KU LEUVEN

FACULTEIT INGENIEURSWETENSCHAPPEN

Forecasting residential PV power using transfer learning with synthetic data

Robbe Vander Eeckt

Thesis submitted for the degree of
Master of Science in Engineering:
Energy

Thesis supervisors:
Prof. dr. ir. Johan Driesen
Prof. dr. ir. Hussain Syed Kazmi

Assessors:
Prof. dr. ir. J. Poortmans
ir. R. Smets

Mentor:
J. Depoortere

Academic year 2024 – 2025

Preface

I would like to express my gratitude to the many individuals who have supported me throughout this challenging journey. First and foremost, my deepest thanks go to Joris Depoortere, whose invaluable guidance was instrumental in bringing this thesis to a successful conclusion. I am also immensely grateful to Aiko for patiently listening to me and providing encouragement when I was stuck. Finally, I would like to thank my family for their unwavering support, which enabled me to complete this work.

Robbe Vander Eeckt

Contents

Abstract

Solar power forecasting is essential for optimizing energy use in residential households. Machine learning models are promising for this power forecasting because they can capture its non-linear characteristics. Nonetheless, these models require a significant amount of data that is unavailable for new installations. To overcome the limited data availability, this thesis proposes a transfer learning model using Long Short-Term Memory (LSTM) networks trained on synthetic photovoltaic (PV) generation provided by the Photovoltaic Geographical Information System. Using the metadata about a PV installation, this synthetic data simulates the past production of the new PV installation. The model utilizes Numerical Weather Predictions (NWP) and autoregressive covariates. This proposed model is compared to benchmarks, including models trained only on sites' actual PV power, physical models, and TL models with no weather covariates.

The research investigates the effect of physics-informed variables on the accuracy of transfer learning. Furthermore, it examines the usage of reanalysis data to train with synthetic PV data due to the low accessibility of historical Numerical Weather Prediction output. Walk-forward validation is employed for forecasting the actual PV power to simulate real-life conditions and the impact of increasing target data.

The results demonstrate that models trained with historical Numerical Weather Prediction data achieve higher zero-shot forecasting accuracy. Contrary to expectations, including physics-informed variables did not enhance performance; in fact, it showed a slight decrease. Additionally, models trained on reanalysis data catch up with those trained on historical NWP data once limited target data becomes available.

Discussion highlights include the impact of Storm Darcy in February 2021, which caused instability in machine learning models and the suitability of other ML models. The proposed LSTM-based transfer learning model can provide accurate forecasts even with no or limited actual PV power data, proving its potential for practical applications in solar power forecasting for residential households.

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations

ANN	Artificial Neural Network
ARIMA	Auto-Regressive Integrated Moving Average
biLSTM	Bidirectional LSTM
CNN	Convolution Neural Network
CPTEC	Centro de Previsão do Tempo e Estudos Climáticos
CSI	Clear Sky Index
CSP	Clear-Sky Power
DHI	Diffuse Horizontal Irradiance
DNI	Direct Normal Irradiance
DNN	Deep Neural Network
DWD	Deutscher Wetterdienst
ECMWF	European Centre for Medium-Range Weather Forecasts
ENS	Ensemble Forecast Suite
ERA5	Fifth generation of European Reanalysis
GB	Gradient Boosting
GHG	Greenhouse Gas
GHI	Global Horizontal Irradiance
GRIB	Gridded Binary
GRU	Gated Recurrent Unit
HP	Hyperparameters
HRES	High Resolution Forecasts
ICON	Icosahedral Nonhydrostatic
IFS	Integrated Forecast System
LCOE	Levelized Cost Of Electricity
LSTM	Long Short-Term Memory

MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Square Error
MSL	Mean Sea Level
NWP	Numerical Weather Prediction
PV	Photovoltaics
PVGIS	Photovoltaic Geographical Information System
PVPF	Photovoltaic Power Forecast
RA	Reanalysis
RES	Renewable Energy Source
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SVM	Support Vector Machine
SVR	Support Vector Regression
TIGGE	The International Grand Global Ensemble
TL	Transfer Learning
UM	Unified Model

Symbols

GHI	Global Horizontal Irradiance
DNI	Direct Normal Irradiance
DHI	Direct Horizontal Irradiance
ϕ	Tilt angle of surface
R_b	geometric relation of tilted and horizontal beam irradiance
F_i	brightness coefficients
γ	Incident angle of irradiance
I_{et}	Extraterrestrial irradiance
K_{nc}	Correlation coefficient of direct irradiance and airmass
A, B, C	Correlations of direct irradiance and clearness index
T_{PV}	Temperature of PV panel
h	Altitude
$\hat{P}(t+k)$	Forecasted PV output at timestep k
f_t	Forget gate
i_t	Input gate
$\tilde{c}_t$	New cell state
c_t	Cell state

o_t	Output gate
h_t	Output of LSTM cell
x_t	Input at time t
W_n	Weight matrices
b_n	Bias vectors
σ	Sigmoid activation function
tanh	Hyperbolic tangent activation function
$L(.)$	Loss function
w	Vecor with the model's weights
λ	Weight decay
β	Learning rate
k	Number of folds
MSE_i	MSE for fold i
D	Period
P_{CS}	Clear sky power
k^*	Clear sky PV index
ξ_h	D-lag autocorrelation coefficient
m	Slope
R^2	Coefficient of determination
SS	Skill score
τ_i	Difference Lagrangian multipliers
$\mathcal{K}$	Kernel function for SVM
c	constant linear kernel SVM
T	Temperature
p	Pressure at MSL
q	Specific humidity
V	Wind speed
α	Wind Direction
C	Total Cloud Cover
P_{24h}	24-h lagged PV power
M	Month of the year
H	Hour of the day
u	Horizontal wind velocity
v	vertical wind velocity
θ	Solar zenith angle
X_{mm}	Min-max scaled variable

Chapter 1

Introduction

As a result of greenhouse gas (GHG) emissions, mainly CO_2 and CH_4, the world is already 1.1 °C warmer in 2011-2020 compared to 1850-1900 [1]. However, energy-related GHG emissions reached an all-time high at 41.3 Gt CO_2-eq in 2022 [2], accounting for 71.95 % of total global GHG emission [3]. Furthermore, the electrical power system is responsible for 35.35 % of these energy-related emissions. With the electricity demand expected to rise further [4], the carbon intensity of electricity production should be reduced and eventually go to zero.

Renewable energy sources (RES) technologies have risen significantly over the last few years to answer this challenge. As no fuel-based resources are necessary, RES eliminates GHG emissions and fuel costs. The absence of fuel cost results in a Levelised Cost Of Electricity (LCOE) comparable to or lower than the LCOE of conventional, fossil-based technologies [5]. Due to increasing fossil fuel and electricity prices, the global installed power of residential photovoltaic (PV) installations has risen from 39 GW in 2015 to 195 GW in 2022 [6] and is expected to have risen further in 2023. However, these distributed PV installations also challenge the existing power system as reports about PV installations shutting down have tripled in Belgium [7].

To address these issues of uncontrolled injection of PV panels and residential electrification with heat pumps and electric vehicles, power system operators want to incentivize more flexible use of electricity [8]. One option is to offer dynamic contracts, which consist of hourly electric prices coupled with the electricity price on day-ahead spot markets. Furthermore, grid operators in Flanders and Germany have introduced capacity-based tariffs to limit power consumption during peak hours [9]. A capacity-based tariff obligates the consumer to pay a fee proportional to their maximum offtake over a time interval. With the dynamic contract and new tariff components, a Home Energy Management System can schedule flexible appliances to minimize the overall cost for the user and help the grid. Additionally, energy communities, which are groups of consumers who join together on energy-related matters, are becoming more popular [10]. It encompasses peer-to-peer selling of abundant solar energy. All these use cases will become increasingly important when society moves to a net-zero society. Nevertheless, their success relies on accurate solar power forecasting.

1.1 Solar power and forecasting

Solar power forecasting is essential for efficient integration in the power system. Consequently, this scientific domain has seen a vast increase in publications over the last years [11]. Nevertheless, these publications mainly focus on forecasting for utility-scale solar PV installations. Studies like [12], [13], [14] and others all develop models for PV installations with a peak installed capacity over 20 kW. *Leloux et al.* examined residential PV installations in France, which they defined as installations under 10 kW. This focus could be due to a higher value of accurate forecasts for utility-scale PV installations as these plant operators must participate in the day-ahead electricity market. Nonetheless, the smaller size of residential installations leads to a higher variability in irradiance when a cloud passes by. This variability is aggregated in utility-scale PV systems due to their bigger size. This makes it more challenging to forecast the PV output of residential PV installations.

Moreover, mentions of residential PV installations are mainly in studies done on an aggregated level, where the net energy is of interest for distribution grid operators [15]. Consequently, the assessment of solar power forecasting in residential PV installations has not been thoroughly studied. A notable exception is [16], which studies solar power forecasting for two residential PV installations for 12 h-ahead forecasting. Due to the increasing electrification in households with more flexible appliances, accurate solar power forecasting for individual household PV installations has become increasingly crucial.

1.2 Machine Learning

Machine learning (ML) models show great promise in the forecasting domain, such as economic forecasting [17], weather forecasting [18], and electrical load forecasting [19]. Consequently, the solar forecasting domain shows a considerable interest in machine learning [20]. Furthermore, various papers, like [21], [13], and[22], report increased accuracy when ML methods are utilized to forecast PV power due to the ability to model non-linear characteristics. An endless amount of possibilities exist for ML models for time series forecasting. One of these ML models is the Long short-term memory (LSTM) architecture, which is a popular choice in the literature [23], [24] and a competitive algorithm in solar power forecasting [25], [26], [27].

The LSTM is part of the deep learning algorithms. These algorithms need a significant amount of data to generalize well. There is a current hiatus in research where models are trained on data for a minimum of a year to represent the seasonality component accurately. However, this data is not yet present for new installations, which limits the usability of these deep learning algorithms in those situations. Additionally, physics-informed features are shown to increase the PV forecasts of ML models [28], [29].

1.3 Research Objectives

Solar forecasting with ML has become an important research domain. However, the focus is mainly on utility-scale PV systems, as accurate forecasts are needed to participate efficiently in the energy markets. Furthermore, many research papers use a significant amount of training data of at least a year, which is not available for new installations. Additionally, if the proposed models use weather covariates to forecast the power, historical outputs of Numerical Weather Prediction (NWP) models are necessary. These are not easily accessible. Therefore, historical weather data, due to higher accessibility, can be more attractive to train on. The current literature does not mention how this affects forecasting performance.

This thesis aims to resolve some of these remarks. It proposes an LSTM architecture that performs day-ahead forecasts for residential PV installation, utilizing predictions of weather variables and historical PV generation as input features. A transfer learning (TL) approach is proposed to overcome limited data availability. The approach starts with training a base model on synthetic PV data generated via a tool of the Photovoltaic Geographical Information System (PVGIS) [30].

First, this thesis investigates the effect of adding physics-based information to the model in terms of performance and training time. Second, this thesis studies the potential of a TL model trained on weather reanalysis data instead of historical NWP runs. The resulting models are compared to relevant benchmarks. Finally, a structured analysis of the results is presented to understand the obtained results better.

Chapter 2

Literature Review

This chapter aims to present the reader with state-of-the-art research in solar power forecasting using machine learning. It starts with an introduction to the solar forecasting field, where a short overview of irradiance forecasting and photovoltaic power forecasts with statistical and physical methods will be given. The following section will present an overview of current research in photovoltaic power forecasts with machine learning, focusing more on day-ahead forecasting. Finally, the last section will conclude the current state of the scientific domain and link it with this thesis' research objectives.

2.1 Introduction to solar power forecasting

This section discusses some fundamental concepts of the solar forecasting domain. It will briefly examine solar irradiance concepts and forecasting methods. Next, it will give a short overview of power forecasting methods apart from machine learning.

2.1.1 Solar Irradiance

The solar irradiance which enters the top of the atmosphere differs from the irradiance measured on the ground. This difference lies in the absorption and reflection of this irradiation by the different constituents of the atmosphere [31]. These interactions result in two components originating from the original extraterrestrial radiation. These two components are the Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DHI) [32]. The geometrical sum of those components gives the Global Horizontal Irradiance (GHI). The irradiance differs during the day, depending on the sun. To model this deterministic variability, the clear sky irradiance was introduced.

The clear sky irradiance and index are critical concepts in solar forecasting. The clear sky irradiance is defined as the irradiance observed if no clouds are present that attenuate the irradiation. The clear sky index (CSI) results from the clear sky irradiance as a normalization of the actual irradiance. Hence, the CSI splits up the irradiance in a deterministic component following clear sky conditions and a stochastic component, quantified by the CSI and dependent on cloud motion

[33]. It can be leveraged to deseasonalize the irradiance time series, as the seasonal and diurnal variation in irradiance is captured by clear sky irradiation. *Yang et al.* recommended this deseasonalistion for irradiance forecast in [34]. However, *Visser et al.* mentioned in [28] that small irradiance values at sunrise or sunset could lead to instabilities, and they recommend describing the seasonality in the input features instead for PV power forecasting.

Photovoltaic panels convert this irradiance into electric power. Most proposed PVPF methods in the literature rely on irradiance forecasts to forecast the PV output. The following section will discuss the techniques for day-ahead irradiance forecasts.

2.1.2 Day-Ahead Irradiance Forecasts

Different options present themselves to forecast the irradiance, each with its characteristic timescale and spatial resolution. Figure 2.1 summarizes sources of irradiance data. For each irradiance source, methods exist to extrapolate the input to forecast future irradiance value. Figure 2.1 shows that only NWP and perhaps remote sensing are relevant for forecasting irradiance for lead times over five hours.

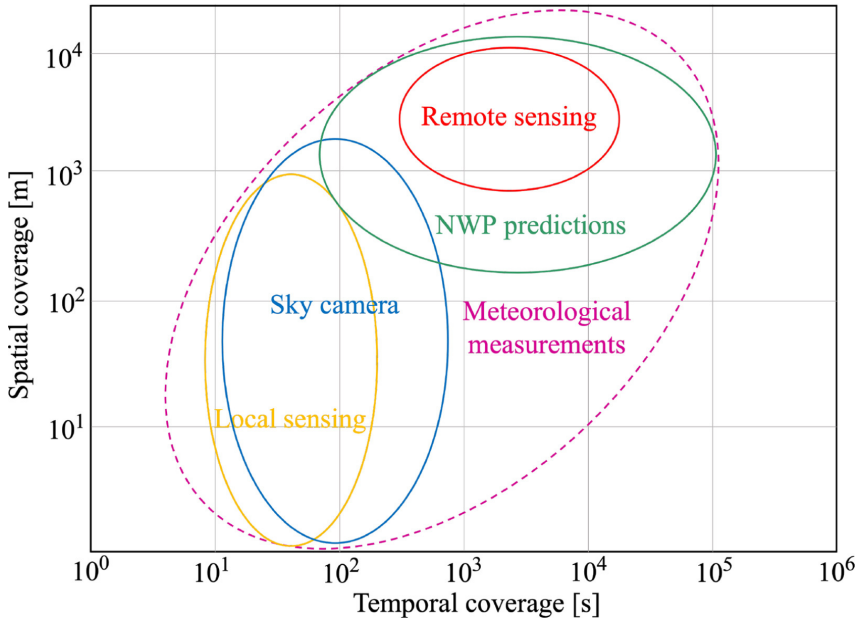


Figure 2.1: Relation of spatial coverage, temporal coverage, and model for irradiance. Source: [35].

Remote sensing consists of satellite imagery, which enables a global field of view. To forecast future irradiance, the motion of clouds is analyzed and extrapolated. Cloud formations and other variables, like water vapor content and aerosol concentration, taken from NWP [36] are utilized to estimate irradiance. Figure 2.2 shows the calculated irradiances starting from a satellite image. These models achieve a high spatial coverage but a limited resolution of 2 to 4 km due to a possibly larger

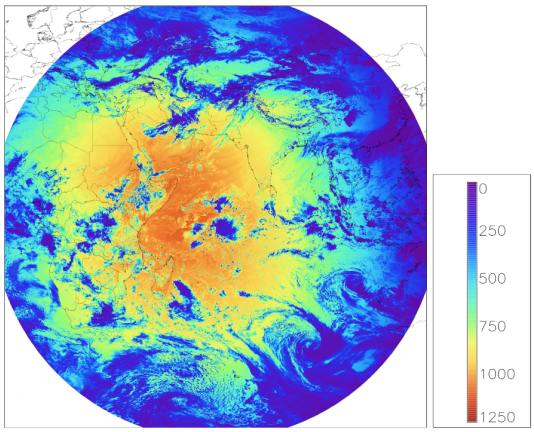


FIGURE 2.2: Irradiance determined by a satellite image. Source: [36]

field of view. Satellite images are sampled at higher frequencies, up to resampling of 5 to 15 min [35], which lowers the lower limit of temporal coverage as opposed to older works like [32] and [37]. Remote sensing methods are increasingly popular for intraday irradiance forecasting. *Huang et al.* provide a systematic review of irradiance estimation from satellite-based images in [38]. They briefly mention machine learning algorithms for estimating irradiance but note they cannot generalize well due to dependence on training data. Nonetheless, *Chu et al.* [35] report that machine learning can enhance performance for remote-sensing-based models.

Numerical Weather Predictions are another viable method to forecast solar irradiance. This method relies on knowing the initial state of the atmosphere with adequate accuracy and then using physical equations to forecast the state of the atmosphere in the future.Vilhelm Bjerknes proposed this idea in 1904 in [39]. Figure 2.3 shows a schematic overview of the NWP methodology, which remains the core idea of NWP. Due to a discretization of non-linear equations governing the atmosphere, spatial resolution is around the range of 1-20 km. Therefore, these grids lack the spatial resolution to accurately resolve the micro-scale physics concerning cloud formation [32]. Nevertheless, it allows for higher temporal coverage, and *Lorenz et al.* [40] show that NWP-based forecasts are more performant than satellite-based for forecast horizons over 4 hours. Solcast, a popular solar irradiance forecasting company, mentions using solely NWP data to forecast irradiance for lead times over 4 hours [41]. NWP seems to be the best option for day-ahead forecasts of irradiance.

Furthermore, advancements have been made to enhance the suitability of NWP for irradiance forecasting. *Jimenez et al.* [42] introduce Solar-WRF, which includes additional irradiance-related atmospheric interactions, to increase accuracy for irradiance forecasting. *Bai et al.* [43] show that the forecasting errors are not independent and show some correlation with other forecasted variables. Some research attempts to post-process NWP outputs with neural networks to improve accuracy. *Lima et al.* achieved around 10 % improvements in accuracy [44], and *Bakker et al.* achieved probabilistic irradiance forecasts using NWP as input [45]. The investigated models reported a 2 %-10 % decrease in Root Mean Square Error (RMSE) when evaluated

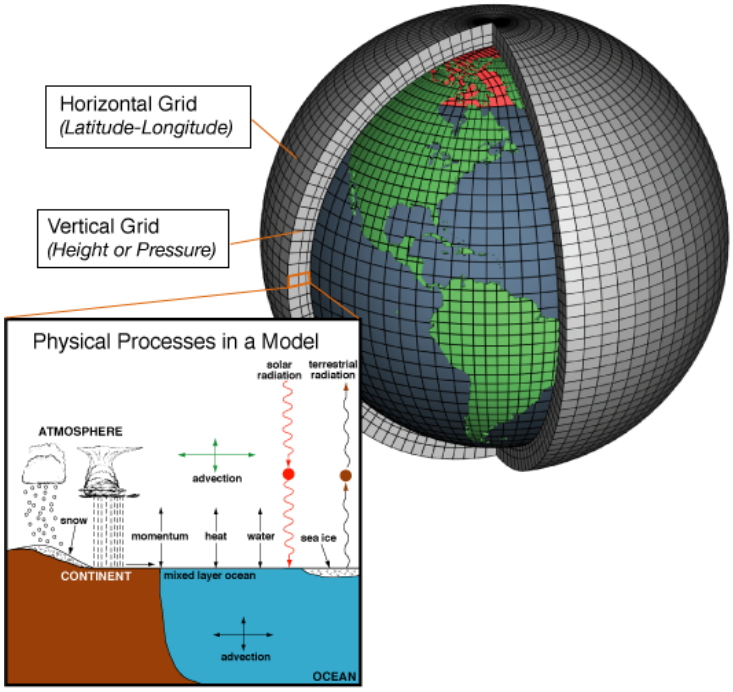


Figure 2.3: Summary of NWP methodology. Source: NOAA [46].

deterministically. Additionally, *Dou et al.* reported a decrease in RMSE of 5.7-9.3 % when using their method for processing day-ahead NWP predictions. Their method consisted of clustering the weather using a Variational Mode Decomposition model to decompose the time series and an encoder-decoder architecture to correct the NWP output.

This section gave an overview of different methodologies to forecast irradiance on a day-ahead timescale. Nonetheless, these sources of irradiance forecasts are not independent. Subsequently, a photovoltaic power forecast model can utilize this irradiance forecast to forecast the PV output.

2.1.3 Photovoltaic power forecast

A photovoltaic power forecast (PVPF) model attempts to forecast the power output of a PV installation. Similar to irradiance forecasts, different options exist depending on the model's spatial and temporal resolution.

Like irradiance forecasting, statistical methods can be applied for PV Power Forecasts, for which the autoregressive integrated moving average (ARIMA) model is popular. *Fara et al.* show that the ARIMA method outperforms Artificial Neural Networks (ANN) in their test setup [47]. *Vagropoulos et al.* [48] conclude that for intra-day forecasting, SARIMA methods, an acronym for seasonal ARIMA, are preferred over ANN. Still, the seasonal ARIMA with exogenous covariates model, SARIMAX, and ANN were the most performant for day-ahead predictions. The

SARIMAX model is similar to the ARIMA model but can include seasonality and other covariates. Furthermore, these statistical methods are used abundantly in literature for benchmarking the performance of PV forecasts with ML [49], [50], [51].

Physical irradiance-to-power conversion models can also forecast PV power using a forecast of input variables, such as irradiance. This type of method relies on parametrized physical equations, where these parameters are estimated using the manufacturers' datasheets. However, the parameters in this method are static and do not include the system's health or age. *Li et al.* propose a grey-box model based on physical equations, but the model's parameters are optimized using data-driven techniques [52]. They report a decrease in Mean Absolute Error (MAE) of 17.4 % compared to conventional physical models and claim that limited training data of 3 days is sufficient for a performant model. Next, *Mayer et al.* introduced an ensemble physical model constructed by collecting different models for each step in the model chain [53]. Like ensemble NWP, they introduced random perturbation to the optimal model chain, which uses different models for some steps of the model chain. They compared this method using deterministic NWP and ensemble NWP and concluded that including both ensemble models resulted in the best metrics for probabilistic forecasts.

In conclusion, various methods exist for PVPF. One option has not yet been introduced. Machine learning models show great promise in the field of PVPF. The subsequent section will explore the added value of these ML models for PVPF.

2.2 Machine Learning for Solar Power Forecasting

Previous sections introduced different concepts related to solar forecasting. The main scope of this thesis is PV power forecasting with machine learning. ML methods are popular for solar power forecasting due to their ability to model non-linear process characteristics. Various review papers focus solely on machine learning for solar power forecasting, e.g., [20] and [37]. These review articles show the wide variety of ML methods depending on forecast horizon, type of predictor variables, and more. This section will limit its scope to literature relevant to day-ahead solar power forecasting. However, exciting forecasting methods for other timescales will be mentioned. I structured this section by the main trends noticed in the solar power forecasting literature and their relevance to this thesis. The first section will discuss the different ML methods and architectures for PVPF. The following section will discuss transfer learning and its role in solar forecasting. Finally, the last section addresses physics-informed ML algorithms.

2.2.1 Different ML methods

During the literature review, I noticed that a considerable amount of literature consisted of researchers proposing new variants of models to improve performance. This section will give an overview of different trends in the proposed ML methods. First, it will discuss state-of-the-art ML techniques for univariate and multivariate PV power forecasting. Multivariate PV power forecasting uses predictor variables like

irradiance forecast to estimate PV power better. Finally, the role of hyperparameter tuning for PVPF with ML will be briefly discussed.

Popular ML methods for solar forecasting are Support Vector Machines (SVM), ensemble methods like Random Forest Regression (RF) or Gradient Boosting (GB), Artificial Neural Networks, and, more recently, deep learning algorithms like Convolutional Neural Networks (CNN), LSTMs, and transformers. Several papers compare these ML algorithms for forecasting solar power. *Visser et al.* compared linear, ensemble, deep learning, and physical models on their suitability for day-ahead PV forecasting in [27]. It concluded that the best-performing model depended on the user's objective. A physical model performed the best for the economic metrics, but RF achieved the lowest MAE. *Voyant et al.* [37] discussed different machine learning methods for solar irradiance forecasting. It concluded that combining different methods will always lead to superior performance. *Kim et al.* concluded that for short-term solar forecasting, the LSTM architecture performed best overall [50]. *Markovic et al.* compared a variety of linear and ML models [54]. The Kernel Ridge regression model performed the best overall, outperforming RF and SVM models. However, the authors mentioned that an ANN performed almost as well at a fraction of the training time.

These papers compare different ML models and show that the optimal ML model depends on the case and the considered evaluation framework. Furthermore, these reviews did not include deep learning models apart from [27] and [50]. Deep learning models are increasingly popular for solar power forecasting.

In 2016, *Gensler et al.* reviewed deep learning models for solar power forecasting in [55] and concluded that an auto-decoder LSTM model showed the highest performance. Later, *Massoudi et al.* reviewed this convergence of photovoltaic power and deep learning models in [56]. They mentioned that an LSTM is excellent for modeling the temporal dependence of solar energy. However, they lack spatial modeling, which a CNN is more suited for. Therefore, they mention hybrid models such as the CNN-LSTM, which can accurately model the temporal and spatial dimensions. However, they also mention that these hybrid models have a high computational cost. Therefore, the trade-off should be considered.

These hybrid deep learning models are also gaining traction in literature. A significant amount of variants are introduced, each with its own specifics. Nonetheless, hybrid models containing a CNN and LSTM seem to be a popular choice. *Liu et al.* proposed a CNN-bidirectional LSTM (biLSTM) model with an attention head for intra-hour forecasting in [57]. They performed ablation studies that showed that each component had added value for accurately forecasting PV power. *Wang et al.* introduced a variant on the CNN-LSTM and used an LSTM and CNN to compare with their hybrid. However, they didn't use the solar zenith filtering, where forecasts for a zenith angle of the sun over 85° are filtered, which is customary for solar power and irradiance forecasting [27]. The LSTM performed second best but suffered the most from the absence of zenith filtering, as its nighttime forecasts were negative values. Other papers that use a hybrid model, which is a variant of a CNN-LSTM, are [58], [59], and [60].

Ensemble Machine Learning (EML) models are being investigated in addition to

hybrid deep learning methods. EML models consist of multiple base learners whose results a meta-learner combines to achieve a more accurate forecast [61]. Different types of EML models, like bagging, boosting, and stacking, exist depending on how the base learners are combined. The aforementioned RF and GB are popular EML models. *Khan et al.* introduce an ensemble stacking model with an LSTM and ANN as base learners for solar power forecasting. Using this ensemble model, they reported improvements of around 10 % in RMSE. In [62], *Alkandari et al.* proposed an ensemble method of an LSTM, Gated Recurrent Untit (GRU), and statistical models. Their results showed that an ensemble model, including statistical models, performed better than an ensemble model with only ML models. Finally, *Chakraborty et al.* concluded that ensemble methods have a higher performance than linear and k-NN regression [63]. Nonetheless, the error metrics differed for the various tested ensemble methods.

In conclusion, endless ML architectures and configurations can be devised to forecast PV generation. Deep learning, hybrid deep learning, and ensemble methods seem to be the most prominent future directions for the type of ML model. Nonetheless, *Alcaniz et al.* mention that actual differences in performance between methods are not high once the technique can capture non-linearities [20]. They argue that for developing a model, more emphasis should be placed on the applicability of ML models in the real world concerning computational resources, limiting the need for training data and its interpretability. Finally, all proposed multivariate models use irradiance forecasts for day-ahead solar power forecasting. Except for [64], where *Boussif et al.* proposed a multi-modal framework that combines past time series and satellite images for day-ahead forecasting.

Hyperparameter Tuning

The design of an ML model comprises various hyperparameters that specify the architecture of the ML model. Adjusting these hyperparameters can enhance the performance of the ML model. This process of adjusting hyperparameters is known as hyperparameter tuning or hyperparameter optimization. *Yang et al.* [65] summarizes different hyperparameter tuning methods.

Hyperparameter tuning can be essential for going from an adequate to a good ML model for PVPF. A lot of papers describe the usage of hyperparameter tuning without emphasizing it. *Tahir et al.* reviewed various hyperparameter tuning methods for solar forecasting in [66]. They tested a Bayesian and random grid search tuning method on the performance of various ensemble models. Although a Bayesian optimizer is more advanced, random grid search outperformed the Bayesian optimizer a few times. Nonetheless, tuned models always outperformed their standard configurations.

2.2.2 Transfer Learning

Zhuang et al. define transfer learning as " Given some/an observation(s) corresponding to $m^S \in r\mathbb{N}^+$ source domain(s) and task(s) (i.e., $(\mathcal{D}_{S_i}, \mathcal{T}_{S_i})|i = 1, ..., m^S$),

and some/an observation(s) about $m^T \in r\mathbb{N}^+$ target domain(s) and task(s) (i.e., $(\mathcal{D}_{T_j}, \mathcal{T}_{T_j})|j = 1, ..., m^T$), transfer learning utilizes the knowledge implied in the source domain(s) to improve the performance of the learned decision functions $f^{T_j}(j = 1, ..., m^T)$ on the target domain(s) " ([67], p.1)

They define the concepts of domain and task in [67]. Transfer learning can be seen as the machine learning equivalent of a child learning how to ride a bicycle quickly after riding a tricycle as a kid. The effectiveness of transfer learning depends on the similarity of the source and target domain and how this transfer learning is applied. There exist a multitude of transfer learning strategies, which are excellently summarised in [67], [68] and [69]. TL can be subdivided into model-based approaches, which share a model over domains, and feature-based approaches, which train a model on shared source and target data. *Weber et al.* reviewed transfer learning for time series forecasting in [70]. They observed that an LSTM and CNN architecture are the most popular architectures for TL in time series forecasting

Transfer learning has become increasingly investigated in the solar forecasting community. *Abubakr et al.* [71] presented a transfer learning model for forecasting solar irradiance. The source domain consisted of irradiance measurements, and they fine-tuned the transfer model to limited observations in the target domain, which showed slight improvement. *Sheng et al.* [72] opted for the data-based approach to transfer learning. The authors proposed to merge source and target domain data to train their proposed Transfer Support Vector Regression model. *Tang et al.* [73] showed a complex ensemble model where they applied transfer learning to generalize to a new site. They applied different strategies and reported that the strategy with fine-tuning of the feature extraction and mapping components performed the best. *Genovese et al.* applied TL with an ANN to generalize forecasting over different PV cell technologies [74]. Their source model was trained on generation for different PV technologies, and they reported increased performance of forecasting target data consisting of the PV power of one technology. *Tang et al.* [75] used Shappley values to determine the similarity between a model trained on limited target data and three source models trained on different source domains. Selecting a source model based on the similarity between Shapply values decreased the RMSE with 3 % compared to using another source model.

Literature shows that the LSTM is a widespread architecture for TL in solar power forecasting. *Sarmas et al.* [76] tested three different TL strategies for solar power forecasting with an LSTM architecture. The source domain was a PV installation for which abundant generation data was available, and the target domain consisted of other installations with different orientations and tilts. The tree techniques were freezing some layers of the model, fine-tuning all of the weights, and resetting the last layer. These techniques improved performance over a model only trained on target data, while results between the strategies were not significant enough to draw conclusions. *Zhang et al.* used an LSTM, trained on PV generation of other sites, to forecast the PV generation of a new installation [77]. They reported that these methods are advantageous when data is limited to 21 days. The performance became similar when the target domain data included more than 21 days. Furthermore, they found that training the source model on data from PV installations further

away decreased performance and that it was best to transfer all parameters to the source domain. Finally, *Kim et al.* proposed an LSTM-TL model for building energy consumption using synthetic data for training the source model. This model is trained on the simulated energy consumption of a building under different weather conditions. The TL with simulated source data resulted in a model that could forecast energy consumption when presented with new weather conditions in the target domain.

2.2.3 Physics-Informed Machine Learning

Karniadakis et al. define Physics-Informed Machine Learning (PIML) as :

" The process by which prior knowledge stemming from our observational, empirical, physical, or mathematical understanding of the world can be leveraged to improve the performance of a learning algorithm " ([78], p.2)

They state that this prior knowledge can be embedded in an ML model via three separate methods in [78]. These methods are via data, which adheres to the physical laws, via custom kernels, or via a term based on the physical equations in the loss term. PIML enhances a model's robustness, allowing it to cope better with imperfect or limited data. PIML can also play its part in weather and climate modeling [79] or power system stability modeling [80].

PIML has gained interest in the solar forecasting community. It should be noted that PIML in solar forecasting is currently limited to engineering features and transforming the target variable. *Pombo et al.* showed that introducing physics-informed features, like the cell temperature, the forecasted power of a physical model, the type of day, and the solar position increased the model's performance [29]. It also introduced a more advanced thermodynamic model for the cell temperature. *Paletta et al.* [81] concluded that applying a clear sky normalization of the irradiance could improve the generalizability of a model to a new site. However, they mention the challenge of constructing an accurate clear sky model for PV power as shading should be considered in a distributed PV plant.

2.3 Conclusion

The previous section reviewed the current literature on machine learning for solar power forecasting. This section situates this thesis's research within the existing literature by analyzing its trends and gaps.

First, different models exist in the literature, and all seem to have some improved performance over others. Nevertheless, review papers address that further improvements are minor once the model reaches a point where it can capture non-linearities. Furthermore, the LSTM architecture seems to be a popular choice for research concerning transfer learning, making it an appropriate choice of ML model for this thesis.

Second, the literature around TL focuses on source domains consisting of other PV installations. The main drawback is that these installations differ from the target installation in tilt, orientation, or weather conditions. Furthermore, some papers

mention the importance of having a PV installation for the source domain close to the target installation. To the best of my knowledge, no literature mentions the use of synthetic or simulated PV data for training a source model.

Next, adding physics-informed covariates seems to improve ML models' performance for forecasting PV generation. Nonetheless, research reporting the use of physics-informed features for solar forecasting is sparse, and the available literature does not precisely discuss the results for an LSTM or TL approach.

To conclude, investigating TL with an LSTM model that uses physics-informed input features and synthetic PV generation in the source domain addresses gaps in the current literature. The current literature does not mention using reanalysis data to train the source model. The findings of this thesis can help future research decide whether it's worth finding historical NWP runs to train their ML models.

Chapter 3

Methodology

This chapter discusses the methodology used in this thesis. The first section examines the forecasting scenario for this thesis. The second section gives a structured overview of how the proposed transfer learning model will be constructed and which modifications will be investigated. Next, the topic of the third section is the proposed ML model and architecture. The fourth section will dive deeper into the specifics of the source and target model. Subsequently, the fifth section mentions the baseline models against which the transfer learning model will be evaluated. Finally, the last section covers the metrics for assessing the model's performance.

3.1 Forecasting Setting

The proposed model aims to effectively predict the solar power production of a residential installation on a day-ahead time scale. Considering the literature for day-ahead forecasts, I suggest a deep learning network that uses transfer learning to overcome the limited data availability. The considered input features are forecasts of weather variables, physis-informed features, and time encodings. The weather covariates result from the NWP run at 00:00. It would be perfectly possible to update the forecasts each time a new NWP output is available. However, this is not further considered in this thesis due to the similar methodology.

The goal is to find an out-of-the-box model that performs well even when no historical production data is available. The model will start with issuing a day-ahead forecast for each day at 00:00 without seeing any historical power production. After 30 days, the model will fine-tune its parameters with these 30 days of the historical output. Fine-tuning means the model will adjust its weights and biases to suit better the target domain, which is the actual PV power. Section 3.4.3 discusses this method further.

3.2 Transfer Learning model

This thesis aims to quantify the effect of two modifications compared to the base model concerning the accuracy and computational intensity of day-ahead forecasting

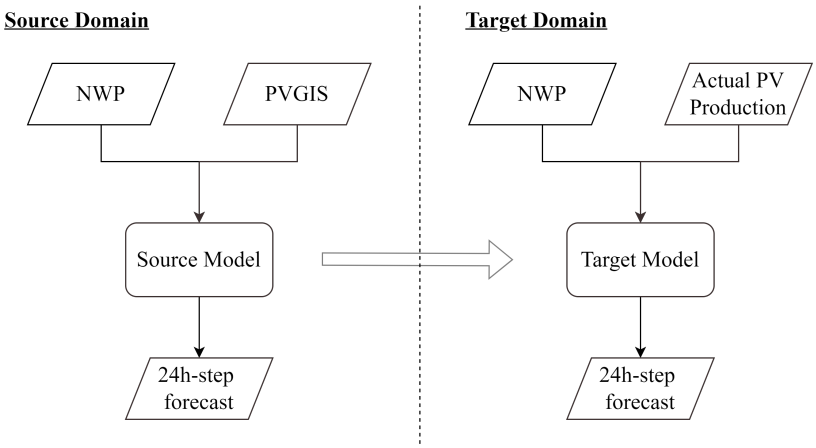


Figure 3.1: Starting point of the transfer learning model

of PV production. Figure 3.1 shows the starting point of the proposed model. An ML model is trained on synthetic PV generation and historical NWP runs. This model is utilized in the target domain to forecast the PV power of the first month. These are called zero-shot forecasts, as the model has not seen any information about the PV generation in the target domain. Modifications are introduced, and the original structure will be changed, which will be discussed in the following sections.

The first modification is developing a more physics-informed model, which section 3.2.1 will discuss. The second modification is training the source model with weather covariates provided by reanalysis data instead of historical NWP runs. Section 3.2.2 dives deeper into this modification. The possible combinations of these two modifications lead to 4 cases that will be assessed.

3.2.1 Physics-Informed Model

The first modification is to add physics-based information to the model to improve its forecasting accuracy and make training less intensive. The model was already physics-informed using NWP, whose forecasts are based on physical equations. However, there exist more physics-informed improvements. Figure 3.2 shows the changes compared to the starting point. The following sections will discuss each modification separately. The first section discusses the addition of physics-informed variables to the model. Next, the second section discusses the post-processing of the model's output concerning the clear sky irradiance. Finally, the last section discusses limiting the synthetic source data for the PV generation to the rated inverter output.

Physics-informed variables

Chapter 2 presented several studies that show significant improvement in forecasts if physics-informed variables are added as input features. Therefore, the physics-informed cases include the GTI and the temperature of the PV panel, T_{PV}.

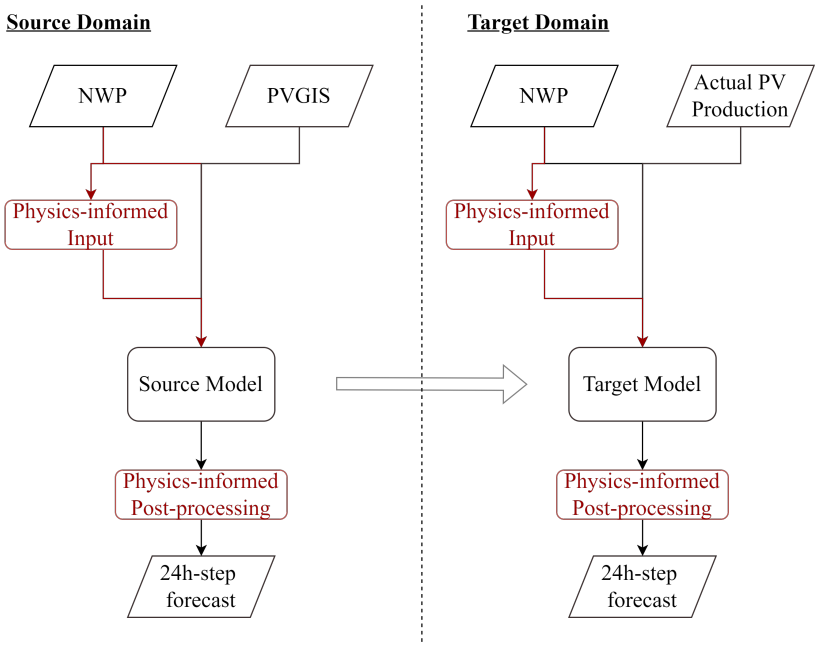


FIGURE 3.2: Transfer Learning Model with physical-based information

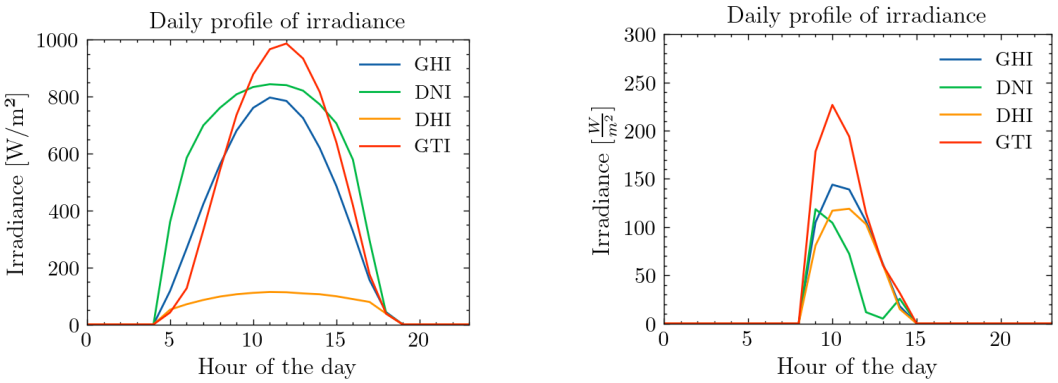
The GTI is the irradiance the surface with the same orientation as the PV panel, receives. Various models exist, which differ in their method for accounting for the diffuse irradiance [82]. The Perez model is a prominent model that considers the anisotropy in the atmosphere [83]. It superimposes on an isotropic sky the circumsolar diffuse radiation and the horizon brightening radiation. The circumsolar diffuse radiation accounts for the preference of forward scattering for aerosol particles, which means that the diffuse irradiance is higher in a disk around the sun. Horizon brightening radiation is a diffuse radiation component concentrated near the horizon and is more pronounced in clear skies. Equation 3.1 gives the GTI using the Perez formulation.

$$\begin{aligned} GTI =& R_b\ DNI + \rho \left(\frac{1-\cos\phi}{2}\right) GHI \\ &+ DHI \left[(1-F_1)\left(\frac{1+\cos\phi}{2}\right) + F_1 \frac{a}{b} + F_2 \sin\phi\right] \end{aligned} \tag{3.1}$$

ϕ is the tilt angle of the surface, R_b is the geometrical relation between the tilted and horizontal beam irradiance, ρ is the ground reflectance, and F_1 and F_2 are brightness coefficients for, respectively, the circumsolar and the horizon brightness. These coefficients are determined empirically. The coefficients a and b consider the incident angle of the sun on the tilted slope and are given respectively by equation 3.2 and 3.3.

$$a = \max(\cos 0°, \cos\gamma) \tag{3.2}$$

$$b = \max(\cos 85°, \cos\theta) \tag{3.3}$$



(A) Summer (B) Winter

FIGURE 3.3: Daily profile of all irradiance components

γ is the incidence angle on the surface, and θ the solar elevation angle. Figure 3.3 shows these different irradiation components for a typical summer and winter day.

The DNI and DHI are not always available as NWP output variables but can be estimated using a decomposition model. A decomposition splits the GHI into the DNI and DHI. Perez and Ineichen proposed the dynamic global-to-direct irradiance DIRINT conversion model in [84] as an improvement to the already existing DISC model, developed by Maxwell [85]. However, the DIRINT model uses a stability index, which uses subsequent values of the clearness index. If the time series misses several values, it leads to additional timestamps for which the DNI and DHI cannot be calculated. Due to this problem, this thesis utilizes the DISC model to estimate the DNI and DHI. Equation 3.4 gives the Direct Normal Irradiance using the DISC model.

$$DNI_{disc} = I_{et}(K_{nc} - (A + B \exp mC)) \tag{3.4}$$

K_{nc} is a parameter that is a function of the airmass, I_{et} is the extraterrestrial radiation, and A, B, and C result from experimental correlations in function of the clearness index. All irradiance components are estimated. Only the PV panel temperature must still be calculated.

The Fuentes model provides the temperature of the PV panel, T_{PV}, using a simple heat-transfer model [86]. The temperature results from one-dimensional, linearised heat equations and uses empirical formulas to determine the convective and radiative coefficients. Current literature has proposed more advanced models for this temperature, e.g., [29] and [87]. Nonetheless, *Olukan et al.* [87] showed that, compared to the simplest models, RMSE differences are in the range of 0.5 %. In conclusion, the error in the variables forecast will dominate the error of the thermal model, and therefore, the utilized model will have less effect on the final accuracy.

All mentioned models are implemented using the open-source Python package *pvlib*. *Anderson et al.* [88] developed this package to help simulate the performance of PV installations and to provide other functions related to solar PV installations.

Physical post-processing of ML output

The PV generation is always bounded between 0 kW and the power generated if clear sky irradiance reaches the panel. This PV power, as a result of clear sky irradiance, is called clear sky power (CSP) for the remainder of the thesis. We apply this physics-based post-processing to the ML model's forecast and evaluate the performance based on these processed forecasts. First, a function calculates the clear sky irradiance using the Perez-Ineichen model [89], one of the most performant models for clear sky irradiance [90]. Equations 3.5-3.7 display the Perez-Ineichen model for clear sky irradiance.

$$GHI = 0.84 I_{et} e^{-0.027 AM(e^{-h/800} + e^{-h/1250}(T_L - 1))} \tag{3.5}$$

$$DNI = \frac{I_{et}}{\cos\theta}(0.664 + \frac{0.163}{e^{-h/8000}} e^{-0.09 AM(T_L - 1)} \tag{3.6}$$

$$DNI = \min(DNI, \frac{I_{et}}{\cos\theta}(1 - \frac{(0.1 - 0.2e^{-T_L})}{0.1 + \frac{0.88}{e^{-h/8000}}} \tag{3.7}$$

where h is the altitude. Second, this irradiance is converted to a theoretical power generation using the PVWatts model, where the possible losses are set close to zero. Section 3.5 explains the physical model. The irradiance can be higher than the clear sky irradiance, which could result in an actual PV production exceeding the limit. Nonetheless, this happens right after sunrise or before sunset, moments with low PV production. Therefore, this problem will not influence the RMSE as much.

Inverter limit in source domain

In moderate climates, PV panels will rarely output their rated power. To decrease installation costs, the rated power of the inverters is sometimes lower than the installed PV power. The inverter will limit a possible higher PV output by its rated power. The synthetic data is modeled by considering the installed PV power but does not consider this 'derating' of the PV inverter. Consequently, the physics-informed cases will limit the power in the source domain by this inverter limit to have a higher resemblance between source and target domain data.

3.2.2 Reanalysis data as input for the source model

As mentioned in section 4, reanalysis (RA) data is more accessible than historical NWP runs, and data is available going further back in time. Considering these advantages, it could be interesting to investigate if reanalysis data can be used instead of historical NWP runs to train the source model. Figure 3.4 shows the updated model diagram. Furthermore, both modifications can be applied simultaneously, represented by figure 3.5.

There are some remarks about utilizing reanalysis data in the source domain. First, the model can already capture biases and other errors present in the NWP if runs from the same provider are utilized in both source and target domains. This will happen when reanalysis data is used because it comes from another provider

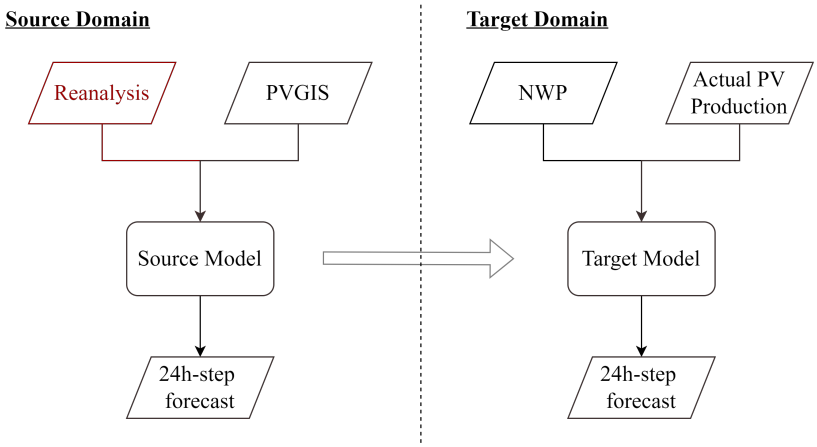


Figure 3.4: Transfer learning model with reanalysis data in the source domain(not physics-informed)

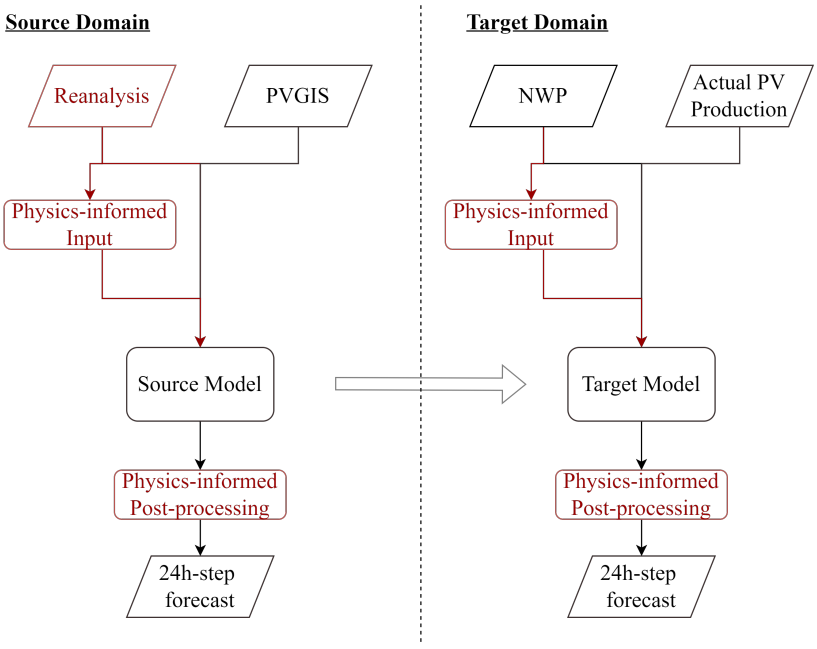


Figure 3.5: Transfer learning model with reanalysis data in the source domain (physics-informed)

and incorporates actual weather measurements. Therefore, the expectation is that this model's zero-shot accuracy will be lower. Second, due to inaccuracies in the underlying NWP model, irradiance from reanalysis data can differ from the actual irradiance measured at the ground. However, the same remark can be made about the NWP runs. Finally, a provider may update their NWP model. Consequently, this can influence the biases and errors of the NWP, negatively impacting the performance of the ML model and leading to worse generalisability. Using reanalysis data avoids this risk, as the same NWP model generates the entire dataset.

In conclusion, there is a lot of uncertainty about the model's performance when using reanalysis data in the source domain. However, its significantly higher availability is a sure advantage.

3.3 Machine learning architecture

Inside the domain of machine learning, a collection of techniques exists called deep learning. The core idea is to have multiple combinations of simple but non-linear transformations on the input data and obtain abstract representations of this data to learn complex functions accurately [91]. The most prominent deep learning techniques are the Convolution Neural Network, the Deep Neural Network (DNN), the Recurrent Neural Network (RNN), and the transformer architecture. Conventionally, the RNN is the most prominent deep learning technique for time series analysis because *Yu et al.* designed it to retain temporal information of the input data [92].

The proposed structure will use an LSTM architecture as the ML model. This architecture has several advantages over other ML models. First, chapter 2 mentioned that an LSTM is an appropriate choice for time series forecasting. Furthermore, this thesis utilizes model-based transfer learning in its simplest form. This approach translates to parameter sharing of the models over domains and fine-tuning these parameters on target data. This philosophy is straightforward to implement with an LSTM, where the learning rate, which regularizes the rate of change of the parameters concerning the error, can be adapted for the target domain. Using the same input features allows immediate utilization of a trained source model in the target domain because of its simplicity. Other popular ML methods for solar power forecasting, e.g., RF and SVM, are not gradient-based models. Therefore, they don't have this regularizing parameter limiting the rate of change. Although model-based transfer learning strategies exist for these methods, they are more involved.

In conclusion, the LSTM architecture facilitates simple parameter sharing over domains and is known to handle time series forecasting very well. These advantages motivate the choice of an LSTM architecture as the chosen ML method. However, section 3.7 explains the analysis of comparing other ML models to the LSTM.

3.3.1 Long-Short Term Memory

The RNN retains temporal information using previous outputs and current input to update its state. However, these RNNs suffer from a critical disadvantage. If the time gap between relevant data is large, the RNN can suffer from an exploding

or vanishing gradient, where the gradient goes respectively to zero or infinity. To resolve these issues, Hochreiter and Schmidhuber introduced the Long-Short Term Memory in [93].

The LSTM cell

Many variants of LSTM cells exist, each with slightly different characteristics. Nevertheless, the most popular of these variants is the LSTM with a forget gate, depicted in figure 3.6. This figure depicts one cell of an LSTM network where each cell is fed with one instant of time series data. The equations correspond to:

$$f_t = \sigma\left(W_{fh}h_{t-1} + W_{fx}x_t + b_f\right) \tag{3.8}$$
$$i_t = \sigma\left(W_{ih}h_{t-1} + W_{ix}x_t + b_i\right) \tag{3.9}$$
$$\tilde{c}_t = \tanh\left(W_{\tilde{c}h}h_{t-1} + W_{\tilde{c}x}x_t + b_{\tilde{c}}\right) \tag{3.10}$$
$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \tag{3.11}$$
$$o_t = \sigma\left(W_{oh}h_{t-1} + W_{ox}x_t + b_o\right) \tag{3.12}$$
$$h_t = o_t \cdot \tanh\left(c_t\right) \tag{3.13}$$

where x_t is the input at time t, c_t is the cell state, f_t is the forget gate, W denotes weight matrices, b is a bias vector, and h_t is the output of one LSTM cell. Furthermore, all W coefficients denote weights and all b coefficient biases. Figure 3.6 shows that the LSTM cell consists of 3 sequential gates. First, the forget gate determines in what order the previous cell state should be remembered based on the input data at time t and the preceding cell's output. Next, the input gate calculates the new cell state using the output of the forget gate and the new input. Finally, the output gate determines the output of the LSTM cell using a combination of the new cell state and the input. An LSTM network consists then of a multitude of these LSTM cells. Nonetheless, various options exist for configuring these LSTM cells into a network.

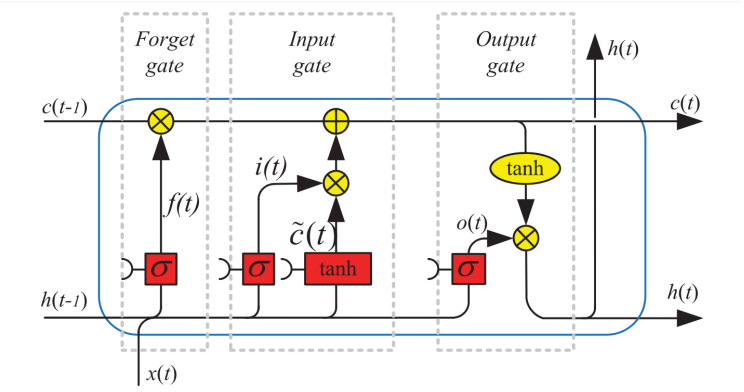


Figure 3.6: Long-Short Term Memory cell. Source: [92]

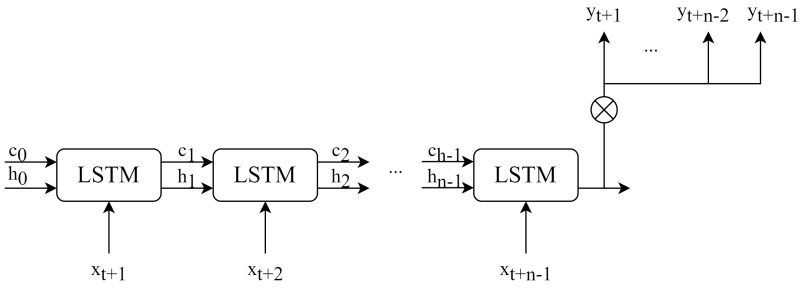


FIGURE 3.7: The LSTM architecture of this thesis

LSTM Architecture for multi-step time series forecasting

A distinction should be made if the dimension of the output is single- or multi-step, which translates to an output of 1 timestep or multiple. A multitude of options exist for multi-step time series analysis with LSTM architectures. This thesis proposes an LSTM network as a feature extractor, where the output of the last LSTM cells is linearly combined with the desired dimension of the multi-step output.

3.3.2 Regularization

A common problem in ML is the overfitting of the model. Overfitting means that the model fits the statistical noise in the training data and cannot generalize well to unseen data. This problem, illustrated in figure 3.8, can be mitigated using regularization techniques. Regularization techniques reduce overfitting to training data [94]. A subset of these techniques adds a component to the loss function to improve generalizability. The L2-regularization is a common technique in this subset.

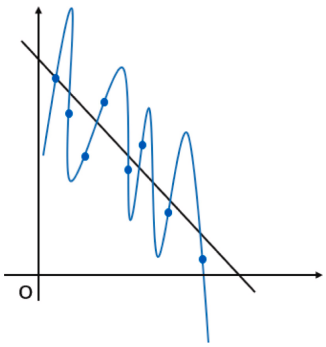


FIGURE 3.8: Example of overfitting. Source: [94]

L2-regularization adds a term to the loss function, penalizing the squared sum of all the weights. Equation 3.14 shows the updated loss function, $L_{l2}(w)$.

$$L_{l2}(w) = L(w) + \lambda \|w\|^2 \tag{3.14}$$

$L(w)$ is the original loss function, λ is the weight decay, which determines the severity of the penalty, and w is a vector containing the model's weights. For gradient-based machine learning models, the gradient of the loss function determines the updated model weights for each iteration. The penalty for the weights will be a linear term in this gradient. Consequently, the L2-regularization can be implemented directly into the weight update without altering the loss function. Equation 3.15 illustrates the weight update with L2-regularization.

$$w_{i+1} = w_i - \beta \nabla L(w) + \lambda w_i \tag{3.15}$$

β is the learning rate. In reality, more advanced optimizers, e.g., *Adam* and *RMSProp*, exist that use additional techniques, like momentum, to update the weights. Nonetheless, the principle of weight decay remains similar to the explanation given.

3.4 Source and Target Model

3.4.1 Overview

The source model should forecast synthetic solar production as accurately as possible. There are some inconsistencies between this data and the actual generation data. As such, a generalizable source model must be obtained that is not overfitted for synthetic PV generation. The transfer model should then do accurate PVPF. After one month, it can update its parameters on the new available target data. However, there is a risk that it overfits on the limited target data and can not generalize well, resulting in poor forecasts for the next month. To limit this risk, hyperparameter tuning should be performed for the model in the target domain. Figure 3.9 shows the overview of the hyperparameter (HP) tuning, training, and evaluation pipeline.

3.4.2 Source Model

For each case, a different model is needed. For example, a source model designed for forecasting based on NWP input will not perform the same if the input is reanalysis data. As such, a general pipeline is proposed to have the most performant, generalizable models possible. The hyperparameter tuning process is conducted with the open-source Python library *Optuna* [95], which contains a vast amount of built-in functions that facilitate the development of a hyperparameter tuning framework.

The first step is the hyperparameter tuning of the source model. This step uses a random search algorithm that randomly chooses 500 sets of hyperparameters and selects the set that obtains the lowest overall Mean Square Error (MSE), where each iteration is called a trial. The hyperparameter tuning process uses 5-fold shuffled cross-validation to assess the error over the whole dataset. The source dataset spans three years and will be split into five shuffled folds, allowing a more uniform MSE over the folds. Consequently, minimizing the overall MSE, MSE_{tot}, across folds will result in an adequate set of hyperparameters. Equation 3.16 shows the total MSE

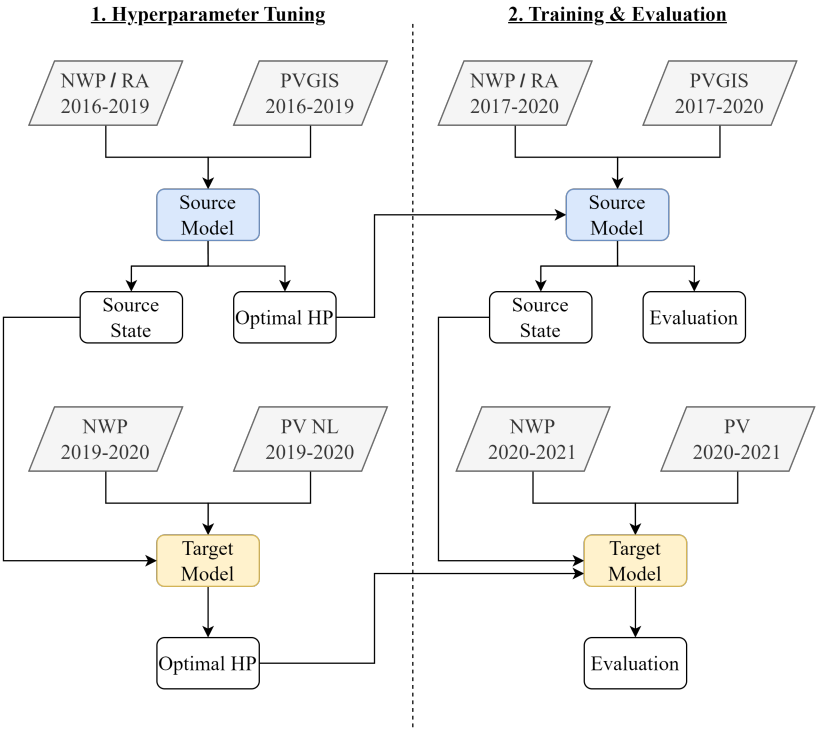


FIGURE 3.9: Overview of the development pipeline of the source and target model

over all folds.

$$MSE_{tot} = \frac{1}{k}\sum_{i=1}^{k} MSE_i \tag{3.16}$$

k is the number of folds for cross-validation, and MSE_i is the MSE for fold i. There will be trials with an intermediate test MSE, which is higher than the MSE of other trials, hinting at an inadequate set of hyperparameters. With *Optuna*, a median pruner is implemented, which interrupts trials with an intermediate MSE above the median MSE of all trials that reached the intermediate point. Table 3.1 presents the configurable hyperparameters and their search spaces.

TABLE 3.1: The configurable hyperparameters and their respective search space

Hyperparameter	Search space
Optimizer	[Adam, RMSProp, SGD]
Nb. of nodes	[4, 1000]
Learning rate	$[10^{-6}, 10^{-1}]$
Batch size	[4,128]
λ	$[10^{-9}, 10^{-1}]$

In conclusion, the model with the optimized hyperparameters is trained and

transferred to the target domain. The source data for training differs from the source data for hyperparameter tuning because the training window slides to the day before the start date of the evaluation dataset. This sliding emulates the real-life scenario better, as all available data until the day before the inference is used to train the source model. It allows for a more unbiased estimate of the performance of the source model. The source model will be evaluated by its RMSE of forecasts for the test data and the time it took to train the model. The trained model will be saved and transferred to the target domain for zero-shot forecasting.

3.4.3 Target model

Ultimately, the goal is to have a model that can forecast accurately the day-ahead solar power in the target domain. This thesis proposes a model that is already trained on a source domain and will fine-tune its weights and biases on the new, available historical generation data of the PV installation. Consequently, the model's accuracy is simulated with the walk-forward validation scheme, as mentioned in section 3.1. The first section outlines how the parameters are tuned for the target model to achieve maximal accuracy. The following section explains how the model's performance will be simulated and reported.

Validation schemes

A popular strategy for assessing overfitting is k-fold cross-validation [96]. However, the non-stationarity of time series can pose problems for k-fold cross-validation, where each subset of data is assumed to be sampled from an identical distribution. Non-stationary data has a changing underlying distribution. Therefore, new cross-validation schemes are proposed for time series forecasting. These schemes, which respect the temporal order, are called the forward-validation methods. [97] showed that for perturbations to stationarity, walk-forward methods are preferred. The actual PV power is non-stationary as the PV modules and components degrade over time. Additionally, walk-forward validation methods better resemble the real-life application of forecasting, where models are updated when new data becomes available.

Walk-forward validation schemes have a validation set that always follows later than the training set chronologically. Two popular techniques exist for constructing the different splits. First, the expanding window technique will include the validation data used in the previous iteration into the training set and continue this for N specified iterations. Second, the sliding window technique uses the same length of training data and discards the data furthest in the past for each iteration. The sliding window method allows for faster computation and less variance of the validation metric due to the constant size of training data.

Due to a limited amount of data in the beginning and the yearly seasonal component of solar power, this thesis proposes a walk-forward validation method where the window will expand until it has one year of data and the annual seasonality of PV generation is represented in the dataset. Next, the window will transition into

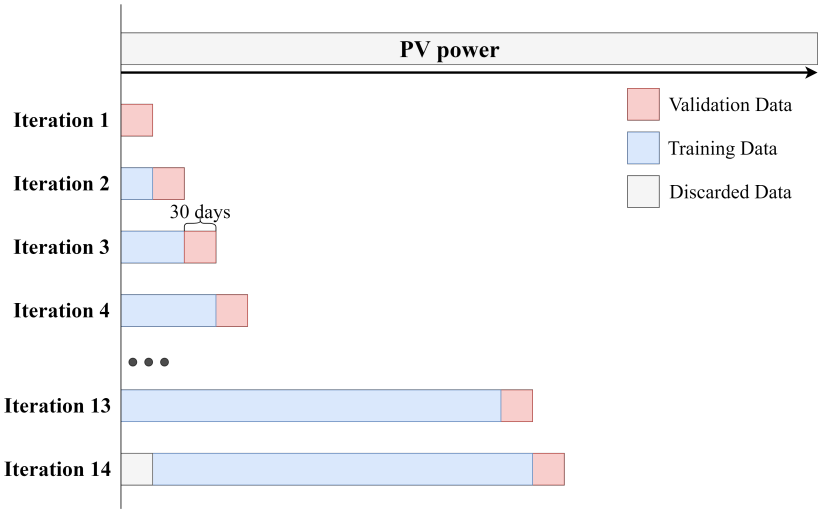


FIGURE 3.10: The window method used in the target domain

a sliding window where the oldest 30 days will be replaced by the most recent 30 days of PV generation. This technique is summarised in figure 3.10.

Hyperparameter Tuning

Initially, the model in the target domain is a copy of the final source model. Some hyperparameters related to the model architecture remain the same and should not be tuned. Nevertheless, the model incorporates the new PV generation data to learn more about the target domain. Fitting too fast on this new data could lead to overfitting, and a too slow fitting could lead to suboptimal accuracy.

Consequently, the parameters associated with this learning process must still be optimized. The learning rate, batch size, and λ are the hyperparameters tuned in this process. Table 3.2 shows their respective search spaces. Compared with the source domain, the search space for the learning rate allows for lower learning rates due to the learning rate typically being orders of magnitude lower in the fine-tuning process. One year of data in the Netherlands is used to fine-tune the parameters in the target domain, as seen in figure 3.9, where the aim is to achieve a minimal MSE using the walk-forward validation method. A random search algorithm will perform 500 trials and base its median pruning on the MSE of the forecasts for the next period after a fine-tuning opportunity.

Training & Evaluation

Finally, all the models can be trained and evaluated in the target domain after hyperparameter tuning and training in the source domain. This thesis assesses the differences in performance and training time between cases.

The performance analysis consists of three parts. First, the models will be assessed by their zero-shot forecast accuracy. Second, the forecast made between

Table 3.2: Optimized hyperparameters and their search space for the model in the target domain

Hyperparameter	Value
Learning Rate	$[10^{-8}, 10^{-2}]$
Batch Size	[1, 64]
λ	$[10^{-9}, 10^{-1}]$

fine-tuning opportunities, which is 30 days, will be grouped for assessment. Third, the accuracy of all forecasts made after the first fine-tuning opportunity will be assessed. Furthermore, the training time will be compared. This duration indicates the computational resources necessary for training these models. It allows a fair comparison between all models, considering the trade-off between performance and computational resources.

3.5 Benchmark models

Relevant benchmarks are developed to test the proposed model. First, this section considers a model trained only on data in the target domain. Second, it discusses the development of a model based on physical equations. Finally, a persistence method is examined.

Target model

This section concerns two benchmark models trained on only data in the target domain. The following procedure is similar to figure 3.9 although without the source domain. It consists of one non-physics-informed and one physics-informed model.

These models' hyperparameters should be tuned to allow fair comparison. It follows the same procedure as the transfer learning models. However, more hyperparameters should be tuned in the target domain due to the absence of a corresponding source model. Table 3.3 shows the configurable hyperparameters.

Table 3.3: The configurable hyperparameters and their respective search space

Hyperparameter	Search space
Optimizer	[Adam, RMSProp, SGD]
Nb. of nodes	[4, 1000]
Learning rate	$[10^{-8}, 10^{-2}]$
Batch size	[1,64]
λ	$[10^{-9}, 10^{-1}]$

Transfer model with no weather covariates

To assess the added value of the forecast of weather variables, a model will be designed that uses only historical weather outputs and encodings of the time. The same pipeline used for other transfer models, as depicted in figure 3.9, will be used to design these models.

Physical model

Next, physic-based methods use physics equations to solve for the PV power using the irradiance and other variables. Different methods exist, which differ from each other in the input variables used or assumptions made to simplify the equations. Nevertheless, equation 3.17 characterizes the physics-based models.

$$\hat{P}(t+k) = F(G\hat{H}I_{t+k}, D\hat{N}I_{t+k}, D\hat{H}I_{t+k}, \hat{X}_{t+k}) \tag{3.17}$$

$\hat{X}_{t+k}$ denotes all forecast of input variables, apart from irradiance, necessary for a specific method.

PVWatts is a physical model for a PV module proposed in [98]. *Dobos et al.* explain in [98] how this model works. Most notably, the efficiency curve is an aggregated curve over all inverter models. Additionally, more in-depth physical models exist, e.g., the SAMD model [99] or the model proposed by *Huld et al.* .[100]. Nevertheless, they rely on multiple coefficients, which vary for the inverters and PV panels. Furthermore, these models show only slight improvements in the order of 1 % [101], which leads to the conclusion that irradiance forecast errors will be more significant than model errors for the physics-based PVPF. The PVWatts model is implemented using *pvlib*, where standard values are assumed for the temperature-dependent loss and the nominal efficiency.

Baseline model

Persistence methods exist for PV power, similar to irradiance forecasts. The persistence method is called a 'naïve' method, which involves no forecasting skill. Therefore, persistence forecasting is the baseline model to which other models can be compared to assess the forecasting skill of the proposed methodology. Persistence methods for PVPF utilize the diurnal pattern of solar irradiance or scale the power with the inverse clear sky index. Equation 3.18 gives the diurnal persistence forecast for day-ahead PV power

$$\hat{P}(t+k) = P(t-D+k) \tag{3.18}$$

$\hat{P}(t+k)$ is the PV output forecast for k steps into the future, and D is the period of the observations, which is 24 h for the diurnal cycle.

Current research motivates the use of a smart persistence method [27]. This method persists the clear sky index and multiplies this with the PV power over clear sky conditions. Equation 3.19 shows this equation for day-ahead forecasting.

$$\hat{P}(t+k) = P_{CS}(t+k)k_t^* \tag{3.19}$$

P_{CS} is the PV output under the clear sky conditions, and k_t^* is the clear sky index for PV power and is given by equation 3.20.

$$k_t^* = \frac{P(t)}{P_{cs}(t)} \tag{3.20}$$

Yang et al. propose a convex combination of persistence and climatology for irradiance forecasting in [102]. The motivation for this convex combination applies to solar power forecasting as well. Due to the time of the forecast issue, 00:00, the clear sky PV index can not be persisted. Therefore, I opted to persist the clear sky PV index of the 24h-lagged PV power, which results in a different clear sky PV index for each hour. The 24h-lagged autocorrelation for the power in the source domain, ξ, gives the weight of the persistence forecast. The climatological clear sky PV index, k_{cl}^*, completes the convex combination with a weight of $1 - \xi$. k_{cl}^* equals the average clear sky PV index of the synthetic PV generation in the source domain. Equation 3.21 gives the forecast for the convex combination of climatology and persistence.

$$\hat{P}(t+k) = P_{cs}(t+k)(\gamma k_t^* + (1-\gamma)k_{cl}^*) \tag{3.21}$$

This method could be elaborated on by periodical fine-tuning, similar to walk-forward validation. However, this is too intensive and will not result in significant differences, so it is discarded for this thesis.

3.6 Evaluation

The RMSE is a standard metric for solar power forecasting. Equation 3.22 gives the formula for calculating the RMSE.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{P}_i - P_i)^2}{N}} \tag{3.22}$$

$\hat{P}_i$ and P_i are the i$^{\text{th}}$-hour respective forecast and observation. The RMSE is utilized over other metrics, following the guidelines discussed in [102]. *Yang et al.* motivate the use of the RMSE-based skill score to facilitate comparison between studies in the solar forecasting literature. Equation 3.23 shows how to calculate the skill score.

$$SS_j = 1 - \frac{RMSE_j}{RMSE_{pc}} \tag{3.23}$$

$RMSE_j$ is the RMSE for case j, $RMSE_{pc}$ is the RMSE of the convex combination of persistence and climatology, and SS_j is the skill score for case j.

3.7 Analysis of the results

An analysis should be performed to meaningfully interpret the results obtained by following the methodology explained in previous sections. First, these allow

an estimate of how robust the findings are when perturbations are inserted into the original test setup. Second, they allow for finding possible improvements or weaknesses in the current test setup. The following sections will present the analyses performed on the results.

Different architectures & ML algorithms

The field of machine learning is constantly evolving. The most optimal models will be outdated next year. Hence, this thesis's findings must be reviewed for multiple ML models and architectures. If similar findings are reported for all ML models, these findings can be postulated with improved confidence. However, running a hyperparameter tuning for all these models would be too intensive. The RF and SVM will be assessed on their zero-shot performance for the first four cases to determine if physics-informed variables and reanalysis source training lead to similar results. Only zero-shot forecasts will be compared due to a more involved transfer learning strategy for these models.

The support vector regression (SVR) method resulted from SVM methods, which rely on mapping to a high-dimensional feature space and performing linear regression within that space [37]. Equation 3.24 shows how the forecast is constructed for (SVR).

$$\hat{P}(t+k) = \sum_{i=1}^{n} \tau_i \mathcal{K}(x_i, x_*) + b \tag{3.24}$$

b is a bias, τ_i is the difference between 2 Lagrange multipliers resulting from the optimization problem, and $\mathcal{K}$ is the kernel function performing the mapping. Equation 3.25 gives the mapping for a linear kernel.

$$\mathcal{K}(x_i, x_j) = x_i^T x_j + c \tag{3.25}$$

c is a constant, and the optimization problem to find the weights is regularized with an L2-regularization constant, similar to equation 3.14.

Random Forest regression is an ensemble machine learning method. It is based on tree classification, where the data is split into two branches based on a loss function. Nonetheless, one tree can be prone to overfitting, so random forest regression uses multiple trees that aim to be uncorrelated by sampling a random subset of all input features [103] per tree. At each node in the forest, a subset of the observations is sampled, and a loss function determines the split. The outcome of the independent trees is averaged to arrive at a forecast. This method is prevalent due to the sampling of features reducing the probability of overfitting. A certain amount of papers, as mentioned in chapter 2, have shown that this method outperforms other methods, like SVR and LSTM, for forecasting PV power.

Both models are constructed using the Python library *Scikit-learn* [104]. For SVR, a linear kerner is utilized with default parameters. The RF is constructed with 1000 estimators, which is the amount of trees. It should be noted that it would be appropriate to perform hyperparameter tuning for these models to have a more representative comparison. The hyperparameter tuning is omitted in this thesis due to time constraints.

Impact of fine-tuning

The transferred source model is fine-tuned after each 30-day period. However, due to the seasonality of the PV generation, it is not easy to assess the effects of the periodical fine-tuning. Therefore, the periodical RMSE is divided by the periodical RMSE of the source model, which does not fine-tune on the target data. Although imperfect, the increase in performance over this source model will allow us to estimate the effect of fine-tuning better and decide if improvements to this procedure are necessary.

3.8 Summary

Table 3.4 summarizes all the cases that will be initially investigated using the methodology mentioned previously. Models based on the TL approach are labeled Transf, and models training only on target data are labeled as Targ.

Table 3.4: Summary of all the researched cases

Case No.	Model Architecture	TL	NWP	Physics-Informed	Source dataset
Transf_{np}	LSTM	✓	✓	-	NWP
Transf_{p}	LSTM	✓	✓	✓	NWP
$\text{Transf}_{ra,np}$	LSTM	✓	✓	-	Reanalysis
$\text{Transf}_{ra,p}$	LSTM	✓	✓	✓	Reanalysis
$\text{Transf}_{ra,nw}$	LSTM	✓	-	-	-
Targ_{np}	LSTM	-	✓	-	-
Targ_{p}	LSTM	-	✓	✓	-
Phys	Physical	-	✓	✓	-

Chapter 4

Data

This chapter discusses the relevant data sources concerning PV power forecasting with ML models. The output of residential PV installations, historical NWP runs, and reanalysis data are necessary for training the proposed models. Finding high-quality data sources is a noteworthy challenge for any data-driven research. Therefore, this chapter aims to give an overview of relevant data sources for these three data types.

Additionally, this data contains missing values due to malfunctions of measurement devices or databases. It is well-known that data quality has a significant impact on the performance of machine learning models [105]. Therefore, it is crucial to check the data quality rigorously and impute missing values carefully.

The first three sections describe data sources for reanalysis, NWP, and residential PV output data, respectively. The fourth section gives an overview of the data used in this thesis. Subsequently, the fifth section will discuss the performed data quality check and imputation. Finally, the sixth section will show some insights into the data and use these insights to decide the input features for the models.

4.1 Reanalysis

Reanalysis or retrospective analysis are climate datasets that combine NWP and historical measurements intending to estimate the past atmospheric states as close to reality as possible [106]. It is invaluable for assessing the state of the climate [107], and some applications label reanalysis data as the ground truth to train the model, e.g. [108]. For this thesis, it could be interesting to train the source model on reanalysis data instead of NWP due to the higher accessibility of this type of data. Therefore, this section lists the most common reanalysis data.

4.1.1 ERA5

The European Center for Medium-Range Weather Forecasts (ECMWF) produced the fifth generation of European Reanalysis (ERA5) [107]. It replaced ERA-interim with the motivation that the underlying NWP models significantly improved in modeling atmospheric processes. It uses the Integrated Forecasting System (IFS) to provide

hourly data from 1940 onwards with a spatial resolution of 0.25°, which translates to 31 km.

In 2021, ECMWF introduced the ERA5-Land dataset [109]. It provides the surface components of the ERA5 dataset, with an increased spatial resolution of 9 km instead of 31 km spatial resolution of ERA5. These datasets can be downloaded via the Climate Data Store of Copernicus [110].

4.2 Historical NWP datasets

This section discusses the different datasets that contain historical NWP model runs. These datasets are more challenging to find as the main interest is to provide the actual weather data. However, there are some archives available that contain NWP model runs from the past. Table 4.1 gives an exhaustive overview of available archives. It doesn't claim to be complete but lists the most representative archives.

4.2.1 TIGGE dataset

The International Grand Global Ensemble (TIGGE) archive consists of ensemble forecast data of 13 NWP centers [111]. It was one of the components of "The Observing System Research and Predictability Experiment" (THORPEX), which was a research collaboration that lasted from 2005 until 2014 [112]. Although this research collaboration has ended, the TIGGE dataset is still online and expected to be available to the general public. It consists of historical NWP forecasts starting in 2006 up until forecasts 48h before real-time. The high amount of data makes this data extremely useful for training a model to handle NWP forecasts. *Bi et al.* used it to compare the skill of *Pangu-Weather* [18] and for research concerning multimodel ensemble forecasts [113]. The spatial resolutions differ for the different models but range between 9 km for the ECMWF and 104 km for the Brazilian meteorological institution CPTEC. However, the main disadvantage of this data is that these global datasets are only available at 6-hourly temporal resolution. A spatiotemporal interpolation of weather variables using neural networks, similar to [114], could be considered but is outside the scope of this thesis. So, in conclusion, this resolution is not granular enough to predict the solar power generation for hourly intervals and, as such, if not further considered.

4.2.2 Met Office

The Met Office is the national meteorological institution of the United Kingdom. It provides the historical outputs of different NWP models divided by the geographical area covered. As such, the Unified Model for Europe (UM-EURO) provides NWP data, which covers the whole of Europe [115]. The dataset ranges from 16 March 2016 until the end of 2021. It has an hourly temporal resolution and a spatial resolution of 0.04 degrees, which translates roughly to a resolution of 4.5 km. The UM runs a global coverage model for which the resolution is 0.234°, which roughly translates to 26 km [116]. The model updates its forecasts every 6 hours. These model updates

are all stored in the CEDA Archive, which belongs to the Natural Environmental Research Council's Environmental Data Service. It aims to provide storage for data concerning environmental research and make it accessible for research that uses this data. As such, it also holds data on the Met Office NWPs. The UM-EURO dataset has a size of 83.3 TB, which makes acquiring this data quite intensive. However, this has proven to be one of the most accessible open-source hourly NWP archives online.

4.2.3 Icosahedral Nonhydrostatic model

The German weather service DWD provides the output of their European model, Icosahedral Nonhydrostatic (ICON) EU, and the global model, ICON Global. These models offer hourly forecasts until 72 hours at a spatial resolution of, respectively, 6.5 km and 13 km [117]. DWD provides these forecasts daily but doesn't store them [118]. Climate Fix stores this data and has a dataset that goes back to 2020 for solar and wind forecasting variables. From March 2023, the database stores all output variables of the ICON model. Open Climate Fix is a non-profit organization focusing on reducing GHG emissions as fast as possible [119]. It provides open-source forecasting models and datasets [120], [121].

4.2.4 ECMWF Integrated Forecast System

The Integrated Forecast System (IFS) of the ECMWF is an exemplary NWP algorithm. It consists of multiple forecasting systems connected in various ways. One subsystem of the IFS is the High-Resolution Forecast (HRES), which gives up to 10 days ahead forecasts at a spatial resolution of 9 km. These are deterministic forecasts with an hourly forecast step until 90 h [122]. Another subsystem is the Ensemble Forecast Suite (ENS), which delivers probabilistic forecasts at similar spatial resolution since its update in 2023 [123] and a similar forecasting step. The ECMWF forecast is hailed as one of the most performant NWP algorithms worldwide. As such, the IFS is also the baseline model for *WeatherBench 2*, which introduces a benchmark to analyze the performance of new data-driven weather models [108]. However, these ECMWF products are not publicly accessible and thus require a license. This license costs 3.000 €/year, making it quite expensive for research purposes. There exists an operational archive [124] which stores their model runs at 00:00 and 12:00 UTC. This archive goes back to 1985, making it the most extensive dataset. However, the data only has an hourly step from 16 November 2011, which makes it still the most extensive NWP archive. However, its annual license fee of €3.000 is a significant disadvantage.

4.2.5 Machine learning algorithms

Recently, machine learning has been introduced as a more than capable alternative for NWP to forecast the future state of the atmosphere. Examples of these are Pangu-Weather [18], by Huawei, and GraphCast [125], by Google Deep Mind. These models are open-source, available on *Github*, and can both be trained if adequate

Table 4.1: Overview of different historical NWP archives

Archive	Spatial Resolution	Temporal Resolution	Coverage Region	Time Range	Data Format
TIGGE	9 - 104 km	6-hourly	Global	2006 , D-2	GRIB2
UM: Euro	4.5 km	hourly	Europe	2016 - 2021	GRIB2
UM: Global	26 km	hourly	Global	2016 - 2021	GRIB2
ICON-EU	6.5 km	hourly	Global	2020 - D	Zarr
ICON-Global	13 km	hourly	Global	2020 - D	Zarr
IFS HRES	9 km	hourly [2]	Global	1985, D	GRIB2
Pangu	[1]	1/6/24-hourly	Global	[1]	[1]
GraphCast	[1]	6/24-hourly	Global	[1]	[1]

[1] Dependent on available data
[2] Starting from 16 November 2011

input data is provided. As such, the spatial resolution solely depends on the input data provided. They provide pre-trained models because these are quite intensive to train. Using pre-trained models, one could simulate the historical NWP outputs with appropriate inputs. In opposition to GraphCast, Pangu-Weather can provide an hourly model that would be suited for this thesis. However, incorporating an additional deep learning network vastly increases the complexity and time required to run. More importantly, only a limited amount of variables are currently available with these models.

4.2.6 Additional historical NWP sources

More meteorological institutions provide NWP forecasts, although with a limited geographical coverage. Additionally, there exist open-source NWP models Weather & Research Forecasting, which allows users to have their own NWP model [126]. For example, Rutgers University Weather Weather & Research Forecasting has NWP outputs for the region around New Jersey [127]. Furthermore, *Blaylock et al.* provides an open-source Python package *Herbie* [128], which provides NWP archives from the National Oceanic and Atmospheric Administration. This package seems to be a promising candidate. However, I found this to late during my thesis.

4.3 Residential PV generation

4.3.1 Photovoltaic Geographical Information System

The Photovoltaic Geographical Information System is an online tool for solar resource assessment developed by the European Commission Joint Research Centre [30]. Figure 4.1 gives a snapshot of the online tool, which provides various insights for estimating the performance of a PV installation. One of the options is to obtain a time series of solar power production of a simulated PV installation for which the

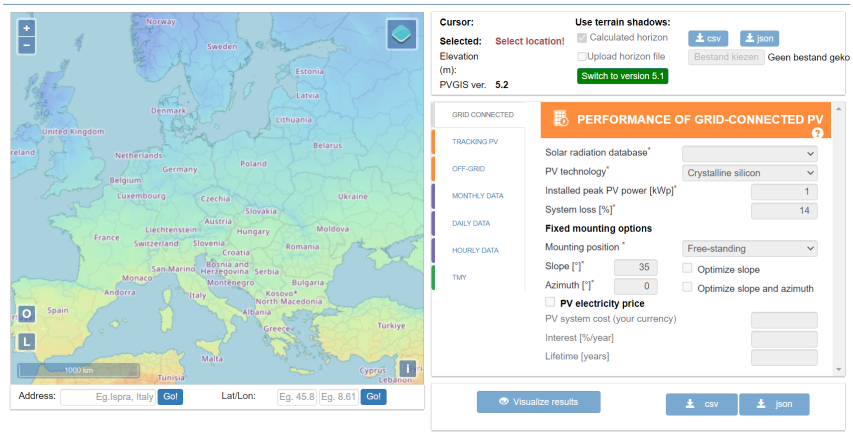


FIGURE 4.1: A snapshot of the PVGIS online tool. Source: PVGIS [30].

location, azimuth, tilt, and peak power can be specified. It estimates the radiance at the selected area using past satellite images, which allows for higher accuracies compared to irradiance values provided by reanalysis. Consequently, PVGIS can simulate the historical solar power of an existing installation if the specifications correspond.

4.3.2 UK

Apart from the historical ICON outputs, Open Climate Fix provides PV generation measurements of 1311 PV systems from 2018 to 2021 located in the UK [129]. They deliver the measurements at a temporal granularity of 2,5 and 30 min and metadata for each installation consisting of orientation, tilt, installed PV power, and location. However, closer inspection of the 30-minute data reveals that several values are missing, and many NaN values are replaced by zeroes in the dataset, resulting in a more challenging data quality check. Moreover, community posts mention incorrect metadata for some installations. Nonetheless, open access to this enormous amount of actual PV generation is infrequent.

4.3.3 NL

A dataset from the Netherlands was made available for this thesis. It consists of the PV generation of 50 households in the Netherlands starting in 2020 and ending in August 2021, with a temporal resolution of 15 minutes. Some installations contained limited data for 2019, but this was not the case for all installations. This data was of high quality, reporting no missing values and only a limited number of outliers.

4.3.4 Costa Rica

This data concerns a PV installation of a university building in San José, Costa Rica. The installation consists of 17 panels of type *CS6U-325P*, with a peak power 325 Wp and a SolarEdge *SE7600A-US* inverter. It spans from 2019 to 2021 and has

Table 4.2: Overview of the residential PV installations

Name	Place	Coördinate [°]	PV Rating [kW]	Inverter Rating [kW]	Tilt [°]	Azimuth [°]
NL_1	Nieuwegein, NL	52.05, 5.07	3.10	2.50	33	5
NL_2	Koudum, NL	52.91, 5.45	2.48	2.00	33	44
NL_3	Stavoren, NL	52.88, 5.36	2.79	2.79	30	76
NL_4	Houten, NL	52.02, 5.18	2.40	2.00	36	-9
NL_5	Jistrum, NL	53.21, 6.07	2.48	2.00	43	-15
NL_6	Chaam, NL	51.5, 4.86	2.40	2.40	30	90
NL_7	Drachten, NL	53.11, 6.10	2.52	2.00	28	-1
CR	San José, CR	9.94, -84.04	5.53	5.53	8.5	-90
UK	Liverpool, UK	53.4, -2.91	2.25	2.25	36	45

an hourly resolution. This dataset was found via the PV power database of the TU Delft [130]. This data did not contain any missing values and a limited amount of outliers.

4.3.5 Other locations

Other PV power datasets, although scarce, indeed exist. However, this thesis focuses on residential PV installations. Furthermore, the orientation and tilt angle of the installation are necessary to estimate the GTI and have accurate source domain generation. The requirement of metadata limited the useable PV power dataset for this thesis. The TU Delft lists available PV power datasets at [131].

4.4 Data Description

This section describes the datasets considered for this thesis. First, it briefly describes the target residential PV installations considered. Second, it discusses the considered providers of the weather covariates.

4.4.1 Residential PV installations

The models should be tested on various residential PV installations for a higher statistical significance. It is opted to test on nine residential PV installations, which results from a tradeoff between the statistical significance of the results and the time to gather and process all this data, for which the quality can be a concern, dependent on the provider. Table 4.2 summarises these nine sites and their most essential metadata. Seven tested sites are in the Netherlands due to the high data quality. The methodology is slightly adjusted for the UK dataset as the data didn't go to 2021. The source data for the final training spans 2016-2019 instead of 2017-2020, which is the period for other installations. Note that this period is always 1 May due to the Met Office data being available from 1 May 2016.

4.4.2 Weather Covariates

For the weather covariates, historical NWP runs and reanalysis data are necessary. Except for Costa Rica, the historical NWP runs are the UM-EURO data from the CEDA archive. The historical NWP outputs for Costa Rica result from the Met Office's UM-Global, which does not provide the DNI and DHI. These are estimated using the DISC model, explained in chapter 3. The reanalysis data is obtained from the *Open-Meteo* platform [132], which blends the ERA5 and ERA5-Land datasets to provide the historical weather.

4.5 Data Acquisition, Quality Check & Preparation

An essential step in all data-driven research is checking data quality, cleaning the data, and, if necessary, imputation or transformations on input data. This section covers the data quality check and cleaning process for PV generation measurements and weather covariates. Next, it discusses transformations on the input features to make the data uniform across different data sources.

4.5.1 Data Quality Check

PV generation measurements

PV generation measurement datasets are not readily available. Due to malfunctions of measurement devices or the PV installation, data at specific timestamps can be missing. Missing data should be imputed via a thorough process that results in consistent values. The data quality check is essential to obtain a performant model after training. Figure 4.2 shows the procedure I followed for all PV production datasets.

The first step tries to eliminate NaN entries. For each NaN value, it is first checked if the preceding and succeeding values exist. If this is the case, a linear interpolation is done to fill up the non-existent values. If more than one subsequent value is missing, the step checks the timestamp of the non-existing values. If these timestamps are at night, it replaces the NaN values with 0 kW. There may still be NaN values after this step. More advanced imputation techniques can be deployed using 24h-shifted power values or clear-sky index interpolation. However, except for the UK, the data sources used in this thesis didn't contain any NaN values after these two steps, so extra steps were unnecessary. The UK dataset transferred NaN values as zeroes, which made the data quality check more intensive. For this dataset, days with a total production of 0 kWh are eliminated. This means that the day after doesn't have an accurate input, as the 24h-lagged power is actually lagged 48 hours. However, I assumed that this would be no problem as these lagged values are independent of other input features.

The subsequent step verifies if the power measurements stay in the feasible range for PV power, i.e., between 0 kW and the installed PV power. It is worth mentioning that limiting negative output to zero might not be the most accurate imputation.

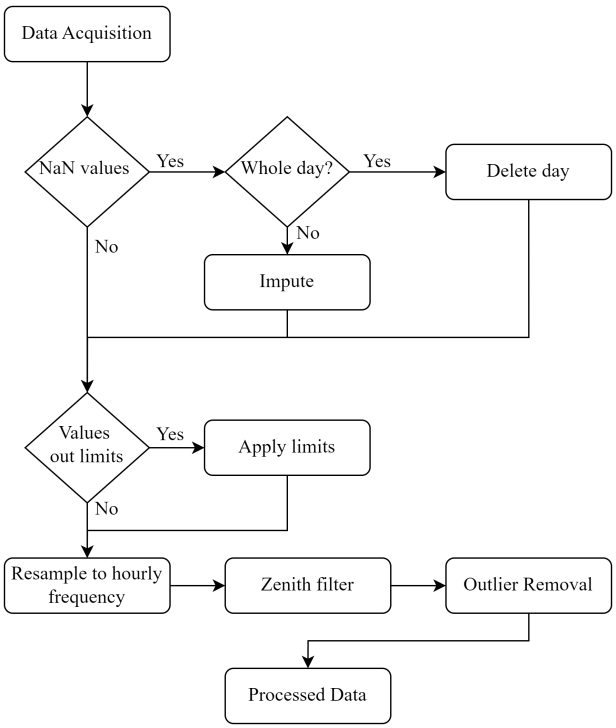


Figure 4.2: Schematic of the data cleaning process for PV measurements

Regardless, there were no negative values in the used datasets, so this remark is not considered.

Afterward, the production measurements are resampled to hourly frequency. It is necessary to resample to an hourly frequency due to the hourly frequency of the synthetic PV generation. Subsequently, a zenith filter is applied to the data. The zenith filter sets the PV power for observations where the sun's zenith angle is greater than 85° to zero. This procedure is customary for solar forecasting as a low sun position leads to inaccurate irradiance estimates. Finally, an outlier procedure is followed, defining outliers as observations where the power is zero but the global irradiance exceeds a threshold of 100W. The resulting dataset is ready to be used in the ML pipeline.

Numerical Weather Predictions

The standard format for NWP is the Gridded Binary or *GRIB* format, which most meteorological institutions use to represent the 2D output of the NWP model. For this thesis, only the values at specific locations are of interest. Therefore, the data acquisition algorithm reads the *GRIB* file per day and variable, selects the output at the nearest grid point to the locations of interest, and then stores this data locally. As one file contains all values for a day, missing timestamps span the whole day. Hence, the data cleaning process is simplified by discarding NaN values, meaning

the model does not include these days for its training. Similar to the process for PV generation, the zenith filter is applied for all irradiance components.

Reanalysis

Open-Meteo provides reanalysis data and makes them more accessible, as most of it is also stored in the *GRIB* format on databases. Like the NWP data, the only step is checking for NaN values and applying a zenith filter.

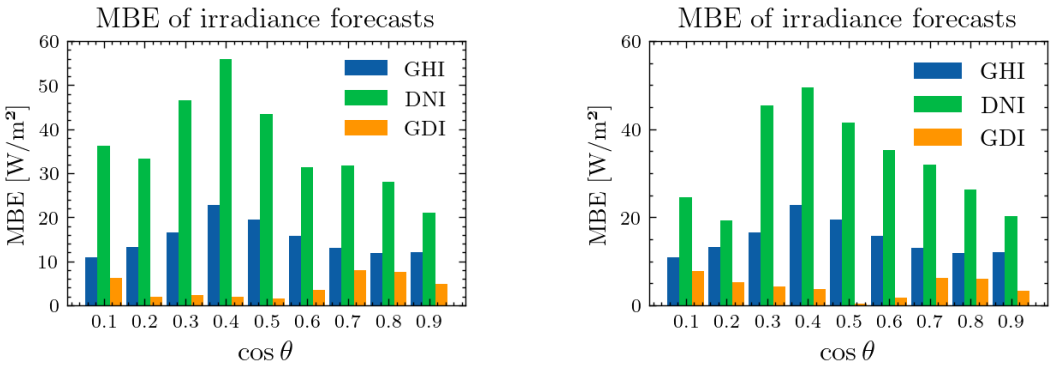
4.5.2 Transformation and Scaling of the data

The data is not yet ready to be used for the regression problem. First, the data should be consistent over all datasets, meaning the same variables should be present in the same unit. Table 4.3 summarizes all possible input variables, their symbol, and units.

TABLE 4.3: All possible covariates and their units as input for the model

Symbol	Variable	Unit
T	Temperature	K
p	Pressure at MSL	Pa
GHI	Global Horizontal Irradiance	W/m^2
DHI	Diffuse Horizontal Irradiance	W/m^2
DNI	Direct Normal Irradiance	W/m^2
q	Relative Humidity	%
V	Wind Speed	m/s
α	Wind Direction (Sine/Cosine)	–
C	Total Cloud Amount	%
P_{24h}	24h-lagged PV power	W
M	Month of the year (Sine/Cosine)	–
H	Hour of the day (Sine/Cosine)	–
GTI	Global Tilted Irradiance	W/m^2
P_{CS}	Clear-Sky Power	W
T_{PV}	Temperature of PV panel	K

Transformations of some features are necessary to make them uniform among different providers. For example, suppose the output of a NWP model gives the horizontal, u, and vertical components, v of the wind velocity. In that case, it should



(A) Original (B) Corrected with closure equations

Figure 4.3: MBE of the irradiance grouped by the cosine of the zenith angle

be transformed to wind speed and direction using the following equation:

$$V = \sqrt{u^2 + v^2}$$
$$\alpha = \tan^{-1}(\frac{v}{u})$$

Next, for the Met Office's NWP output, direct irradiance was defined with respect to a horizontal surface instead of a perpendicular one. Furthermore, it was unclear if the irradiance was defined as an hourly average or instant irradiance. Therefore, I opted to derive the DNI using the closure equation, given by equation 4.1.

$$GHI = DNI + DHI \cos\theta \tag{4.1}$$

Furthermore, DHI values higher than the GHI were limited by the GHI. This method led to smaller differences with the irradiance provided by reanalysis data. Figure 4.3 shows the Mean Biased Error (MBE) of the irradiance components grouped by the cosine of the solar zenith angle.

If all possible covariates are included and uniformized, the data is scaled to allow for improved performance of the ML models. A min-max scaling is applied to the data because an LSTM can be sensitive to the relative scale of the input features. Equation 4.2 describes the min-max scaling.

$$X_{mm} = \frac{X - min}{max - min} \tag{4.2}$$

Where X_{mm} is the scaled input for the ML model, and min and max are, respectively, the minimum and maximum of the feature in the considered source domain.

4.6 Data Visualization And Feature Selection

This section will use visualizations to give insight into the considered datasets and provide a motivation for the feature selection. First, the similarity between the source

domain's synthetic PV generation and the target domain's actual PV generation will be investigated. Second, the accuracy of Met Office's NWP output will be assessed using the ERA5 data. Finally, I will use the insight into the correlation of weather covariates and synthetic PV generation to select the input features for each model.

4.6.1 Similarity of PVGIS and actual PV generation

To have a model only trained on source domain data that can perform zero-shot forecasts with acceptable accuracy, the source domain PV generation, the PVGIS data, should closely resemble the actual PV generation. PVGIS provides synthetic, hourly PV generation until the end of 2020, which allows us to study this similarity. However, it should be noted that this information cannot be utilized to develop the model, as this would be data leakage, where information that would be unknown at the time of forecasting is used. Nonetheless, it can help interpret the results.

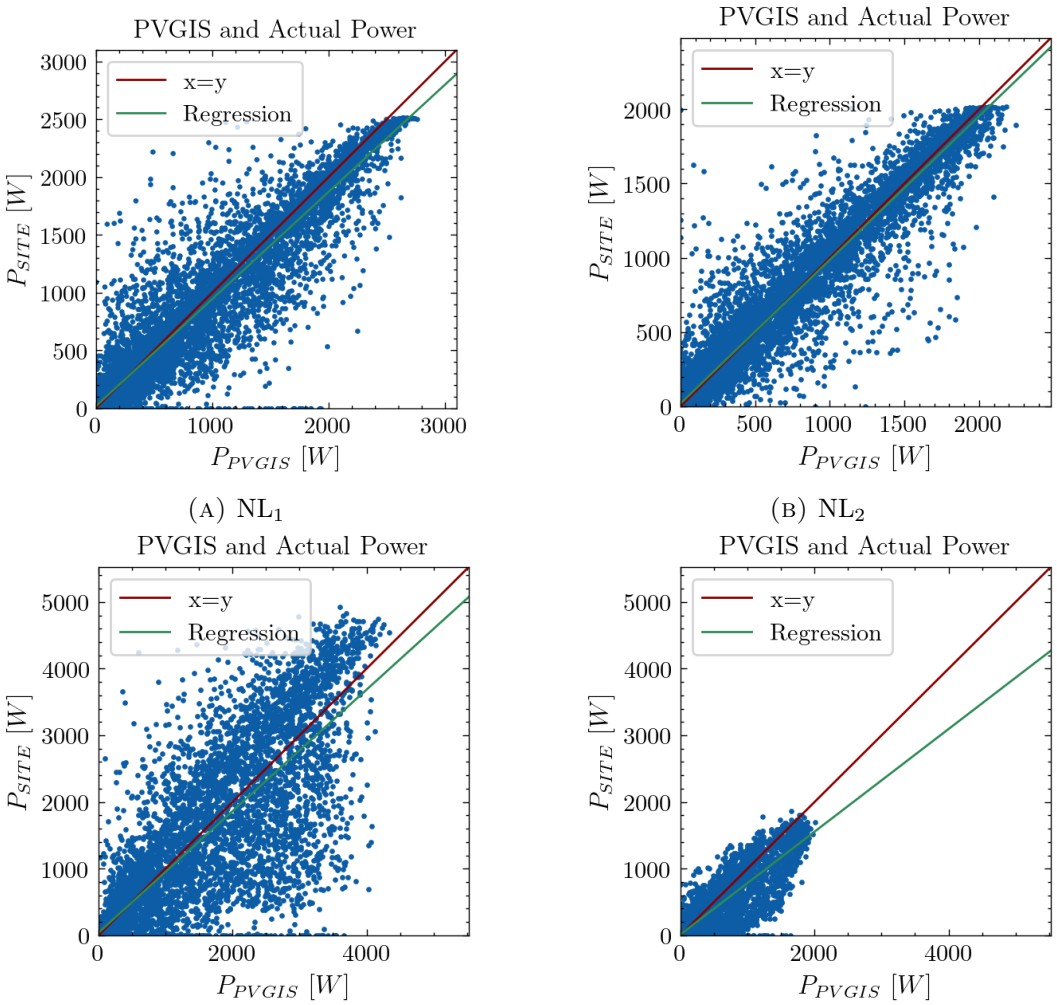
For this purpose, figure 4.4 shows scatter plots of the PVGIS and actual PV generation for all the sites. It could be deduced that the generation data shows higher similarity for the Netherlands sites, as opposed to the sites in Costa Rica and the UK. The slope, m, and coefficients of determination, R^2, of a linear regression between both PV generations, given in table 4.4, confirm this hypothesis.

For Costa Rica, the database used by PVGIS to calculate the synthetic PV power could explain this lower similarity. The database used for Costa Rica is the irradiances as estimated by irradiances given by ERA-5 due to satellite-derived irradiances only being available until 2015. It can be concluded that due to a less accurate irradiance estimate via reanalysis data, as mentioned previously, the accuracy of PVGIS decreases. However, this argument does not hold for the UK. Additionally, the actual slope of the regression line is significantly lower than others. One hypothesis could be that the losses PVGIS estimates to be 14 % are higher for this site. Another hypothesis is that the provided installed PV power or other metadata about the PV installation is inaccurate and should be lower. Nonetheless, I included this site in further analysis due to interest in how all cases handle a lower similarity between source and target domain.

At last, these scatter plots reveal which sites have an inverting rating lower than the installed PV power. The actual PV generation of both sites in the Netherlands seems to be cut off at a particular value. As no metadata is provided about inverter ratings, I assumed that these cut-off values are the inverter ratings, which amount to respectively 2.5 kW and 2 kW for site NL_1 and NL_2. A similar cut-off characteristic was observed for the other Netherlands installations. The same procedure was used for these sites.

4.6.2 Similarity ERA-5 and Met Office NWP

Following the previous section, comparing the reanalysis dataset, ERA-5, and the NWP dataset of Met Office could also be interesting. For this purpose, figure 4.5 shows a scatter plot of the same features for both providers. Some features like temperature and pressure are pretty similar. However, significant differences exist



(a) NL_1 (b) NL_2

(c) CR (d) UK

Figure 4.4: Scatter plot of PVGIS and actual PV generation

between the total cloud amount and irradiances. It is not entirely correct to call these differences errors for two reasons. Firstly, both models' resolutions are different, which can lead to significant deviations for unstable features like irradiance and clouds. Second, reanalysis datasets do not provide entirely accurate irradiance estimates due to parametrizing physical interactions in the atmosphere, which are less interesting for atmospheric modeling. Nonetheless, reanalysis data will be closer to the ground truth. The deviations between ERA5 and UM-EURO are substantial and show the added value of utilizing the learning capabilities of an ML-based methodology that can anticipate errors in the NWP.

TABLE 4.4: Slope and coefficient of determination for linear regression line

Site	m	R^2
NL_1	0.9268	0.9557
NL_2	0.9675	0.9707
CR	0.9114	0.8661
UK	0.7700	0.8926
NL_3	0.9750	0.9738
NL_4	1.0029	0.9717
NL_5	0.9447	0.9674
NL_6	1.0207	0.9574
NL_7	0.9104	0.9566

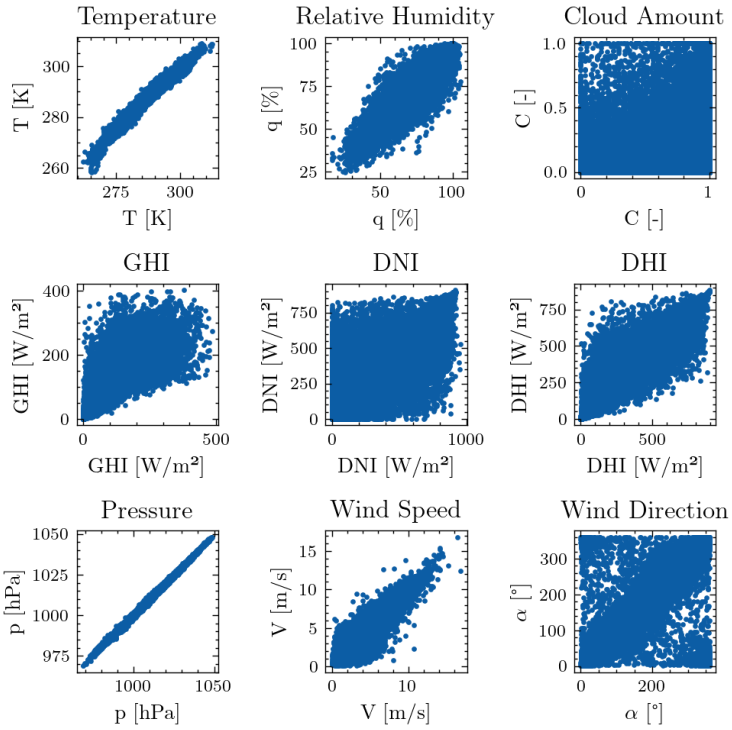


FIGURE 4.5: Scatter plot of the features given by ERA-5 and Met Office's NWP

4.6.3 Features Correlation & Selection

This section aims to visualize the relation between the possible covariates and the PV generation to select only relevant features as input for the models. To ensure consistency, this will be based on correlations between the reanalysis data and the synthetic data, as it is assumed that no actual PV generation data is available at the present moment.

Figure 4.6 shows the correlation matrix of the considered input features and the synthetic PV power generated by PVGIS for the site NL_1. Considering the correlations, physical insights, and limiting the number for computational intensity, table 4.5 shows the input features for all distinct cases. The wind speed shows a low correlation with the power output. However, physical insights tell us that the wind speed influences the temperature of the PV panel and is thus included. The physics-informed cases include the engineered features, GTI and T_{PV}, but omit the P_{CS} due to its irrelevance if other irradiance-related features are provided to the model. Furthermore, the encoded month and hour of the day are included in the model with no weather covariates to let the model learn about the seasonality of the PV power. However, when weather covariates are utilized, this seasonality is also encapsulated in the irradiance forecast. Therefore, these variables are assumed to be obsolete in cases with weather variables.

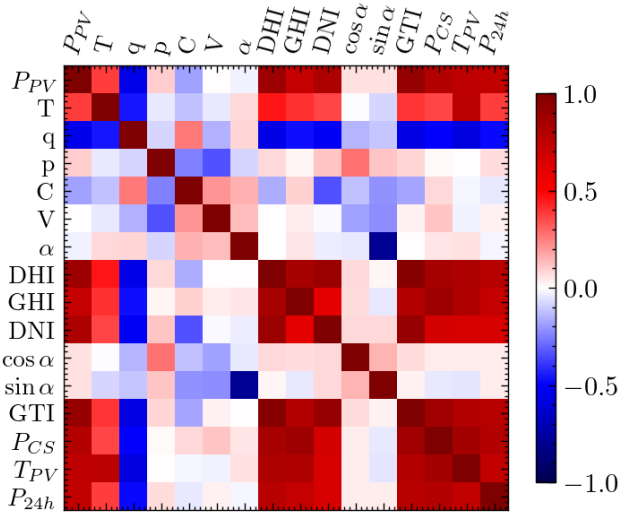


Figure 4.6: Correlation matrix of all features and PVGIS power for NL_1

TABLE 4.5: Input features for the distinct cases

Variable	Non-physics-informed	Physics-informed	No weather
T	✓	✓	-
p	-	-	-
GHI	✓	✓	-
DHI	✓	✓	-
DNI	✓	✓	-
q	✓	✓	-
V	✓	✓	-
α (Sine/Cosine)	-	-	-
C	-	-	-
P_{24h}	✓	✓	✓
M (Sine/Cosine)	-	-	✓
H (Sine/Cosine)	-	-	✓
GTI	-	✓	-
P_{CS}	-	-	-
T_{PV}	-	✓	-

Chapter 5

Results

This section will report the results obtained by following the methodology from chapter 3. The first section presents the results of the source model. These results are the obtained hyperparameters after optimization, the accuracy of the model on test data in the source domain, and the time to train the source models. The following section will discuss the target domain models, splitting the evaluation of the model into the performance for zero-shot forecasts for the first 30-day period and the performance for the remainder of the year with fine-tuning opportunities. The subsequent section will present an analysis of the results. This analysis compares other ML models, the RF and SVM, and examines the influence of NWP errors on the model's performance. The machine learning networks are trained using a GPU, *Nvidia* T400 4GB, allowing faster training times.

5.1 Source model

As mentioned, the source model should be designed to achieve a robust, performant model. This section will show the results for the different cases.

The first step was to tune the hyperparameters for the separate cases. The most optimal trials had *Adam* as optimizer. Due to this dominance, the hyperparameter tuning process was reiterated with *Adam* fixed as optimizer.

Table 5.1 shows the optimal hyperparameters for each case after a random search. Hyperparameter tuning of the non-physics-informed and physics-informed cases results in the same hyperparameters. Furthermore, the cases where the source model is trained on reanalysis data require more units.

After the hyperparameter tuning, the source model is trained on three years of data in the source domain, with a 80 % shuffled train-test split until the day before the first day of forecasting for the target domain. Table 5.2 presents the RMSE on this test data for each case and site.

A comparison of the RMSE shows that source models trained on reanalysis data achieve lower RMSE, although the differences are minor. The RMSE is quite similar for the non-physics-informed and physics-informed cases, although it is generally lower for the non-physics-informed case. The duration of the training process measures the

Table 5.1: Optimal hyperparameters for each case (*Fixed hyperparameters*)

Case	Optimizer	Nb. of layers	Nb. of nodes	Learning Rate	Batch size	λ
Trans_{np}	*Adam*	1	162	1.81×10^{-3}	39	6.79×10^{-8}
Trans_{p}	*Adam*	1	162	1.81×10^{-3}	39	6.79×10^{-8}
$\text{Trans}_{ra,np}$	*Adam*	1	799	5.43×10^{-4}	106	9.47×10^{-9}
$\text{Trans}_{ra,p}$	*Adam*	1	799	5.43×10^{-4}	106	9.47×10^{-9}
Trans_{nw}	*Adam*	1	533	3.10×10^{-4}	107	1.65×10^{-9}

Table 5.2: RMSE on the test data of the source domain for the different locations

Site	Transf_{np}	Transf_{p}	$\text{Transf}_{ra,np}$	$\text{Transf}_{ra,p}$	Transf_{nw}
NL_1	229.56	237.05	**226.64**	229.56	349.50
NL_2	198.36	192.82	**183.16**	185.44	250.37
NL_3	189.74	187.42	**177.69**	196.22	255.57
NL_4	182.79	187.13	192.64	**179.34**	279.53
NL_5	197.45	205.64	197.09	**194.67**	276.81
NL_6	145.41	142.20	146.40	**141.17**	199.10
NL_7	181.28	179.02	186.16	**178.76**	264.79
CR	266.17	288.73	**214.92**	222.68	311.78
UK	174.38	181.59	**166.12**	174.79	248.71

computational intensity. Table 5.3 summarizes the average training time over all sites. These show a trend that could be expected with the results of the hyperparameter tuning, where the reanalysis-trained models take longer due to more units.

Table 5.3: Average Training time of the source model

Case Nb.	Training Time [s]
Transf_{np}	20.01
Transf_{p}	20.15
$\text{Transf}_{ra,np}$	50.04
$\text{Transf}_{ra,np}$	50.33
Transf_{nw}	26.66

5.2 Target model

The following part concerns the evaluation and training of the model in the target domain. As mentioned, hyperparameter tuning is necessary for a model that accurately fine-tunes to the new available target data. First, a summary of the hyperparameter tuning results will be discussed. Second, the performance and training time of the

models in the target domain will be addressed. The performance will be split up into zero-shot accuracy and the accuracy of the walk-forward validation starting from one month of training data.

Table 5.4 shows the optimized hyperparameters, where the hyperparameters that result from the source model are put in italic font. This table indicates that two parameters influencing the overfitting have changed compared to the source domain. The learning rate is orders of magnitude smaller than the learning rate in the source domain. Furthermore, except for Trans_{np}, the optimal weight decay has increased significantly.

TABLE 5.4: Optimal hyperparameters for each case (*Fixed hyperparameters*)

Case No.	Optimizer	Nb. of layers	Nb. of nodes	Learning Rate	Batch size	λ
Trans_{np}	*Adam*	*1*	*162*	9.16×10^{-6}	56	2.18×10^{-8}
Trans_{p}	*Adam*	*1*	*162*	2.67×10^{-6}	55	1.65×10^{-4}
$\text{Trans}_{ra,np}$	*Adam*	*1*	*799*	1.46×10^{-6}	11	1.46×10^{-6}
$\text{Trans}_{ra,p}$	*Adam*	*1*	*799*	1.67×10^{-5}	36	1.54×10^{-6}
Trans_{nw}	*Adam*	*1*	*533*	2.24×10^{-7}	3	2.21×10^{-8}
Targ_{np}	*Adam*	*1*	593	7.95×10^{-4}	46	4.58×10^{-5}
Targ_{p}	*Adam*	*1*	623	5.88×10^{-4}	26	4.45×10^{-5}

5.2.1 Zero-shot accuracy

The transfer model is assessed by utilizing the RMSE-based skill scores. The RMSE will be omitted because the skill score leads to the same conclusions and allows for comparison over different sites. Appendix A reports the RMSE, MAE, and correlation for the interested reader. Table 5.5 shows the skill score of the zero-shot forecasts of each model at each site. First, the models trained with historical NWP runs outperform the benchmarks for all sites. Second, one of the benchmark models outperforms the reanalysis-trained models only for NL_2, CR, and UK. For the UK sites, the negative skill scores show that the convex persistence-climatology baseline outperforms the reanalysis-trained models. Third, lower skill scores are observed for Costa Rica and the UK, sites with lower resemblance between source and target domain. Costa Rica's sunny climate results in less variable PV power generation. This decrease in variability leads to a higher accuracy of the persistence method, making it harder to outperform the baseline of the skill score. Additionally, the forecast of weather covariates resulted from a UM-Global with a lower spatial resolution. Finally, NL_6 stands out due to the reanalysis-trained models' higher accuracy than the historical NWP-trained models and an exceptionally high skill score for the physical model.

Figure 5.1 presents a scatter plot of the source and target domain resemblance and the zero-shot skill score of model Transf_{np}. The absolute deviation from unity of the slope of the fitted curve, explained in section 4.6.1, is chosen to characterize the similarity of the source and target domain. The figure shows that, except for

Table 5.5: Skill score of zero-shot forecasts

Site	Transf_{np}	Transf_{p}	$\text{Transf}_{ra,np}$	$\text{Transf}_{ra,np}$	Transf_{nw}	Phys
NL_1	0.3228	**0.3274**	0.1870	0.2009	0.0444	0.1307
NL_2	**0.3968**	0.2967	0.1769	0.1963	0.1586	0.2308
NL_3	**0.3810**	0.3668	0.2694	0.2831	0.1075	0.1646
NL_4	0.2758	**0.3714**	0.2185	0.2138	0.0616	0.1864
NL_5	**0.3769**	0.3710	0.2739	0.2728	0.0911	0.1229
NL_6	0.3899	0.3934	**0.4375**	0.4123	0.0956	0.3436
NL_7	**0.2624**	0.2464	0.2139	0.2188	0.0898	-0.1084
CR	**0.1859**	0.1742	0.1701	0.1471	0.1705	-0.4055
UK	0.0085	**0.1114**	-0.0421	-0.1044	-0.0278	-0.2324

one outlier, site NL_4, there is a correspondence between the resemblance and the skill score.

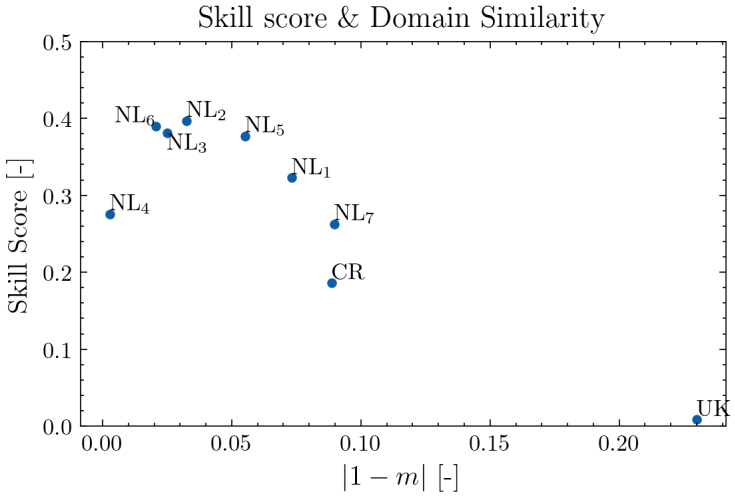


Figure 5.1: Scatter plot of the source and target domain resemblance and the zero-shot skill score of Transf_{np}

5.2.2 Walk-forward validation

Next, these models and two additional models only trained in the target domain, cases Targ_{np} and Targ_{n}, are compared using the walk-forward validation scheme. Table 5.6 summarizes the skill score for all sites.

The skill scores for the proposed cases are higher than all benchmarks for the sites in the Netherlands. For Costa Rica, the physics-informed transfer model performs worse than one of the benchmarks. In the UK, a target model performs slightly

better than the transfer models. Sites with a lower resemblance between source and target domain show more difficulties with outperforming the benchmarks.

Comparing the skill scores between cases shows that physics-informed cases perform worse overall, although with only a small skill score decrease. Physics-informed cases have a slight negative influence on the accuracy of the LSTM model. The reanalysis-trained models show insignificant skill score differences over all sites, even performing better for some sites.

Finally, one can notice that the site NL_6 performs substantially worse than other sites in the Netherlands. The following section will delve deeper into the reason for this significant performance drop.

TABLE 5.6: Average skill score of the model in the target domain, excluding the first month

	Transf_{np}	Transf_{p}	$\text{Transf}_{ra,np}$	$\text{Transf}_{ra,np}$	Transf_{nw}	Targ_{np}	Targ_{p}	Phys
NL_1	**0.3887**	0.3822	0.3861	0.3863	0.1019	0.3075	0.3545	0.1850
NL_2	0.3995	0.3849	**0.4038**	0.3991	0.1125	0.3191	0.3333	0.2401
NL_3	**0.4393**	0.4307	0.4239	0.4194	0.1300	0.3468	0.3615	0.2929
NL_4	**0.4150**	0.4055	0.4000	0.4068	0.0976	0.3690	0.3771	0.2750
NL_5	**0.3796**	0.3747	0.3777	0.3744	0.0867	0.3170	0.3480	0.1358
NL_6	**0.3216**	0.3191	0.3196	0.3187	0.1208	0.2386	0.1864	0.1605
NL_7	**0.3799**	0.3598	0.3649	0.3611	0.0944	0.2821	0.3406	0.0626
CR	**0.2824**	0.2611	0.2799	0.2767	0.1953	0.2594	0.2751	-0.1815
UK	0.3723	0.3337	0.3857	0.3579	0.1558	**0.3866**	0.3737	-0.0142

The models are compared in terms of computational resources. The average training times in table 5.7 show no differences between the non- and physics-informed cases. This was already expected during the design of the source model due to the same architectures resulting from a random grid search.

TABLE 5.7: Average training time of the model in the target domain

Case	Training Time
Transf_{np}	3.75
Transf_{p}	3.77
$\text{Transf}_{ra,np}$	40.87
$\text{Transf}_{ra,np}$	18.62
Transf_{nw}	43.82
Targ_{np}	9.06
Targ_{p}	10.41

Subsequently, figure 5.2 presents the periodic skill scores for 3 NL sites, CR, and UK. NL_6 is included to analyze its lower skill score compared to other NL sites. Appendix A presents the same figures for the remaining sites. First, these figures

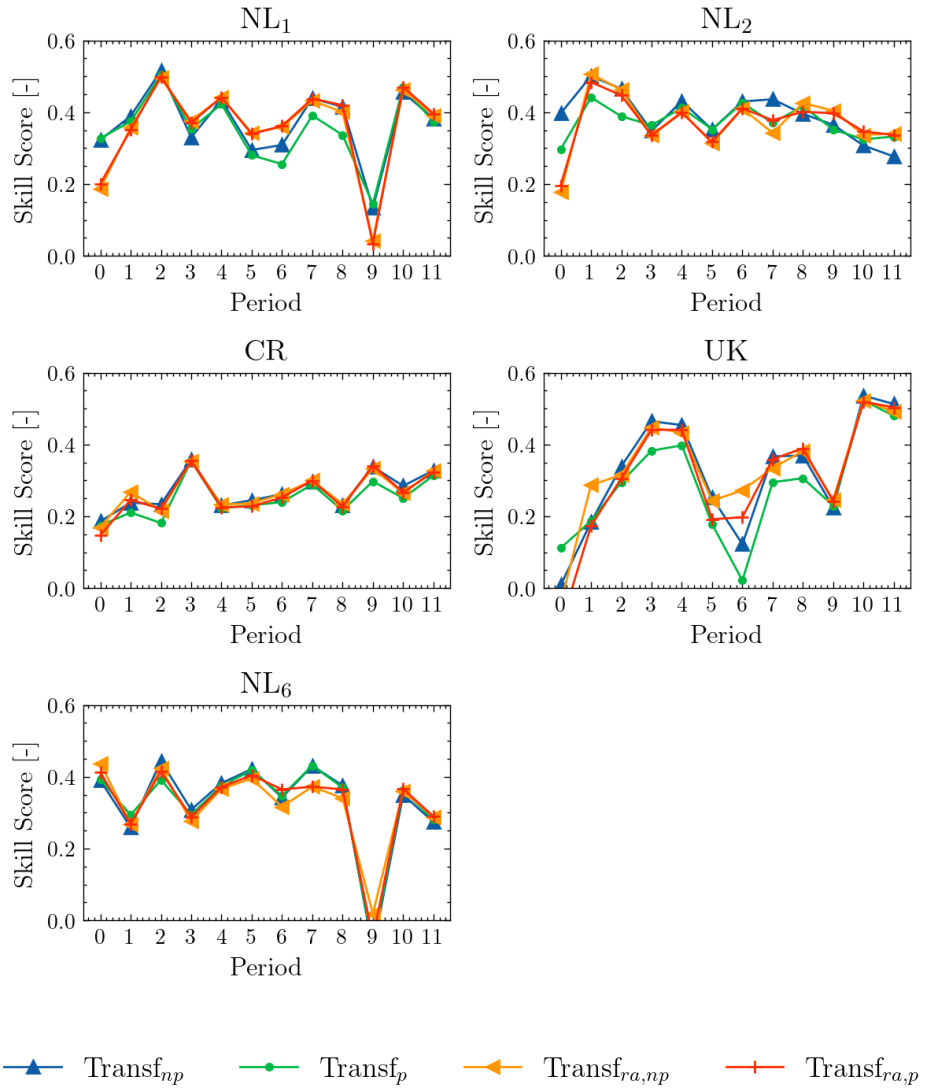


Figure 5.2: Periodic skill scores of proposed models in the target domain

show that the periodic skill scores are rather unstable. It is difficult to assess if the changes in skill score can be attributed to overfitting, the inherent variability of irradiance and PV power, or the accuracy of the NWP forecasts. This will be further explored in the succeeding section. Second, reanalysis-trained models catch up with the historical NWP models after the first-fine-tuning opportunity. Third, the UK, where the zero-shot forecast was inadequate, has the most unstable skill scores. Finally, a significant performance drop is noticed in period 9 for NL_1, NL_6 and UK. The following section will try to find the reason for this significant drop in periodic skill scores.

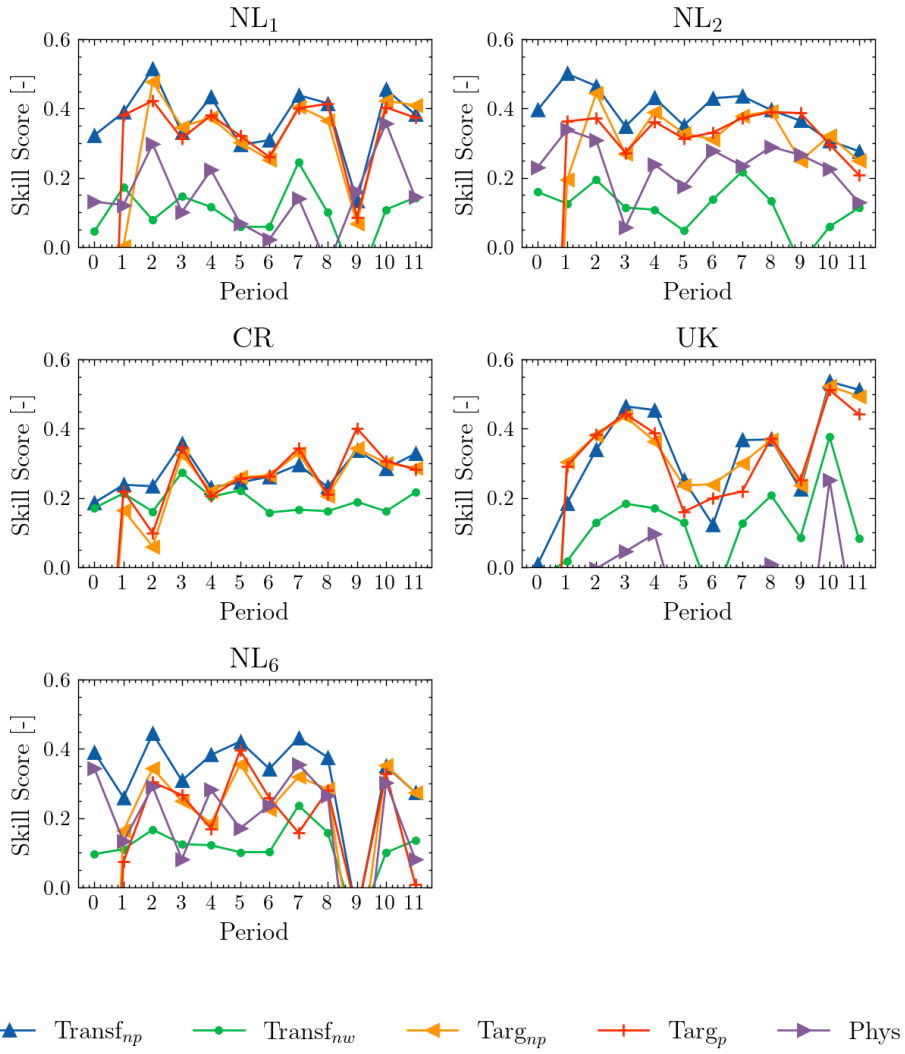


FIGURE 5.3: Periodic skill scores of model Transf$_{np}$ and the benchmarks in the target domain

Figure 5.3 shows a similar figure but only with Transf$_{np}$ of the proposed models and the benchmark models. One can observe that unstable behavior is noticed for the benchmark models, even the physical model. An influence that can not be attributed to the periodical fine-tuning seems to exist. This influence could be the variability of PV power or the accuracy of NWP forecasts. Next, the figure shows that the target models show similar performance after one or two months compared to the proposed TL model. The physics-informed model shows higher skill scores for the first month. Furthermore, the benchmark models also show the performance drop for sites NL$_1$ and NL$_6$.

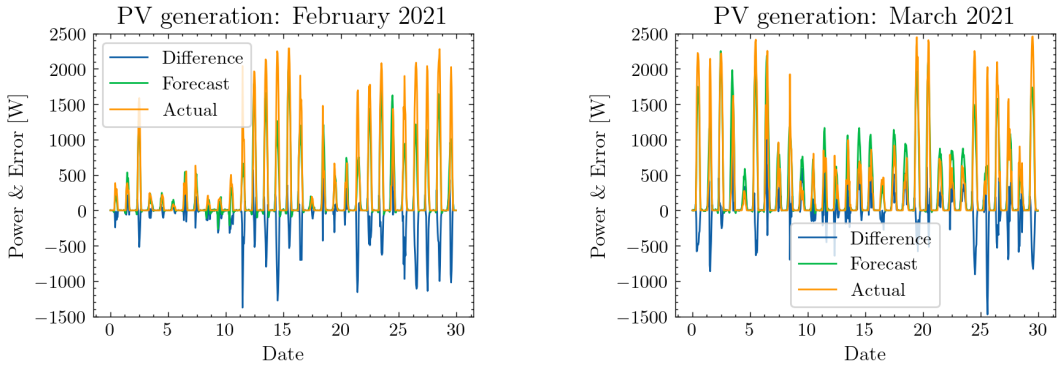
5.3 Result Analysis

This section will address some results that were obtained and provide further explanation. First, the performance drop in the ninth period will be discussed. Second, the periodic results will be compared to a source model with no fine-tuning to isolate its effect. Third, to asses whether NWP forecast errors influence the performance, the overall performance in the target domain will be compared to the NWP forecast errors. Finally, other popular ML models for solar power forecasting will be evaluated on their zero-shot performance and compared to the proposed model.

5.3.1 Analysis of performance drop

The previous section mentioned that the skill score dropped significantly in period 9 for some sites in the Netherlands and the UK. For NL_1, the skill score of the ML models even dropped close to zero, and the physical model performed better. This section will analyze the cause of this significant performance drop to find a possible weakness in the model.

The ninth period covers 28 January until 27 February 2021. Figure 5.4 shows the model's forecast and the actual PV generation for February and March. The model significantly underestimates the PV power in February compared to March, leading to a lower skill score. The focus will be on site NL_1 to analyze this phenomenon.

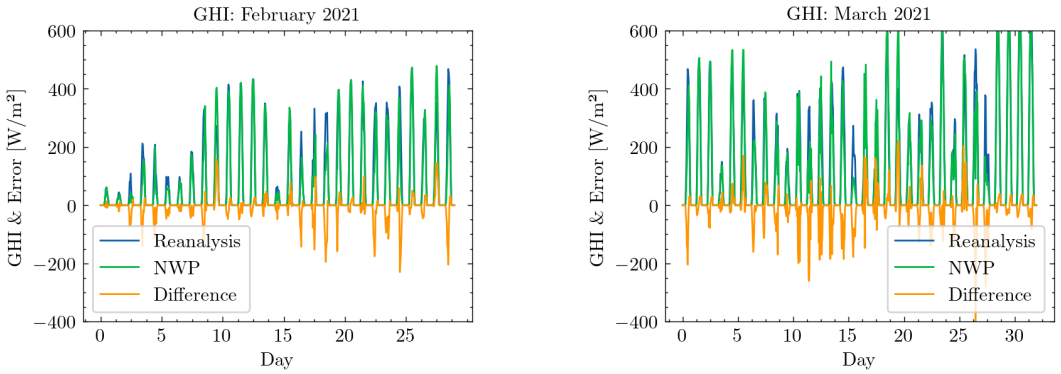


(A) February 2021 (B) March 2021

Figure 5.4: The forecasts, actual PV generation, and error for February and March 2021

The first hypothesis is that the irradiance forecasting error is higher than other months, resulting in higher errors in the PV forecast. However, this hypothesis cannot explain why the physical model suffers less, as it depends on the same irradiance forecasts. Figure 5.5 shows that the errors in the forecast of the GHI for February 2021 are not significantly different than for March 2021

A second hypothesis could be that February 2021 had weather conditions for which the model was unprepared, meaning it did not encounter this in the training process. During this period, the snowstorm Darcy passed through the Netherlands,



(A) February 2021 (B) March 2021

FIGURE 5.5: The forecasted irradiance, actual irradiance, and difference for February and March 2021

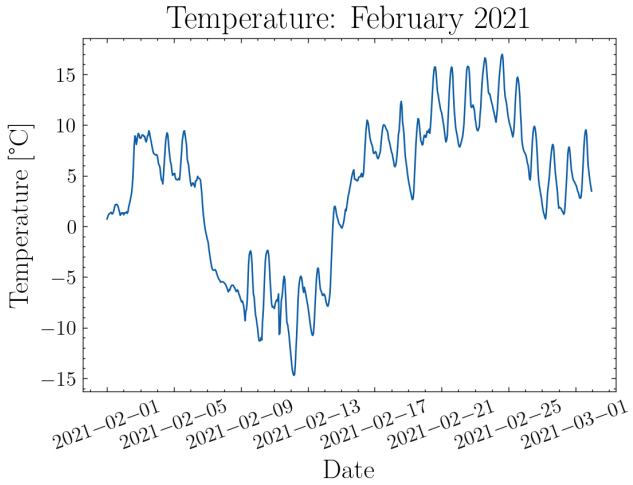


FIGURE 5.6: The temperature for February 2021

UK, and other West-European countries [133]. This resulted in heavy snowfall and reported temperatures of −23 °C in Scotland [134]. The Dutch Meteorological Institute mentioned that a snowstorm of this severity was last observed in 2010, whereas the model's training extended until 2016. Figure 5.6 shows the temperature at site NL_1 for February 2021. It shows that the temperature reached a minimum of −15 °C, while the minimum observed temperature of the training data until February was −7 °C. Furthermore, this means that the input was scaled to a negative value instead of between 0 and 1, which is the aim of min-max scaling. Table 5.8 shows the minimum input temperature after min-max scaling. Sites NL_1 and NL_6 have the most negative inputs and showed the most significant performance drop. It seems plausible that these extreme weather conditions resulted in an inaccurate model.

Table 5.8: Minimum scaled input temperature in February 2021

Site	T_{mm} [-]
NL_1	-0.0876
NL_2	-0.0569
NL_3	-0.0191
NL_4	-0.0610
NL_5	-0.0005
NL_6	-0.0841
NL_7	-0.0202

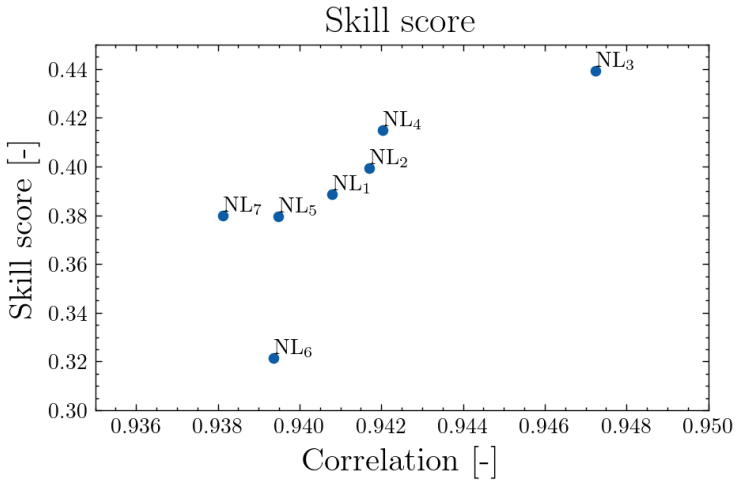


Figure 5.7: Scatter plot of the irradiance correlation and the skill score of case Transf_{np}

5.3.2 Influence of NWP errors

This section analyzes the differences in skill scores between different NL sites. The UK and CR sites were discarded due to the low similarity between the source and target domains. It seems plausible that the NWP error influences the skill scores of the models. The correlation between the GHI provided by NWP and reanalysis data represents the NWP error. Figure 5.7 shows a dependence except for site NL_6. Sites NL_1 and NL_6 suffered from the performance drop in period 9. The hyperparameter tuning pipeline used generation data from the site NL_1. If the performance drop were not present, NL_1 would probably be an outlier having a higher skill score compared to the correlation of irradiance. Nonetheless, discarding the outliers, it can be observed that a relation may exist between the accuracy of the NWP forecast and the model's performance.

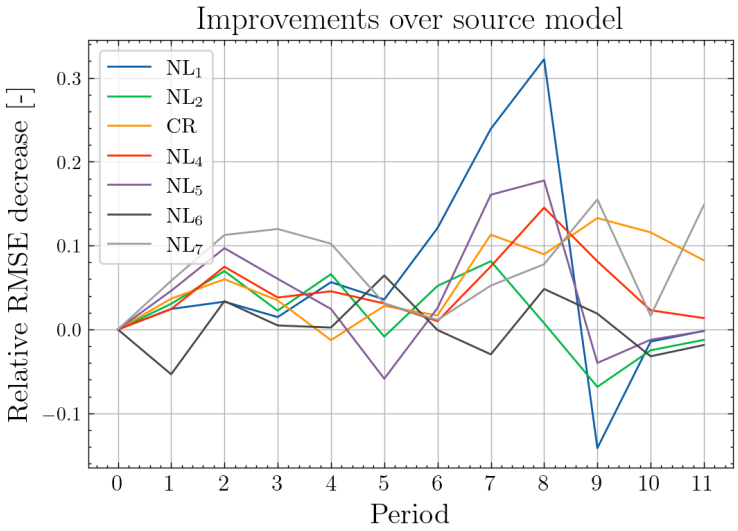


FIGURE 5.8: The relative increase in performance compared to the non-fine-tuned source model

5.3.3 Performance of fine-tuning

Assessing whether periodical fine-tuning results in a better or overfitted model is complex. The increase in performance compared to a non-fine-tuned source model could indicate the effect of fine-tuning. Figure 5.8 shows this relative increase for the Transf_{np} model for all sites. It indicates unstable and even negative improvements compared to the source model. Site NL_1 shows the most stable increase until period 9, which is expected due to the hyperparameter tuning based on this site. The transfer model seems to suffer more from extreme weather conditions, although this effect is less pronounced for site NL_6. In conclusion, the improvements over the non-fine-tuned model do not show a stable converging behavior, although the overall effect is positive.

5.3.4 Other ML models

Two other popular ML methods, SVM and RF, are evaluated on their zero-shot accuracy after training on the source data for the first 4 cases and all sites. The skill scores for the RF method are given in table 5.9. Similarly, table 5.10 gives the same results for the SVM.

These tables show that the physics-informed methods, except for the UK and CR sites, perform significantly better in terms of accuracy for both methods. Furthermore, the physics-informed reanalysis-trained model achieves zero-shot performance, close to the non-physics-informed NWP-trained model. These observations show the added value of physics-informed variables for these models, which was not present for the LSTM model.

To assess the proposed model with respect to these other two ML models, the RMSE for the four cases and all sites are presented in figure 5.9. The UK is excluded

Table 5.9: Skill score of zero-shot forecasts for RF

Site	Transf_{np}	Transf_{p}	$\text{Transf}_{ra,np}$	$\text{Transf}_{ra,p}$
NL_1	0.269753	**0.347558**	0.040807	0.22386
NL_2	0.268531	**0.380853**	0.145425	0.288148
NL_3	0.172657	**0.379624**	0.056903	0.292801
NL_4	0.252852	**0.284106**	0.057661	0.184284
NL_5	0.265528	**0.375873**	0.07585	0.278884
NL_6	0.319432	**0.37249**	0.229289	0.334611
NL_7	0.050721	**0.197301**	-0.18897	0.119028
CR	**0.145367**	0.145217	0.053333	0.075221
UK	**0.064653**	0.064294	-0.00392	0.046176

Table 5.10: Skill score of zero-shot forecasts for SVM

Site	Transf_{np}	Transf_{p}	$\text{Transf}_{ra,np}$	$\text{Transf}_{ra,p}$
NL_1	0.1871	**0.3001**	0.0441	0.1811
NL_2	0.2056	**0.3734**	0.1387	0.2396
NL_3	0.0935	**0.3681**	0.0555	0.2698
NL_4	0.2236	**0.2641**	0.0773	0.2053
NL_5	0.1968	**0.3690**	0.1192	0.2853
NL_6	0.1937	**0.3784**	0.1719	0.3646
NL_7	0.0686	**0.2564**	0.0080	0.1633
CR	0.1707	**0.1724**	0.0394	0.0859
UK	**0.1551**	0.0110	0.0035	-0.013

due to all zero-shot skill scores being around zero. It shows that the best-performing model depends on the place and the case. Except for NL_1, the proposed model has the highest accuracy over all cases. Nevertheless, the differences between the models, especially RF, are minor, while no hyperparameter tuning was performed for these models. Regardless, the proposed model allows for effortless fine-tuning in the target domain, which is more complicated to achieve with the other ML models. As such, it can be concluded that the proposed model is the best solution for this forecasting problem overall.

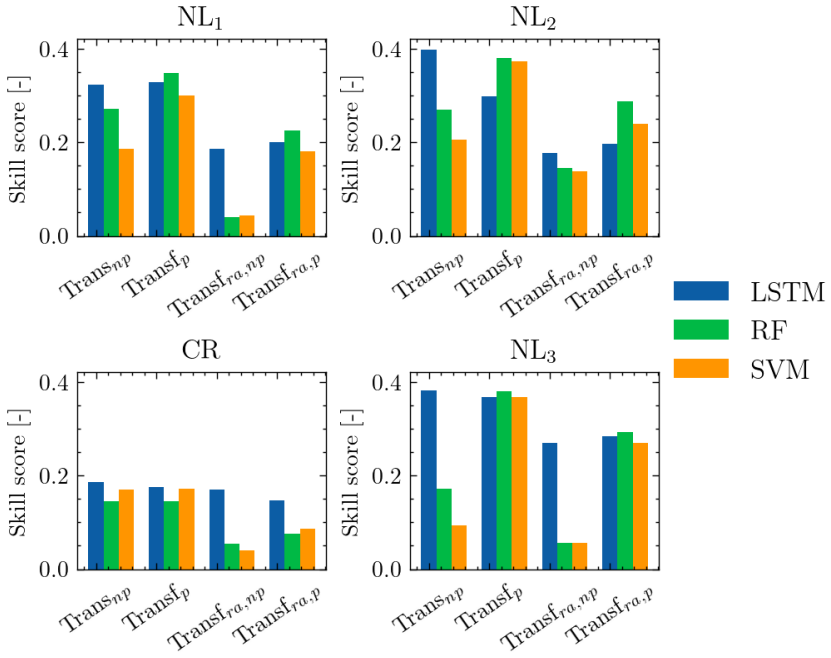


FIGURE 5.9: Bar chart for comparison of ML models

Chapter 6

Discussion

This chapter interprets the results and links them with the research questions. The first research question was whether a TL model with weather covariates can provide better forecasts than physical, target, and TL models without weather covariates. Second, it was examined whether adding physics-based features could enhance the TL model's performance and training time. Third, due to the low accessibility of historical NWP runs, I examined whether a source model trained on reanalysis data performs well. Furthermore, this section will summarise new insights into the results and the possibilities for future research.

The results showed that the proposed model's forecasts exceed the benchmarks in skill scores. These findings confirm that a TL model, where the source domain consists of synthetic PV generation, can most certainly provide reliable forecasts. The model shows an increased performance even for sites where a lower resemblance between source and target PV generation was observed. It can be concluded that the proposed model is a viable solution to forecast the PV power of residential installations with limited data availability. However, models trained only on actual PV generation seem to catch up after two months. This limits the applicability of transfer learning once two months of actual PV generation is available.

The physics-informed variables and post-processing were expected to increase performance and decrease the computational resources of the model. However, the results refute this hypothesis, as the site performance was sometimes worse. It seems that the LSTM can extract the engineered variables itself. This decrease in performance, although small, could be explained by the law of parsimony in ML, which states that the smallest possible set of features should be utilized. Nonetheless, for the models trained in the target domain, the physics-informed model seems to learn quicker and is consequently more accurate after only training on one month of data. Further research is necessary to confirm or deny this claim as this was not investigated for this thesis.

Concerning reanalysis data, the zero-shot forecasts are mostly more inaccurate. This is plausible due to the weather covariates from another provider and the presence of forecast errors, which it could not train on in the source domain. However, regarding performance over the whole year, these models seem to have slightly lower

skill scores. Considering the significantly higher accessibility of reanalysis data, it can be concluded that the reanalysis datasets are a viable alternative for TL if limited target data is available. Nonetheless, a trend of making historical NWP runs more accessible can be observed with initiatives mentioned in chapter 4. In the coming years, historical NWP output may become better accessible. However, when the target domain NWP differs from the source NWP provider, training on reanalysis data to have a less biased source model can still be a viable option.

The results presented skill scores grouped for each period between fine-tuning. The expectation was a stable converging curve, which showed that the model adapted to the target domain. However, the actual periodical skill scores behaved rather unstable. Some sites even showed a significant drop in performance in period 9. A storm that resulted in unseen negative temperatures could explain this drop. However, the ML models being a black box makes it difficult to confirm this hypothesis. Further research could concentrate on developing a more stable model that handles unseen data better. Weather measurements as an additional input feature could make the model more robust. A possibility is to construct an encoder-decoder architecture, where the encoder is fed with historical time series containing past weather measurements, and the decoder uses the encoder input and weather forecasts to construct a forecast for each timestamp. However, I have no proof for this hypothesis, and additional research is necessary.

It was difficult to assess the instability of periodic skill scores due to the seasonality of PV production. To better evaluate the effect of the periodic fine-tuning, the performance increase of the fine-tuned models was assessed with a non-fine-tuned source model. The expected stable increase in the model's performance was not observed; some sites even showed a negative skill score for some periods. Nonetheless, the overall effect of the fine-tuning seemed to be positive. Improvements are most certainly possible. First, a stacked LSTM model where some layers are frozen in the target domain could be a more stable approach. Second, a learning rate dependent on the amount of target data available could enhance performance. Third, the hyperparameter tuning process was split up for the source and target domain. An incorporated approach that uses the target domain accuracy as a metric could lead to more stable and performant TL models. Finally, the fine-tuning process could depend on previous performance by slowing the fine-tuning if the model is relatively accurate.

After training on source domain data, other ML models were assessed on their zero-shot performance in the target domain. The results show high skill scores for the physics-informed cases. It seems that the physics-informed variables can enhance performance for some ML models. Furthermore, their skill scores were comparable to those of the proposed model without tuning the hyperparameters. Therefore, future research could adapt the proposed TL approach for these ML methods, possibly resulting in more performant models.

An economic assessment of this increase in forecasting accuracy is complex due to its dependence on household appliances and the presence of a battery system. However, future research should attempt to asses the economic importance to decide if future research is of interest for residential PV installations. Nonetheless, except

for the spatial dimensions and possible tracking, utility-scale PV systems are similar to household PV systems, and the results obtained should be generalizable to utility-scale PV systems.

Chapter 7

Conclusion

This thesis presented a transfer learning model for forecasting the power output of residential photovoltaic installations, demonstrating significant improvements over conventional benchmark models. The proposed TL model uses Long Short-Term Memory networks and synthetic PV generation to enhance forecasting accuracy even with no actual PV generation to train on.

The integration of Numerical Weather Predictions of weather variables has led to a significant increase in the performance of the ML model in the target domain, compared to a model that does not utilize weather covariates. Furthermore, the proposed model outperforms the physical model or models trained on target data. The comparison was done using skill scores based on the convex combination of persistence and climatology, following the recommendations of the solar forecasting community.

Contrary to expectations, including physics-informed variables did not enhance the TL model's zero-shot forecast accuracy or overall accuracy, resulting even in a slight decrease. Furthermore, the physics-informed features did not result in lower training times. Models trained on reanalysis data achieved acceptable zero-shot forecasts and caught up with those trained on historical Numerical Weather Prediction data once limited target data became available. This indicates that reanalysis-trained models are a viable option when historical NWP outputs are scarce.

The analysis of results also reveals challenges, particularly the impact of unseen weather conditions. Storm Darcy in February 2021, which resulted in unseen negative temperatures, deteriorated model performance by introducing negative temperatures unseen by the model during training. This instability underscores the need to incorporate more diverse weather scenarios or develop more stable models in future research to ensure stability under such conditions.

Furthermore, while the proposed models show strong performance, there is potential for better fine-tuning within the target domain. More adaptive fine-tuning techniques or other transfer learning strategies could further enhance model accuracy, especially as more target data becomes available over time.

In addition to the proposed models, other machine learning models were trained on synthetic data and assessed on their zero-shot performance. These promising

results advocate that implementing a similar methodology for these ML models, especially the random forest, could result in a better-performing model.

In conclusion, the developed TL models are valuable for forecasting the PV output of new residential installations with limited data availability. These models can rapidly adapt and deliver reliable forecasts with minimal initial data, supporting more efficient and effective integration of solar power into the energy grid. Future research should focus on refining these models, incorporating additional covariates, and improving fine-tuning techniques while exploring other ML models with TL to continue enhancing forecast accuracy. Finally, these results should be placed in a broader context to examine the necessity of future research. The economic implications of increasing forecasting accuracy on the scale of individual households can justify research focusing on residential installations.

Appendices

Appendix A

Additional metrics

This appendix includes additional metrics. These metrics lead to similar conclusions as the ones based on the skill score and are therefore left out of the result section. However, to avoid leaving out valuable information, the RMSE, MAE, and correlation are included in this section.

A.1 RMSE

Table A.1 shows the RMSE of the zero-shot forecasts. Next, table A.2 shows the RMSE over all other periods. Figures A.1 and A.2 show, respectively, the skill scores of the proposed models and for the proposed model Transf_{np} against the benchmark models for the other sites.

TABLE A.1: RMSE of zero-shot forecasts in the target domain

Site	Transf_{np}	Transf_{p}	$\text{Transf}_{ra,np}$	$\text{Transf}_{ra,np}$	Transf_{nw}	Phys	Persis
NL_1	306.41	**304.34**	367.87	361.58	432.410	393.34	452.47
NL_2	**249.14**	290.50	339.98	331.94	347.54	317.69	413.03
NL_3	**266.19**	272.29	314.18	308.27	383.77	359.22	430.01
NL_4	266.42	**231.25**	287.48	289.21	345.20	299.30	367.85
NL_5	**247.19**	249.51	288.06	288.46	360.56	347.93	396.69
NL_6	261.48	260.00	**241.09**	251.87	387.60	281.32	428.59
NL_7	**281.80**	287.90	300.34	298.47	347.73	423.46	382.05
CR	**774.10**	785.25	789.17	811.04	788.78	1762.11	950.90
UK	327.67	**293.64**	344.37	364.96	339.65	407.28	330.46

A.2 Correlation

Table A.3 shows the correlation between forecasts and actual PV power after the first finetuning opportunity.

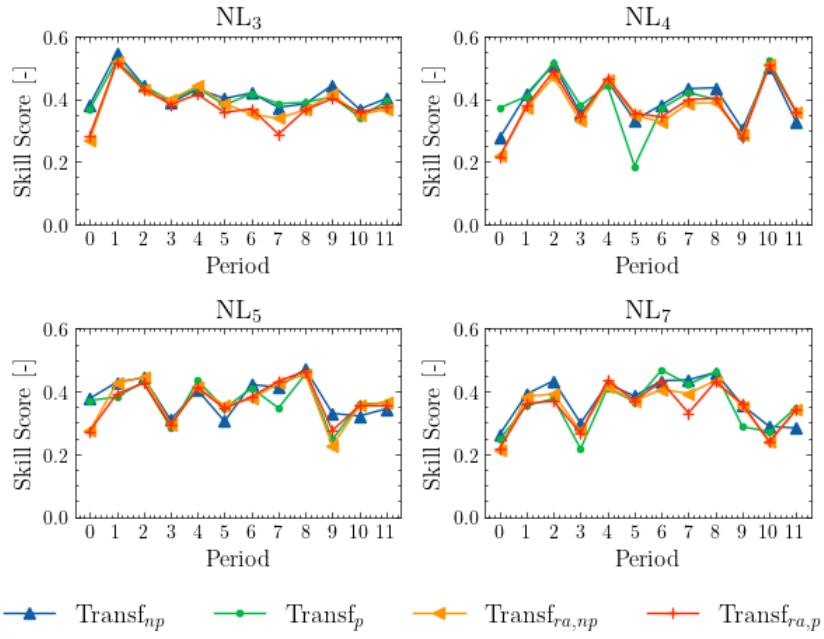


Figure A.1: Periodical skill scores for additional sites

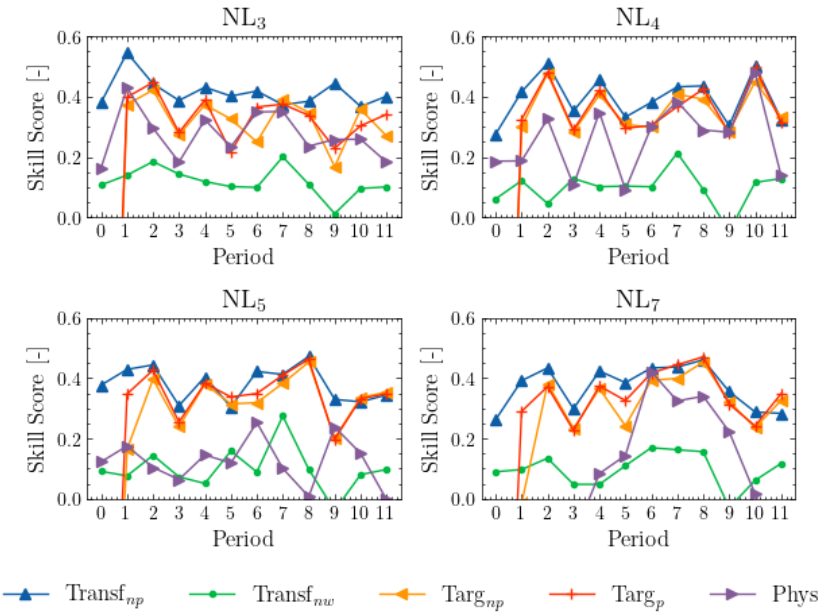


Figure A.2: Periodical skill scores of Transf$_{np}$ and benchmark modelsfor additional sites

TABLE A.2: Average RMSE of the Walk-Forward validation in the target domain, excluding the first month

Site	Transf_{np}	Transf_{p}	$\text{Transf}_{ra,np}$	$\text{Transf}_{ra,np}$	Transf_{nw}	Targ_{np}	Targ_{p}	Phys
NL_1	**314.13**	317.48	315.49	315.39	461.52	355.86	331.73	418.84
NL_2	257.18	263.42	**255.35**	257.37	380.08	291.61	285.54	325.44
NL_3	**241.74**	245.46	248.42	250.33	375.11	281.65	275.32	304.89
NL_4	**259.78**	264.01	266.44	263.42	400.72	280.23	276.61	321.97
NL_5	**253.39**	255.38	254.16	255.51	373.00	278.96	266.28	352.97
NL_6	**226.75**	227.58	227.43	227.72	293.88	254.50	271.94	280.61
NL_7	**241.91**	249.75	247.76	249.23	353.29	280.05	257.23	365.69
CR	**776.90**	799.97	779.64	783.06	871.16	801.84	784.77	1279.13
UK	222.29	235.97	**217.54**	227.42	298.98	217.25	221.79	359.18

TABLE A.3: Correlation of the walk-forward validation in the target domain, excluding the first month

Site	Transf_{np}	Transf_{p}	$\text{Transf}_{ra,np}$	$\text{Transf}_{ra,np}$	Transf_{nw}	Targ_{np}	Targ_{p}	Phys
NL_1	**0.884975**	0.883211	0.883402	0.883438	0.732417	0.850008	0.870895	0.853677
NL_2	0.885107	0.878983	**0.88693**	0.884656	0.729112	0.850834	0.856984	0.851645
NL_3	**0.906034**	0.902965	0.900928	0.899441	0.759447	0.87175	0.877005	0.885259
NL_4	**0.88675**	0.882793	0.880512	0.883536	0.705583	0.867854	0.872061	0.854591
NL_5	**0.880091**	0.877934	0.879279	0.877953	0.715377	0.853041	0.867208	0.839388
NL_6	**0.870582**	0.868618	0.870137	0.869234	0.771569	0.841776	0.817659	0.833095
NL_7	**0.885763**	0.876965	0.879067	0.877513	0.735125	0.8448	0.870286	0.828017
CR	**0.791565**	0.777263	0.790988	0.788986	0.73035	0.777461	0.789991	0.773157
UK	0.862489	0.848213	0.866918	0.855119	0.726864	**0.867477**	0.860883	0.771492

A.3 MAE

Table A.3 shows the MAE between forecasts and actual PV power after the first finetuning opportunity.

Table A.4: MAE of the walk-forward validation in the target domain, excluding the first month

Site	Transf_{np}	Transf_{p}	$\text{Transf}_{ra,np}$	$\text{Transf}_{ra,np}$	Transf_{nw}	Targ_{np}	Targ_{p}	Phys
NL_1	218.59	220.24	**218.02**	218.13	346.86	251.02	240.04	278.87
NL_2	**180.22**	184.54	182.79	181.91	289.31	207.68	204.39	215.00
NL_3	**163.72**	166.38	170.47	172.39	269.63	198.56	190.71	199.79
NL_4	**184.53**	187.72	187.79	185.87	298.68	205.60	201.55	211.38
NL_5	**178.35**	179.28	178.83	179.17	280.37	198.93	189.30	235.58
NL_6	**149.56**	151.21	153.02	152.60	209.88	177.78	183.99	175.39
NL_7	**169.65**	177.75	175.49	178.94	265.04	194.32	183.92	250.30
CR	**581.43**	604.11	589.07	590.37	658.30	584.29	581.69	962.77
UK	158.92	172.74	153.46	159.95	223.78	**157.56**	160.40	247.03

Bibliography

[1] J. S. Kikstra, Z. R. J. Nicholls, C. J. Smith, J. Lewis, R. D. Lamboll, E. Byers, M. Sandstad, M. Meinshausen, M. J. Gidden, J. Rogelj, E. Kriegler, G. P. Peters, J. S. Fuglestvedt, R. B. Skeie, B. H. Samset, L. Wienpahl, D. P. van Vuuren, K.-I. van der Wijst, A. Al Khourdajie, P. M. Forster, A. Reisinger, R. Schaeffer, and K. Riahi, "The IPCC Sixth Assessment Report WGIII climate assessment of mitigation pathways: From emissions to global temperatures," *Geoscientific Model Development*, vol. 15, no. 24, pp. 9075–9109, Dec. 2022.

[2] "CO2 Emissions in 2022 – Analysis," https://www.iea.org/reports/co2-emissions-in-2022, Mar. 2023.

[3] United Nations Environment Programme, *Emissions Gap Report 2023: Broken Record – Temperatures Hit New Highs, yet World Fails to Cut Emissions (Again)*. United Nations Environment Programme, Nov. 2023.

[4] "Electricity Market Report 2023 – Analysis," https://www.iea.org/reports/electricity-market-report-2023, Feb. 2023.

[5] "2023 Levelized Cost Of Energy+," https://www.lazard.com/research-insights/2023-levelized-cost-of-energyplus/.

[6] IEA, "Solar," https://www.iea.org/energy-system/renewables/solar-pv.

[7] Winckelmans, "Fluvius grijpt in om uitval zonnepanelen te voorkomen | De Standaard," *De Standaard*, Apr. 2024.

[8] ELIA, "CCMD," https://www.eliagroup.eu/en/ccmd.

[9] "Capacity tariffs explained – gridX," https://www.gridx.ai/knowledge/capacity-tariffs, Apr. 24.

[10] V. Z. Gjorgievski, S. Cundeva, and G. E. Georghiou, "Social arrangements, technical designs and impacts of energy communities: A review," *Renewable Energy*, vol. 169, pp. 1138–1156, May 2021.

[11] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy Forecasting: A Review and Outlook," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 376–388, 2020.

[12] M. P. Almeida, O. Perpiñán, and L. Narvarte, "PV power forecast using a nonparametric PV model," *Solar Energy*, vol. 115, pp. 354–368, May 2015.

[13] W. Khan, S. Walker, and W. Zeiler, "Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach," *Energy*, vol. 240, p. 122812, Feb. 2022.

[14] J. G. da Silva Fonseca Jr., T. Oozeki, T. Takashima, G. Koshimizu, Y. Uchida, and K. Ogimoto, "Use of support vector regression and numerically predicted cloudiness to forecast power output of a photovoltaic power plant in Kitakyushu, Japan," *Progress in Photovoltaics: Research and Applications*, vol. 20, no. 7, pp. 874–882, 2012.

[15] S. E. Razavi, A. Arefi, G. Ledwich, G. Nourbakhsh, D. B. Smith, and M. Minakshi, "From Load to Net Energy Forecasting: Short-Term Residential Forecasting for the Blend of Load and PV Behind the Meter," *IEEE Access*, vol. 8, pp. 224 343–224 353, 2020.

[16] K. Bot, A. Ruano, and M. d. G. Ruano, "Short-Term Forecasting Photovoltaic Solar Power for Home Energy Management Systems," *Inventions*, vol. 6, no. 1, p. 12, Mar. 2021.

[17] R. P. Masini, M. C. Medeiros, and E. F. Mendes, "Machine learning advances for time series forecasting," *Journal of Economic Surveys*, vol. 37, no. 1, pp. 76–111, 2023.

[18] K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, "Accurate medium-range global weather forecasting with 3D neural networks," *Nature*, vol. 619, no. 7970, pp. 533–538, Jul. 2023.

[19] W. Guo, L. Che, M. Shahidehpour, and X. Wan, "Machine-Learning based methods in short-term load forecasting," *The Electricity Journal*, vol. 34, no. 1, p. 106884, Jan. 2021.

[20] A. Alcañiz, D. Grzebyk, H. Ziar, and O. Isabella, "Trends and gaps in photovoltaic power forecasting with machine learning," *Energy Reports*, vol. 9, pp. 447–471, Dec. 2023.

[21] K. Zaouali, R. Rekik, and R. Bouallegue, "Deep Learning Forecasting Based on Auto-LSTM Model for Home Solar Power Systems," in *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, Jun. 2018, pp. 235–242.

[22] A. Sharma and A. Kakkar, "Forecasting daily global solar irradiance generation using machine learning," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 2254–2269, Feb. 2018.

[23] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF Models for Sequence Tagging," Aug. 2015.

[24] Y. Zhang, Q. Liu, and L. Song, "Sentence-State LSTM for Text Representation," May 2018.

[25] L. Wang, M. Mao, J. Xie, Z. Liao, H. Zhang, and H. Li, "Accurate solar PV power prediction interval method based on frequency-domain decomposition and LSTM model," *Energy*, vol. 262, p. 125592, Jan. 2023.

[26] J. Zhang, Y. Chi, and L. Xiao, "Solar Power Generation Forecast Based on LSTM," in *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Nov. 2018, pp. 869–872.

[27] L. Visser, T. AlSkaif, and W. van Sark, "Operational day-ahead solar power forecasting for aggregated PV systems with a varying spatial distribution," *Renewable Energy*, vol. 183, pp. 267–282, Jan. 2022.

[28] L. Visser, T. AlSkaif, J. Hu, A. Louwen, and W. Van Sark, "On the value of expert knowledge in estimation and forecasting of solar photovoltaic power generation," *Solar Energy*, vol. 251, pp. 86–105, Feb. 2023.

[29] D. V. Pombo, M. J. Rincón, P. Bacher, H. W. Bindner, S. V. Spataru, and P. E. Sørensen, "Assessing stacked physics-informed machine learning models for co-located wind–solar power forecasting," *Sustainable Energy, Grids and Networks*, vol. 32, p. 100943, Dec. 2022.

[30] E. Commission, "Photovoltaic Geographical Information System (PVGIS)," https://joint-research-centre.ec.europa.eu/photovoltaic-geographical-information-system-pvgis_en, Feb. 2024.

[31] R. M. Goody and Y. L. Yung, *Atmospheric Radiation: Theoretical Basis*. Oxford University Press, Dec. 1995.

[32] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, "Solar forecasting methods for renewable energy integration," *Progress in Energy and Combustion Science*, vol. 39, no. 6, pp. 535–576, Dec. 2013.

[33] P. Lauret, R. Alonso-Suárez, J. Le Gal La Salle, and M. David, "Solar Forecasts Based on the Clear Sky Index or the Clearness Index: Which Is Better?" *Solar*, vol. 2, no. 4, pp. 432–444, Oct. 2022.

[34] D. Yang, W. Wang, C. A. Gueymard, T. Hong, J. Kleissl, J. Huang, M. J. Perez, R. Perez, J. M. Bright, X. Xia, D. Van Der Meer, and I. M. Peters, "A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality," *Renewable and Sustainable Energy Reviews*, vol. 161, p. 112348, Jun. 2022.

[35] Y. Chu, Y. Wang, D. Yang, S. Chen, and M. Li, "A review of distributed solar forecasting with remote sensing and deep learning," *Renewable and Sustainable Energy Reviews*, vol. 198, p. 114391, Jul. 2024.

[36] A. G. Amillo, T. Huld, and R. Müller, "A New Database of Global and Direct Solar Radiation Using the Eastern Meteosat Satellite, Models and Validation," *Remote Sensing*, vol. 6, no. 9, pp. 8165–8189, Sep. 2014.

[37] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569–582, May 2017.

[38] G. Huang, Z. Li, X. Li, S. Liang, K. Yang, D. Wang, and Y. Zhang, "Estimating surface solar irradiance from satellites: Past, present, and future perspectives," *Remote Sensing of Environment*, vol. 233, p. 111371, Nov. 2019.

[39] B. V, "Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik," *Meteor. Z.*, vol. 21, pp. 1–7, 1904.

[40] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 1, pp. 2–10, Mar. 2009.

[41] "Irradiance Data Methodology," http://solcast.com/irradiance-data-methodology.

[42] P. A. Jimenez, J. P. Hacker, J. Dudhia, S. E. Haupt, J. A. Ruiz-Arias, C. A. Gueymard, G. Thompson, T. Eidhammer, and A. Deng, "WRF-Solar: Description and Clear-Sky Assessment of an Augmented NWP Model for Solar Power Prediction," *Bulletin of the American Meteorological Society*, vol. 97, no. 7, pp. 1249–1264, Jul. 2016.

[43] M. Bai, P. Yao, H. Dong, Z. Fang, W. Jin, Xusheng Yang, J. Liu, and D. Yu, "Spatial-temporal characteristics analysis of solar irradiance forecast errors in Europe and North America," *Energy*, vol. 297, p. 131187, Jun. 2024.

[44] F. J. L. Lima, F. R. Martins, E. B. Pereira, E. Lorenz, and D. Heinemann, "Forecast for surface solar irradiance at the Brazilian Northeastern region using NWP model and artificial neural networks," *Renewable Energy*, vol. 87, pp. 807–818, Mar. 2016.

[45] K. Bakker, K. Whan, W. Knap, and M. Schmeits, "Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation," *Solar Energy*, vol. 191, pp. 138–150, Oct. 2019.

[46] N. O. a. A. A. US Department of Commerce, "Breakthrough article on the First Climate Model," https://celebrating200years.noaa.gov/breakthroughs/climate_model/.

[47] L. Fara, A. Diaconu, D. Craciunescu, and S. Fara, "Forecasting of Energy Production for Photovoltaic Systems Based on ARIMA and ANN Advanced Models," *International Journal of Photoenergy*, vol. 2021, p. e6777488, Aug. 2021.

[48] S. I. Vagropoulos, G. I. Chouliaras, E. G. Kardakos, C. K. Simoglou, and A. G. Bakirtzis, "Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting," in *2016 IEEE International Energy Conference (ENERGYCON)*, Apr. 2016, pp. 1–6.

[49] D. Lee and K. Kim, "Deep Learning Based Prediction Method of Long-term Photovoltaic Power Generation Using Meteorological and Seasonal Information," *Journal of Society for e-Business Studies*, vol. 24, no. 1, Mar. 2019.

[50] E. Kim, M. S. Akhtar, and O.-B. Yang, "Designing solar power generation output forecasting methods using time series algorithms," *Electric Power Systems Research*, vol. 216, p. 109073, Mar. 2023.

[51] S. Atique, S. Noureen, V. Roy, S. Bayne, and J. Macfie, "Time series forecasting of total daily solar energy generation: A comparative analysis between ARIMA and machine learning techniques," in *2020 IEEE Green Technologies Conference(GreenTech)*, Apr. 2020, pp. 175–180.

[52] B. Li, X. Chen, and A. Jain, "Enhancing Power Prediction of Photovoltaic Systems: Leveraging Dynamic Physical Model for Irradiance-to-Power Conversion," Feb. 2024.

[53] M. J. Mayer and D. Yang, "Pairing ensemble numerical weather prediction with ensemble physical model chain for probabilistic photovoltaic power forecasting," *Renewable and Sustainable Energy Reviews*, vol. 175, p. 113171, Apr. 2023.

[54] D. Markovics and M. J. Mayer, "Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction," *Renewable and Sustainable Energy Reviews*, vol. 161, p. 112364, Jun. 2022.

[55] A. Gensler, J. Henze, B. Sick, and N. Raabe, "Deep Learning for solar power forecasting — An approach using AutoEncoder and LSTM Neural Networks," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. Budapest, Hungary: IEEE, Oct. 2016, pp. 002 858–002 865.

[56] M. Massaoudi, I. Chihi, H. Abu-Rub, S. S. Refaat, and F. S. Oueslati, "Convergence of Photovoltaic Power Forecasting and Deep Learning: State-of-Art Review," *IEEE Access*, vol. 9, pp. 136 593–136 615, 2021.

[57] W. Liu and Z. Mao, "Short-term photovoltaic power forecasting with feature extraction and attention mechanisms," *Renewable Energy*, vol. 226, p. 120437, May 2024.

[58] K. Wang, X. Qi, and H. Liu, "A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network," *Applied Energy*, vol. 251, p. 113315, Oct. 2019.

[59] R. A. Rajagukguk, R. A. A. Ramadhan, and H.-J. Lee, "A Review on Deep Learning Models for Forecasting Time Series Data of Solar Irradiance and Photovoltaic Power," *Energies*, vol. 13, no. 24, p. 6623, Jan. 2020.

[60] A. Agga, A. Abbou, M. Labbadi, Y. E. Houm, and I. H. Ou Ali, "CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production," *Electric Power Systems Research*, vol. 208, p. 107908, Jul. 2022.

[61] M. Zounemat-Kermani, O. Batelaan, M. Fadaee, and R. Hinkelmann, "Ensemble machine learning paradigms in hydrology: A review," *Journal of Hydrology*, vol. 598, p. 126266, Jul. 2021.

[62] M. AlKandari and I. Ahmad, "Solar power generation forecasting using ensemble approach based on deep learning and statistical methods," *Applied Computing and Informatics*, vol. ahead-of-print, no. ahead-of-print, Jan. 2020.

[63] D. Chakraborty, J. Mondal, H. B. Barua, and A. Bhattacharjee, "Computational solar energy – Ensemble learning methods for prediction of solar power generation based on meteorological parameters in Eastern India," *Renewable Energy Focus*, vol. 44, pp. 277–294, Mar. 2023.

[64] O. Boussif, G. Boukachab, D. Assouline, S. Massaroli, T. Yuan, L. Benabbou, and Y. Bengio, "Improving day-ahead Solar Irradiance Time Series Forecasting by Leveraging Spatio-Temporal Context," Oct. 2023.

[65] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, Nov. 2020.

[66] M. F. Tahir, M. Z. Yousaf, A. Tzes, M. S. El Moursi, and T. H. M. El-Fouly, "Enhanced solar photovoltaic power prediction using diverse machine learning algorithms with hyperparameter optimization," *Renewable and Sustainable Energy Reviews*, vol. 200, p. 114581, Aug. 2024.

[67] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.

[68] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, p. 9, May 2016.

[69] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[70] M. Weber, M. Auch, C. Doblander, P. Mandl, and H.-A. Jacobsen, "Transfer Learning With Time Series Data: A Systematic Mapping Study," *IEEE Access*, vol. 9, pp. 165 409–165 432, 2021.

[71] M. Abubakr, B. Akoush, A. Khalil, and M. A. Hassan, "Unleashing deep neural network full potential for solar radiation forecasting in a new geographic location with historical data scarcity: A transfer learning approach," *The European Physical Journal Plus*, vol. 137, no. 4, p. 474, Apr. 2022.

[72] H. Sheng, B. Ray, J. Shao, D. Lasantha, and N. Das, "Generalization of solar power yield modeling using knowledge transfer," *Expert Systems with Applications*, vol. 201, p. 116992, Sep. 2022.

[73] Y. Tang, K. Yang, S. Zhang, and Z. Zhang, "Photovoltaic power forecasting: A hybrid deep learning model incorporating transfer learning strategy," *Renewable and Sustainable Energy Reviews*, vol. 162, p. 112473, Jul. 2022.

[74] A. Genovese, V. Bernardoni, V. Piuri, F. Scotti, and F. Tessore, "Photovoltaic Energy Prediction for New-Generation Cells with Limited Data: A Transfer Learning Approach," in *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, May 2022, pp. 1–6.

[75] Z. Tang, Y. Tang, A. Qiao, J. Liu, and J. Gao, "Transfer Learning Based Photovoltaic Power Forecasting with XGBoost," in *2023 Panda Forum on Power and Energy (PandaFPE)*, Apr. 2023, pp. 1781–1785.

[76] E. Sarmas, N. Dimitropoulos, V. Marinakis, Z. Mylona, and H. Doukas, "Transfer learning strategies for solar power forecasting under data scarcity," *Scientific Reports*, vol. 12, no. 1, p. 14643, Aug. 2022.

[77] T. Zhang, K. Zheng, and Q. Chen, "Short-Term Photovoltaic Prediction Based on Transfer Learning and LSTM," in *2023 IEEE 7th Conference on Energy Internet and Energy System Integration (EI2)*, Dec. 2023, pp. 3438–3443.

[78] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, May 2021.

[79] K. Kashinath, M. Mustafa, A. Albert, J.-L. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, R. Wang, A. Chattopadhyay, A. Singh, A. Manepalli, D. Chirila, R. Yu, R. Walters, B. White, H. Xiao, H. A. Tchelepi, P. Marcus, A. Anandkumar, P. Hassanzadeh, and n. Prabhat, "Physics-informed machine learning: Case studies for weather and climate modelling," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2194, p. 20200093, Feb. 2021.

[80] G. S. Misyris, A. Venzke, and S. Chatzivasileiadis, "Physics-Informed Neural Networks for Power Systems," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*. Montreal, QC, Canada: IEEE, Aug. 2020, pp. 1–5.

[81] Q. Paletta, Y. Nie, Y.-M. Saint-Drenan, and B. Le Saux, "Improving cross-site generalisability of vision-based solar forecasting models with physics-informed transfer learning," *Energy Conversion and Management*, vol. 309, p. 118398, Jun. 2024.

[82] P. G. Loutzenhiser, H. Manz, C. Felsmann, P. A. Strachan, T. Frank, and G. M. Maxwell, "Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation," *Solar Energy*, vol. 81, no. 2, pp. 254–267, Feb. 2007.

[83] R. Perez, R. Seals, P. Ineichen, R. Stewart, and D. Menicucci, "A New Simplified version of the Perez diffuse irradiance model for tilted surfaces," *Solar Energy*, vol. 39, pp. 221–231, Dec. 1987.

[84] P. Ineichen, R. R. Perez, R. D. Seal, E. L. Maxwell, and A. Zalenka, "Dynamic global-to-direct irradiance conversion models," *ASHRAE transactions*, vol. 98, no. 1, p. 354, 1992.

[85] E. L. Maxwell, "A quasi-physical model for converting hourly global horizontal to direct normal insolation," Solar Energy Research Inst., Golden, CO (USA), Tech. Rep. SERI/TR-215-3087, Aug. 1987.

[86] M. K. Fuentes, "' A Simplified Thermal Model for Flat-Plate Photovoltaic Arrays."

[87] T. A. Olukan and M. Emziane, "A Comparative Analysis of PV Module Temperature Models," *Energy Procedia*, vol. 62, pp. 694–703, Jan. 2014.

[88] K. S. Anderson, C. W. Hansen, W. F. Holmgren, A. R. Jensen, M. A. Mikofski, and A. Driesse, "Pvlib python: 2023 project update," *Journal of Open Source Software*, vol. 8, no. 92, p. 5994, Dec. 2023.

[89] P. Ineichen and R. Perez, "A new airmass independent formulation for the Linke turbidity coefficient," *Solar Energy*, vol. 73, no. 3, pp. 151–157, Sep. 2002.

[90] F. Antonanzas-Torres, R. Urraca, J. Polo, O. Perpiñán-Lamigueiro, and R. Escobar, "Clear sky solar irradiance models: A review of seventy models," *Renewable and Sustainable Energy Reviews*, vol. 107, pp. 374–387, Jun. 2019.

[91] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[92] Y. Yu, X. Si, C. Hu, and J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures," *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019.

[93] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," *Neural computation*, vol. 9, pp. 1735–80, Dec. 1997.

[94] Y. Tian and Y. Zhang, "A comprehensive survey on regularization strategies in machine learning," *Information Fusion*, vol. 80, pp. 146–166, Apr. 2022.

[95] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '19. New York, NY, USA: Association for Computing Machinery, Jul. 2019, pp. 2623–2631.

[96] D. Berrar, "Cross-Validation," in *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier, 2019, pp. 542–545.

[97] M. Schnaubelt, "A comparison of machine learning model validation schemes for non-stationary time series data," FAU Discussion Papers in Economics, Working Paper 11/2019, 2019.

[98] A. Dobos, "PVWatts Version 5 Manual," National Labratory of U.S. Department of Energy, Golden, Technical Report NREL/TP-6A20-62641, 1158421, Sep. 2014.

[99] J. Kratochvil, W. Boyson, and D. King, "Photovoltaic array performance model." Sandia National Laboratories, Albuquerque, Tech. Rep. SAND2004-3535, 919131, Aug. 2004.

[100] T. Huld, G. Friesen, A. Skoczek, R. P. Kenny, T. Sample, M. Field, and E. D. Dunlop, "A power-rating model for crystalline silicon PV modules," *Solar Energy Materials and Solar Cells*, vol. 95, no. 12, pp. 3359–3369, Dec. 2011.

[101] I. de la Parra, M. Muñoz, E. Lorenzo, M. García, J. Marcos, and F. Martínez-Moreno, "PV performance modelling: A review in the light of quality assurance for large PV plants," *Renewable and Sustainable Energy Reviews*, vol. 78, pp. 780–797, Oct. 2017.

[102] D. Yang, S. Alessandrini, J. Antonanzas, F. Antonanzas-Torres, V. Badescu, H. G. Beyer, R. Blaga, J. Boland, J. M. Bright, C. F. M. Coimbra, M. David, Â. Frimane, C. A. Gueymard, T. Hong, M. J. Kay, S. Killinger, J. Kleissl, P. Lauret, E. Lorenz, D. van der Meer, M. Paulescu, R. Perez, O. Perpiñán-Lamigueiro, I. M. Peters, G. Reikard, D. Renné, Y.-M. Saint-Drenan, Y. Shuai, R. Urraca, H. Verbois, F. Vignola, C. Voyant, and J. Zhang, "Verification of deterministic solar forecasts," *Solar Energy*, vol. 210, pp. 20–37, Nov. 2020.

[103] "What Is Random Forest?" https://www.ibm.com/topics/random-forest, Oct. 2021.

[104] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.

[105] A. Jain, H. Patel, L. Nagalapatti, N. Gupta, S. Mehta, S. Guttula, S. Mujumdar, S. Afzal, R. Sharma Mittal, and V. Munigala, "Overview and Importance of Data Quality for Machine Learning Tasks," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Virtual Event CA USA: ACM, Aug. 2020, pp. 3561–3562.

[106] G. P. Compo, J. S. Whitaker, P. D. Sardeshmukh, N. Matsui, R. J. Allan, X. Yin, B. E. Gleason, R. S. Vose, G. Rutledge, P. Bessemoulin, S. Brönnimann, M. Brunet, R. I. Crouthamel, A. N. Grant, P. Y. Groisman, P. D. Jones, M. C. Kruk, A. C. Kruger, G. J. Marshall, M. Maugeri, H. Y. Mok, Ø. Nordli, T. F. Ross, R. M. Trigo, X. L. Wang, S. D. Woodruff, and S. J. Worley, "The Twentieth Century Reanalysis Project," *Quarterly Journal of the Royal Meteorological Society*, vol. 137, no. 654, pp. 1–28, 2011.

[107] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut, "The ERA5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.

[108] S. Rasp, S. Hoyer, A. Merose, I. Langmore, P. Battaglia, T. Russel, A. Sanchez-Gonzalez, V. Yang, R. Carver, S. Agrawal, M. Chantry, Z. B. Bouallegue, P. Dueben, C. Bromberg, J. Sisk, L. Barrington, A. Bell, and F. Sha, "WeatherBench 2: A benchmark for the next generation of data-driven global weather models," Aug. 2023.

[109] J. Muñoz-Sabater, E. Dutra, A. Agustí-Panareda, C. Albergel, G. Arduini, G. Balsamo, S. Boussetta, M. Choulga, S. Harrigan, H. Hersbach, B. Martens, D. G. Miralles, M. Piles, N. J. Rodríguez-Fernández, E. Zsoter, C. Buontempo, and J.-N. Thépaut, "ERA5-Land: A state-of-the-art global reanalysis dataset for land applications," *Earth System Science Data*, vol. 13, no. 9, pp. 4349–4383, Sep. 2021.

[110] Copernicus Climate Change Service, "Complete ERA5 global atmospheric reanalysis," 2023.

[111] ECMWF, "TIGGE archive," https://confluence.ecmwf.int/display/TIGGE.

[112] R. Swinbank, M. Kyouda, P. Buchanan, L. Froude, T. M. Hamill, T. D. Hewson, J. H. Keller, M. Matsueda, J. Methven, F. Pappenberger, M. Scheuerer, H. A. Titley, L. Wilson, and M. Yamaguchi, "The TIGGE Project and Its Achievements," *Bulletin of the American Meteorological Society*, vol. 97, no. 1, pp. 49–67, Jan. 2016.

[113] R. Hagedorn, R. Buizza, T. M. Hamill, M. Leutbecher, and T. N. Palmer, "Comparing TIGGE multimodel forecasts with reforecast-calibrated ECMWF ensemble forecasts," *Quarterly Journal of the Royal Meteorological Society*, vol. 138, no. 668, pp. 1814–1827, 2012.

[114] O. Antonić, J. Križan, A. Marki, and D. Bukovec, "Spatio-temporal interpolation of climatic variables over large region of complex terrain using neural networks," *Ecological Modelling*, vol. 138, no. 1, pp. 255–263, Mar. 2001.

[115] Centre for Environmental Data Analysis, "NWP-Euro: Operational Numerical Weather Prediction (NWP) output from the European Atmospheric High Resolution Model; part of the Met Office Unified Model (UM)."

[116] ——, "NWP-Global: Operational Numerical Weather Prediction (NWP) output from the UK Met Office Global Atmospheric Unified Model (UM). Centre for Environmental Data Analysis."

[117] D. Reinert, Prill, H. Frank, M. Denhard, M. Baldauf, C. Schraff, C. Gebhardt, C. Marsigli, J. Förstner, G. Zängl, and L. Schlemmer, "DWD Database Reference for the Global and Regional ICON and ICON-EPS Forecasting Systemb," Deutscher Wetterdienst, Offenbach Am Main, Tech. Rep., Apr. 2024.

[118] DWD, "NWP forecast data," https://www.dwd.de/.

[119] "Open Climate Fix," https://openclimatefix.org/.

[120] Open Climate Fix, "Dwd-icon-global."

[121] ——, "Dwd-icon-eu."

[122] "Dissemination schedule - Data and Charts - ECMWF Confluence Wiki," https://confluence.ecmwf.int/display/DAC/Dissemination+schedule.

[123] G. Lentze, "Model upgrade increases skill and unifies medium-range resolutions," https://www.ecmwf.int/en/about/media-centre/news/2023/model-upgrade-increases-skill-and-unifies-medium-range-resolutions, Jun. 2023.

[124] "ECMWF," https://www.ecmwf.int/, Mar. 2024.

[125] R. Lam, A. Sanchez-Gonzalez, M. Wilson, P. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, A. Merose, S. Hoyer, G. Holland, O. Vinyals, J. Stott, A. Pritzel, S. Mohammed, and P. Battaglia, "Learning skillful medium-range global weather forecasting," https://www.science.org/doi/10.1126/science.adi2336.

[126] C. Skamarock, B. Klemp, J. Dudhia, O. Gill, Z. Liu, J. Berner, W. Wang, G. Powers, G. Duda, D. Barker, and X.-y. Huang, "A Description of the Advanced Research WRF Model Version 4.3," 2021.

[127] M. Optis, A. Kumler, G. Scott, M. Debnath, and P. Moriarty, "Validation of RU-WRF, the Custom Atmospheric Mesoscale Model of the Rutgers Center for Ocean Observing Leadership," Tech. Rep. NREL/TP–5000-75209, 1599576, Feb. 2020.

[128] B. K. Blaylock, "Herbie: Retrieve Numerical Weather Prediction Model Data," Zenodo, May 2024.

[129] Open Climate Fix, "Uk_pv."

[130] A. Alcañiz, A. V. Lindfors, M. Zeman, H. Ziar, and O. Isabella, "Effect of Climate on Photovoltaic Yield Prediction Using Machine Learning Models," *Global Challenges*, vol. 7, no. 1, p. 2200166, 2023.

[131] "PV power databases," https://www.tudelft.nl/ewi/over-de-faculteit/afdelingen/electrical-sustainable-energy/photovoltaic-materials-and-devices/dutch-pv-portal/pv-power-databases.

[132] P. Zippenfenig, "Open-Meteo.com Weather API," Zenodo, Jan. 2024.

[133] KMNI, "Code rood voor sneeuwstorm Darcy," https://www.knmi.nl/over-het-knmi/nieuws/code-rood-voor-seeuwstorm-darcy, Feb. 2021.

[134] M. Office, "Severe winter weather and storm Darcy, February 2021," Met Office, London, Tech. Rep., Feb. 2021.