradiance directly from the dataset. Both features are calculated from the irradiance data, with clear sky irradiance being the maximum within a temporal interval of three months

[3] As this implementation was not directly verified against the original implementation, it is only assumed to perform similarly, not identically. A relatively close match to the published error results for BSRN stations contained in both subspaces, the one selected by Ayet and Tandeo and this and a previous test with a different dataset, led to the assumption that the implementation is viable.

around the observation time and the cloud index being a ratio of the actual values at a given timestep and the clear sky irradiance from the corresponding interval. Based on the calculated values, four features[4] are extracted from a connected correlation-mask around the point of interest, with the 256x64 spatial feature space. The results are then used to perform a k-nearest neighbour search within the range of three months and three hours around the input date to find k historical analogs. To alleviate suboptimal spatial analog matching in the reduced feature space, an optimization of the found cloud-pattern-analogs' spatial correlation to the input data is performed by translating the analogs to their optimal fit to the observation. Finally following dimensionality reduction of the analog ensemble via Principal Component Analysis (PCA), a Linear Regression is fitted on the analogs and their temporal successors, which then serves as an estimator applied to the observation's principial components. The length of future steps to be estimated is thereby a dynamically changeable parameter, for the comparison it will be fixed to 12 timesteps in 15-min increments, corresponding to three hours. This may impact performance, as the original lead times were between 1 and 6 hours, with hourly step-size. The original method also supported probabilistic output, which was here omitted and used in its deterministic configuration as the other models. A statistical downscaling approach as a post-processing step utilizing local ground data was also applied in the original publication to enhance forecast quality. But as this requires additional local historical datasets, it is considered an extension beyond the scope of this comparison, which is focused on the viability of deriving forecasts from large-scale satellite data.

While there are some detriments to using the analog-search algorithm, it can still be a viable option in cases where sufficient historical data is available and training efforts in advance are not possible. As and additional benefit, an input data source with reduced temporal resolution could be used as an input source with a historical temporal resolution of more frequent timesteps, warranting that the captured feature is similar enough. As the dataset can be switched to a different geographical and climatic region, it has the general benefit of flexibility in contrast to neural network solutions. The analog search is also intrinsically more robust to missing timesteps, as it does not require or is able to process more than a single timestep as input. The limitation of only one timestep makes it per default more resilient to missing timesteps but also limits its next output to the last valid captured timestep.

[4] Cloud fraction, cloud spread, clear sky intensity and cloud intensity. These features are deemed viable indicators for an analog match, based on the assumption that "histograms discriminate between different cloud regimes" [17].

A.1.4. Tables with timestep-dropout for local forecasting

Model	MAE ↓ Budapest	MAE ↓ Payerne	MAE ↓ Sonnblick	RMSE ↓ Budapest	RMSE ↓ Payerne	RMSE ↓ Sonnblick
Persistence(GHI)	148.43	147.59	159.55	202.49	206.92	213.38
Analog Search	100.15	90.45	123.83	151.87	133.32	181.49
IrradianceNet	61.00	65.88	91.94	106.77	**106.05**	141.62
with timestep-dropout	61.18	66.14	92.90	107.18	107.70	142.05
PFST-LSTM	62.13	66.12	87.48	108.65	106.80	**135.64**
with timestep-dropout	61.30	66.23	**87.38**	107.17	106.80	**135.64**
PhyDNet-SST	61.41	67.23	90.21	107.71	108.55	140.81
with timestep-dropout	61.37	66.04	88.42	107.84	106.98	137.95
PhyDNet-dual	64.68	68.24	89.91	111.20	108.58	140.31
with timestep-dropout	62.92	67.53	88.39	109.05	108.45	138.12
IrradPhyDNet	61.43	67.10	89.35	107.09	108.65	139.20
with timestep-dropout	**60.92**	**65.72**	91.43	**106.52**	106.54	141.08

Table A.3.: Results on BSRN site GHI

Model	MAE ↓ Budapest	MAE ↓ Payerne	MAE ↓ Sonnblick	MAE ↓ Hohe Warte	RMSE ↓ Budapest	RMSE ↓ Payerne	RMSE ↓ Sonnblick	RMSE ↓ Hohe Warte
Persistence	144.06	151.10	160.09	142.7	194.32	208.19	214.74	195.12
Analog Search	86.57	70.95	81.87	109.74	132.42	104.66	121.23	159.73
IrradianceNet	38.85	43.11	46.43	56.12	**74.26**	79.46	82.64	90.14
with timestep-dropout	39.62	43.56	46.02	55.35	75.02	80.23	83.58	88.81
PFST-LSTM	39.30	42.83	47.08	57.05	76.20	79.79	84.36	90.29
with timestep-dropout	38.46	42.24	46.27	57.32	74.42	78.03	82.65	90.20
PhyDNet-SST	**38.29**	42.15	46.64	**54.97**	74.90	78.99	84.89	89.68
with timestep-dropout	38.43	42.75	46.60	55.41	74.65	78.99	83.52	89.99
PhyDNet-dual	42.30	45.19	49.59	62.05	78.34	81.27	86.15	97.70
with timestep-dropout	41.01	45.11	48.44	60.02	76.85	81.96	86.17	94.66
IrradPhyDNet	38.55	42.84	46.50	55.32	74.71	78.95	83.85	**89.24**
with timestep-dropout	38.81	**41.89**	**45.04**	55.65	74.62	**77.34**	**81.30**	89.85

Table A.4.: Results on closest grid cells