UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

Spatio-temporal solar forecasting

" Documento Definitivo"

Doutoramento em Ciências Geofísicas e da Geoinformação
Especialidade de Meteorologia

Rodrigo Amaro e Silva

Tese orientada por:
Professor Doutor Miguel Centeno da Costa Ferreira Brito
Doutora Sue Ellen Haupt

Documento especialmente elaborado para a obtenção do grau de doutor
2019

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

Spatio-temporal solar forecasting

Doutoramento em Ciências Geofísicas e da Geoinformação

Especialidade de Meteorologia

Rodrigo Amaro e Silva

Tese orientada por:

Professor Doutor Miguel Centeno da Costa Ferreira Brito

Doutora Sue Ellen Haupt

Júri:

Presidente:

- Doutor João Manuel de Almeida Serra, Professor Catedrático, Faculdade de Ciências da Universidade de Lisboa

Vogais:

- Doutor Philippe Blanc, *Full Professor* da *MINES Paristech*, Sophia Antipolis (França)
- Doutor Hugo Carreira Pedro, *Associate Project Scientist* do *Center for Energy Research* da *University of California San Diego* (Estados Unidos da América)
- Doutor Ricardo Jorge Gomes de Sousa Bento Bessa, Investigador Sénior do Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência – INESC TEC
- Doutor Miguel Centeno da Costa Ferreira Brito, Professor Auxiliar com Agregação da Faculdade de Ciências da Universidade de Lisboa (orientador)

Documento especialmente elaborado para a obtenção do grau de doutor

Doutoramento financiado pela Fundação para a Ciência e Tecnologia (PD/BD/106007/2014)

2019

Agradecimentos / Acknowledgements

Estes últimos cinco anos representaram, sem dúvida, o maior desafio que enfrentei até agora. Imensas frustrações, mas também um crescimento pessoal e científico que superou quaisquer expectativas.

Queria deixar aqui explícito o papel positivo e decisivo de um conjunto vasto de pessoas no que foi fazer (e sobreviver a) este doutoramento.

Aos meus pais, ao meu irmão, à minha família, à Rita. Por tudo.

Ao Miguel Centeno Brito. Primeiro "professor", agora "Miguel".

Ao Instituto Dom Luiz pela oportunidade e, conjuntantemente com o grupo *Energy Transition* da Faculdade de Ciências da Universidade de Lisboa, pelo acolhimento.

À Ana Cunha, ao Javier Mellado, à Catarina Matos, ao Federico Ienna, ao Vasco Conde, ao Fabiano Mota. Companheiros, no bom e no mau. Ana e Javier, sempre juntos.

À malta do "Copos" (e mais alguns). Se voltasse atrás no tempo, voltava a escolher o mesmo curso. Por vossa causa. Não podia pedir mais.

À Joana, à Raquel, à Sara, ao Ivo, ao David, à Marta, ao Ângelo, ao Alex, ao Riccardo e à Filipa. Por tornarem o meu dia-a-dia divertido e me aturarem.

À Catarina Guerreiro, ao Pedro Almeida, ao Roberto… enfim à malta da má vida que foi particularmente importante numa fase inicial, e bem divertida, deste processo.

To Jamie Taylor, for being so open and generous. You provided me the data that unlocked my first published article.

To Sue Ellen Haupt, Laura Clemente-Harding, David John Gagne II, Ricardo Aler and David-Pozo Vázquez. Sue you allowed me to have one of the most relevant professional experiences I have had so far. Everyone, you made me feel integrated in a new environment and every day you made me want to learn more and discuss more about science.

To Elena, Nicola (Nico), Paolo, Francesca and Alex. My Italian crew. You were the friends I needed for my U.S. experience to thrive. I won't forget all those hang outs and amazing trips.

Ao Miguel Heleno e ao Gonçalo Cardoso. Pela aventura de um mês na Califórnia. Gonçalo, não me vou esquecer de como me recebeste. Nem da cerveja que ainda me deves.

Ao Ricardo Bessa. Pelo espírito crítico e inestimável *feedback* num início em que ainda estava a tentar perceber o que era um doutoramento.

To Wilfried van Sark and Atse Louwen. For enabling my two-week experience in Utrecht and providing important feedback to my research.

Ao Emanuel Dutra. Pelo importante *feedback* na ponta final desta experiência.

Ao Guilherme e à Inês. Pelo novo desafio que me proporcionaram. Não imaginam o quão importantes foram para que acabasse o doutoramento.

Abstract

Current and future photovoltaic (PV) deployment levels require accurate forecasting to ensure grid stability. Spatio-temporal solar forecasting is a recent solar forecasting approach that explores spatially distributed solar data sets, either irradiance or photovoltaic power output, modeling cloud advection patterns to improve forecasting accuracy. This thesis contributes to further understanding of the potential and limitations of this approach, for different spatial and temporal scales, using different data sources; and its sensitivity to prevailing local weather patterns.

Three irradiance data sets with different spatial coverages (from meters to hundreds of kilometers) and time resolutions (from seconds to days) were investigated using linear autoregressive models with external inputs (ARX). Adding neighboring data led to accuracy gains up to 20-40 % for all datasets. Spatial patterns matching the local prevailing winds could be identified in the model coefficients and the achieved forecast skill whenever the forecast horizon was of the order of scale of the distance between sensors divided by cloud speed.

For one of the sets, it was shown that the ARX model underperformed for non-prevailing winds. Thus, a regime-based approach driven by wind information is proposed, where specialized models are trained for different ranges of wind speed and wind direction. Although forecast skill improves by up to 55.2 % for individual regimes, the overall improvement is only of 4.3 %, as those winds have a low representation in the data.

By converting the highest resolution irradiance data set to PV power, it was also shown that forecast accuracy is sensitive to module tilt and orientation. Results are shown to be correlated with the difference in tilt and orientation between systems, indicating that clear-sky normalization is not totally effective in removing the geometry dependence of solar irradiance. Thus, non-linear approaches, such as machine learning algorithms, should be tested for modelling the non-linearity introduced by the mounting diversity from neighboring systems in spatio-temporal forecasting.

Keywords: solar forecasting; spatio-temporal solar forecasting.

Resumo

Perspetiva-se um crescimento acentuado da capacidade instalada de tecnologias fotovoltaicas nos próximos anos. Para acomodar a considerável variabilidade temporal que lhe é intrínseca, os operadores da rede elétrica necessitarão de modelos preditivos de produção fotovoltaica. Métodos espácio-temporais para previsão solar são uma abordagem recente que explora séries temporais distribuídas no espaço de recurso solar e/ou produção fotovoltaica, procurando modelar os padrões de advecção das nuvens de forma a gerar previsões de melhor qualidade.

Esta tese contribui para uma melhor compreensão do potencial e limitações destes métodos, para vários escalas de tempo e do espaço, assim como da sua sensibilidade aos padrões meteorológicos locais. Três conjuntos de dados correspondentes a diferentes escalas espaciais (de alguns metros a centenas de quilómetros) e temporais (de alguns segundos até vários dias) são explorados, utilizando modelos lineares regressivos com entradas externas (ou *linear autoregressive models with external inputs*, ARX).

Para os vários casos de estudo, a inclusão de registos passados de locais vizinhos pode melhorar a qualidade de previsão, em termos de erro quadrático médio (RMSE), até 20-40%. Sempre que o horizonte de previsão é da ordem de escala da distância entre sensores dividida pela velocidade do vento, padrões espaciais que correspondem aos ventos predominantes de cada local podem ser identificados nos valores de *forecast skill* (i.e. o valor acrescentado de um dado modelo relativamente a outro de referência, neste caso em termos de RMSE e comparativamente a um modelo de *smart persistence*) obtidos para os vários locais. Neste contexto, a qualidade do modelo está dependente da disponibilidade de informação vizinha adequada (i.e. do sensor a prever e dos seus vizinhos), sendo esta dependência mais acentuada no conjunto de dados com menor resolução temporal (por exemplo, a diferença de *forecast skill* entre os locais com melhor e pior desempenho para um horizonte de previsão de 10 segundos é de 31.7 %). A distribuição espacial dos valores dos coeficientes de regressão reforça esta ideia, demonstrando que: i) quando disponível, o modelo atribui maior relevância ao local para onde se realiza a previsão e a vizinhos que se encontrem na direção de onde sopra o vento; ii) à medida que o horizonte de previsão aumenta, o modelo atribui maior relevância a vizinhos mais distantes; iii) quando não existe informação vizinha adequada, o modelo adopta uma abordagem semelhante à persistência, atribuindo menor relevância à informação vizinha.

Por outro lado, quando o horizonte de previsão ultrapassa esta escala, o desempenho do modelo espácio-temporal torna-se mais homogéneo para os vários sensores, independentemente da existência de sensores a barlavento. Isto porque, neste caso, o modelo atribui igual relevância a toda a informação vizinha, adotando uma abordagem semelhante à de uma média no espaço e no tempo. Deste modo, o modelo de previsão apresenta previsões postura mais conservadoras do estado geral da atmosfera, um análogo à previsão meteorológica com horizontes mais alargados que se aproxima de previsões climatológicas.

Um dos casos de estudo analisados neste trabalho consiste num ano e meio de valores médios de radiação global horizontal a cada 10 segundos, medidos por 16 piranómetros numa área de 1×1 km^2. O modelo ARX explora as correlações espácio-temporais que os ventos predominantes introduzem nos dados, superando, em média, o modelo de persistência. No entanto, quando o *forecast skill* é desagregado em função da direção do vento, são observados maus desempenhos de previsão quando esta se desvia da direção predominante, apresentando valores negativos (até -44.7 %).

Para mitigar este efeito, é proposta uma abordagem que procura explorar regimes de vento através de modelos especializados treinados e aplicados para intervalos específicos de direção e/ou velocidade do vento. Como a distribuição dos vários padrões de vento não é homogénea no tempo, uma separação menos comum dos dados em subconjuntos de treino e teste é utilizada. Quando os modelos são especializados para diferentes intervalos de direção do vento, o *forecast skill* melhora para todos os regimes definidos. No entanto, não só em vários casos estes valores continuam a ser relativamente modestos (num deles mantém-se, inclusivamente, negativo), como em nenhum dos casos o modelo mostra ser capaz de identificar novos padrões espaciais. O modelo mostra ser capaz de selecionar a informação vizinha correspondente à direção de onde o vento tem origem quando a definição de regimes tem também em conta a velocidade do vento, separando cada regime de direção em dois subregimes delimitados por um valor fixo de velocidade do vento. Apesar de, individualmente, os regimes melhorarem o seu *forecast skill* até 55.2 %, o ganho global é de apenas 4.3 %. Isto ocorre devido à reduzida porção dos dados que corresponde aos regimes onde se verificou um maior impacto. Contudo, é de esperar que esta abordagem baseada em regimes tenha um impacto superior em locais onde os ventos predominantes não tenham uma direção tão frequente.

Para este mesmo conjunto de dados foi também avaliado o impacto da inclinação e orientação particular dos sensores/módulos fotovoltaicos numa rede de informação espacialmente distribuída. Este estudo permite explorar o desempenho do modelo espácio-temporal quando são utilizados dados de geração fotovoltaica, em particular em meio urbano.

Devido à inexistência de dados de geração fotovoltaica com resoluções temporais da ordem dos segundos, o conjunto de dados de radiação solar global medidos a cada 10 segundos foi convertido em geração fotovoltaica utilizando modelos de decomposição e transposição *state-of-the-art*. Foram simulados 4004 conjuntos de dados, considerando diferentes valores de inclinação e orientação para duas tipologias de instalação (em telhado ou fachada). O desempenho do modelo de previsão mostrou estar correlacionado com a diferença de inclinação e orientação entre sistemas vizinhos, indicando que a normalização dos dados, relativamente a um perfil de céu limpo, não é totalmente eficaz em remover a sua dependência geométrica. Dependendo do conjunto de valores de inclinação e orientação considerados, o *forecast skill* de um modelo pode variar em termos absolutos até 10 %. Modelos de previsão treinados para sistemas instalados em fachadas (i.e. na vertical) obtêm *forecast skill* inferiores, um resultado atribuído ao fato destes sistemas estarem mais dependentes da componente difusa (anisotrópica e mais complexa de modelar) da radiação solar. Por este mesmo motivo, quando os vizinhos de um sistema vertical correspondem a instalações em telhados, o modelo beneficia que estes tenham uma geometria o mais diversa possível.

Ao longo deste trabalho foram identificadas potenciais linhas de pesquisa futura. O modelo ARX linear deve ser testado com conjuntos de dados solares espacialmente distribuídos correspondentes a locais com ventos predominantes menos frequentes. Devem ser ainda avaliadas abordagens de regressão não-linear, tais como algoritmos de aprendizagem automática (p.e. redes neuronais artificiais ou máquinas de vetores de suporte), comparando-os com o modelo ARX linear. A reconhecida capacidade destes modelos em capturar padrões complexos pode aumentar o valor de conjuntos dados solares espacialmente distribuídos, indo mais além do que procurar correlações espácio-temporais dominantes. Por último, métodos espácio-temporais devem ser implementados em conjuntos de dados de geração fotovoltaica, complementados pelas análises espácio-temporais realizadas ao longo desta tese. O grande potencial desta fonte de dados, devido à elevada densidade de sistemas fotovoltaicos que se espera que venha a ocorrer tanto em meio urbano como em regiões mais isoladas, tem de ser adequadamente avaliado, incorporando condicionantes como

sombreamento por obstruções vizinhas, diferentes eficiências e nível de degradação dos módulos fotovoltaicos e inversores entre sistemas, ou fenómenos não-lineares como *inverter clipping*.

Palavras-chave: previsão solar; previsão solar espácio-temporal.

Contents

1. Introduction 4
   1.1. Motivation 4
      1.1.1. Power systems operation 4
      1.1.2. Impacts of a high PV penetration in power systems 5
      1.1.3. High-resolution forecasting and spatially distributed solar data 7
   1.2. Thesis research questions 9
   1.3. Thesis outline 9
2. Spatio-temporal solar forecasting 12
   2.1. Solar forecasting overview 12
   2.2. Concept 14
      2.2.1. Numerical Weather Prediction 15
      2.2.2. Image-based methods 17
         2.2.2.1. Satellite imagery 17
         2.2.2.2. Sky camera imagery 18
      2.2.3. Autoregressive time series methods 21
         2.2.3.1. Persistence and solar data detrending 21
         2.2.3.2. Linear and non-linear methods 23
      2.2.4. Spatio-temporal time series methods 23
3. Testing various spatio-temporal scales 27
   3.1. Introduction 27
   3.2. Case studies 28
      3.2.1. Oahu and very short-term forecasting 28
      3.2.2. Oklahoma and short-term forecasting 29
      3.2.3. Oklahoma and day-ahead forecasting 29
   3.3. Data pre-processing 30
   3.4. Forecasting model 31
   3.5. Assessment metrics 32
   3.6. Results 33
      3.6.1. Forecast skill as a function of forecast horizon 33
      3.6.2. Spatio-temporal analysis 34
   3.7. Final Remarks 39

4. Integrating wind information in statistical spatio-temporal solar forecasting models 41
4.1. Introduction 41
4.2. The Oahu case study 41
4.3. A proposal for a regime-based forecasting approach 43
4.3.1. Defining regimes based on wind direction 43
4.3.2. Defining regimes based on wind direction and speed 44
4.3.3. Testing the regime-based approach for all sites 46
4.3.4. Impact of wind forecast uncertainty 47
4.4. Final remarks 47
5. Spatio-temporal approaches sensitivity to modules' tilt and orientation 49
5.1. Introduction 49
5.2. Simulating a network of differently arranged surfaces 50
5.2.1. Global irradiance decomposition and transposition 50
5.2.2. Modelling PV optical losses 52
5.2.3. Limitations 53
5.3. Case studies 53
5.3.1. The impact of tilted surfaces: one-neighbor model 54
5.3.2. The impact of diverse inclinations: many-neighbors model 56
5.4. Final remarks 59
6. Conclusions 61

List of Figures

Figure 1.1 – Ramp rate distributions for load and PV ramps for a single house and the aggregate of 482 houses in the month of January 2007 in Ota City, Japan. The arrows illustrate the reduction in variability due to aggregation (Lave et al., 2016). 6

Figure 1.2 – Relation between horizons, models and activities (Diagne et al., 2013). 8

Figure 2.1 – Basis of all NWP models. First, a domain is defined. Secondly, the domain is spatially discretized at the desired resolution. Finally, the NWP predicts desired information by solving equations of motion and thermodynamics (Inman et al., 2013). 15

Figure 2.2 – Image from the Meteosat-9 satellite and corresponding CMVs. Adapted from Cros et al. (2014). 17

Figure 2.3 – Meteosat spatial resolution (MINES ParisTech, 2009). 18

Figure 2.4 – Raw sky camera image (left); processed image for cloud identification (center); and inferred CMV (right). Adapted from Schmidt et al. (2017). 19

Figure 2.5 – Example of spectral reflectance (left) and camera occlusion from dew or rain (right) on a sky camera. Adapted from Siddiqui et al. (2019). 20

Figure 2.6 – On the left, a pyranometer array from Bosch et al. (2013); on the right: a network of reference cells from Bosch and Kleissl (2013). Cloud advection patterns could be derived from each of the data sets. 24

Figure 2.7 – Pyranometer network deployed in Melpitz, Germany, for the HOPE project. Adapted from Macke et al. (2017) 25

Figure 2.8 – From Lorenzo et al. (2015), an interpolated clear-sky index map (left), which is shifted according to a wind vector in order to generate a forecast (right). 25

Figure 3.1 – Classification of forecasting approaches based on spatial and temporal resolution. Adapted from (Lorenz, 2014) and (Inman et al., 2013). 27

Figure 3.2 –Spatial distribution of the pyranometers from the Oahu Solar Measurement Grid. 28

Figure 3.3 – Satellite grid from the NSRDB, covering part of the Oklahoma state (USA) and centered at the (97.54W, 35.41N) coordinates. 29

Figure 3.4 - Spatial distribution of the pyranometers from the Oklahoma Mesonet, centered at the (98.76W, 35.41N) coordinates. 30

Figure 3.5 - Forecast skill distribution for the various sites of each data set (Oahu in blue, NSRDB in red and Mesonet in green. and for several horizons. The dashed line indicates the median value for each horizon. Above each distribution the spatial coverage and time resolution of each data set is indicated. 33

Figure 3.6 – Forecast skill for three different data sets: 10s-ahead for Oahu (top left); 30 min-ahead for Oklahoma NSRDB (top right) and 24 h-ahead for Oklahoma Mesonet (bottom). .. 35

Figure 3.7 – Site relevance (calculated using eq. 3.4) for the ARX model trained with 10-s data. The model was assessed for two different target sites, highlighted in red in each plot, and for 10 s ahead. 36

Figure 3.8 – Site relevance (calculated using eq. 3.4) for the ARX model trained with 10-s data. The model was assessed for two different target sites, one per row and highlighted in red, for 30 s (left), 3 min (center) and 5 min (right) ahead. 37

Figure 3.9 – Oklahoma NSRDB data set forecast skill for 30 min and 3 h ahead forecasts The higher skill region is considerably thinner for the latter, plausibly due to the reduced number of sites that have neighbors that are sufficiently distant. 38

Figure 3.10 – The nWAD metric as a function of the forecast horizon for the Oahu data set (in light blue the peripheral upwind sites and in dark blue the remaining sites), the Oklahoma NSRDB set (in red) and the Oklahoma Mesonet set (in green). 39

Figure 4.1 – Oahu's wind speed and direction histogram at 850 hPa, with predominance of east and northeast winds faster than 5 m/s. 42

Figure 4.2 - Forecast skill (radial axis) as a function of wind direction for the Oahu site highlighted in red Figure 3.7 (left plot). Positive skills (in green) can be seen for northerly and easterly winds, whereas south and west show considerable negative skill values (in red). 43

Figure 4.3 – Example of the regime-specific train-test splitting, in this case for easterly winds. 43

Figure 4.4 - Forecast skill (radial axis) as a function of wind direction for two ARX models: baseline version (left) and one built according to wind direction regimes (right). Positive/negative forecast skill values are shown in green/red. The left plot corresponds to Figure 4.2 but for an alternative train-test configuration. 44

Figure 4.5 - Regression coefficients map for the clustered data with wind speed > 9 m/s. Winds coming from the four cardinal directions are shown, with clear spatial patterns observed. The target sensor is represented by a red circle. 45

Figure 4.6 – Forecast skill (radial axis) according to wind direction and the implications of the wind regime definition. Negative forecast skill values are shown in red, positive forecast skills shown in green. On the bottom line and for the first and last columns, the negative skill value for westerly winds was saturated for visualization purposes. 46

Figure 5.1 – Forecast skill (in %) for a forecast target with a given POA, considering its upwind sensor as a horizontal pyranometer. 54

Figure 5.2 – Overall diffuse fraction (in %) for the site highlighted in red in Figure 3.7 assuming various tilts and orientations. The observed pattern appears to be symmetric to the one shown in Figure 5.1. 55

Figure 5.3 – Forecast skill for a PV surface with a given tilt and orientation, assuming for the upwind site either an optimum or vertical tilt, and three possible orientations Façades appear to be more sensitive to POA than non-vertical surfaces. 56

Figure 5.4 – Forecast skill range for the scenarios 1A and 1B (top row) and 2A and 2B (bottom row), when considering either 2 sites (forecast target and closest upwind sensor) or all 16 sites. Blue and green boxplots correspond, respectively, to rooftop and vertical neighbors. The circles represent an “ideal” case where all sites share the same tilt/orientation. 59

List of Tables

Table 2.1 – List of review and benchmark scientific papers on solar forecasting 13

Table 3.1 – Considered training and test periods for each of the three tested data sets. 32

Table 3.2 – List of data sets considered in this chapter and their spatio-temporal resolutions. The Oahu data set is also considered in the following two chapters. 33

Table 5.1 – Parameters for the Engerer2 decomposition model. Adapted from Engerer (2015). 51

Table 5.2 – Empirical coefficients to calculate the optical losses for each irradiance component. Adapted from Marion (2017). 52

Table 5.3 - Upwind site tilt and orientation maximum deviation, when compared to the forecast target's mounting, which results in a maximum 15 % relative reduction. 56

Table 5.4 – Scenarios tested, where either the forecasting target or the neighboring sites' POA is varied. 57

Table 5.5 – Forecast skill percentiles and interquartile range (IQR) from the 1000 POA sets. Two forecast target tilts, four different scenarios pertaining to the neighbors' POA distribution and two different amounts of sites were considered for the linear ARX model. An "ideal" case, where all the sites share the same POA, was also tested. 58

Nomenclature and Abbreviations

Variable/parameter	Description
PV	Photovoltaics
USA	United States of America
NWP	Numerical Weather Prediction
CMV	Cloud motion vector
NCEP	National Centers for Environmental Prediction
GFS	Global Forecast System
WRF	Weather Research Forecast
NCAR	National Center for Atmospheric Research
MOS	Multiple output statistics
RMSE	Root mean squared error
HRRR	High-Resolution Rapid Refresh
NOAA	National Oceanic and Atmospheric Administration
STEAM	Satellite Earth observation for atmospheric modelling
CSL	Clear-sky library
RBR	Red-blue ratio
RBD	Red-blue deviation
GHI	Global horizontal irradiance
$\widehat{GHI}$	Global horizontal irradiance forecast
h	Forecast horizon
t	Time
$W.m^{-2}$	Watt per squared meter
K_t	Clearness index
TOA_h	Top of atmosphere irradiance for an horizontal surface
K_c	Clear-sky index
GHI_c	Global horizontal irradiance in clear-sky conditions
$\widehat{K_t}$	K_t forecast
AR	Autoregressive
ARMA	Autoregressive moving average
ARIMA	Autoregressive integrated moving average
SARIMA	Seasonal autoregressive integrated moving average
ARX	Autoregressive using exogenous inputs
LASSO	Least absolute shrinkage and selection operator
NREL	National Renewable Energy Laboratory
NSRDB	National Solar Radiation DataBase
c	Regression coefficient (for linear ARX model)
b	Regression bias term (for linear ARX model)
l	Time lag (for linear ARX model)
MBE	Mean bias error

MAE	Mean absolute error
RMSE	Root mean squared eror
KSI	Komolgorov-Smirnov test integral
FS	Forecast skill
WAD	Weighted average distance
nWAD	Normalized weighted average distance
ECMWF	European Centre for Medium-Range Weather Forecasts
POA	Plane-of-array
G_{POA}	Global irradiance at a given plane-of-array
BHI	Beam horizontal irradiance
DHI	Diffuse horizontal irradiance
f_1	Beam horizontal irradiance transposition function
f_2	Diffuse horitonzal irradiance transposition function
f_3	Reflected irradiance transposition function
θ	Irradiance incidence angle
ς	Solar zenith angle
β	Surface tilt or inclination angle
F_1	Degree of circumsolar anisotropy (Perez transposition model)
F_2	Degree of horizon anisotropy (Perez transposition model)
ρ	Ground albedo
K_d	Diffuse fraction
AST	Apparent solar time
α	Solar elevation angle
B_n	Engerer2 diffuse fraction model coefficients
ΔK_{tc}	Deviation between K_t and the corresponding clear-sky situation
K_{tc}	Clearness index in clear-sky conditions
CAMS	Copernicus Atmosphere Monitoring Services
K_{de}	GHI fraction driven by cloud enhancement
OL	Optical losses
a_r	Angular losses coefficient
a_n	Empirical coefficients for optical losses calculation
IQR	Interquartile range
LES	Large eddy simulation

1. Introduction

1.1. Motivation

1.1.1. Power systems operation

An electric power system is composed of loads and generators interconnected by a grid infrastructure and market agents. Its ideal operation is such that, at every given moment, the electric energy demand is met in the most reliable and cost-effective manner.

In practice, economic efficiency is achieved through various markets that are put into place. First, capacity markets address the future total and peak load, creating long-term price signals that attract investors to maintain current power plants and develop new ones. On the other hand, wholesale electricity markets work as platforms that integrate bids from both energy suppliers and producers (i.e. expected load from consumers and generation from power plants), which then defines the wholesale electricity pricing.

The word “expected” is used twice in the previous paragraph: for power generation and for load. It points out how capacity and energy markets are sustained by forecasts since they are operated ahead of time.

Energy markets usually operates at two different time scales:

- Day-ahead, where both suppliers and producers make a single bid for the following day, typically with an hourly resolution. This is where most of the energy is transacted and the slow-starting generators can schedule their operation with enough time in advance.
- Intra-day, where energy suppliers and managers of variable power plants can adjust, several times a day, their initial bids. Additionally, managers of fast starting generators also bid for their generators to match any unexpected load surge. This market operates on shorter time scales, as better forecasts can be made with more up to date information.

These markets promote reliability in the power grid operation by facilitating the match between power generation and load, while generating relevant information for grid operators to base their actions on. However, variability and forecast inaccuracy of both load and power generators must be handled; hence, operating reserve markets are set up, allowing generators with considerable ramping ability, which are either on stand-by or working on sub-optimal conditions (those that do not lead to maximum efficiency and/or equipment longevity), bid for

the availability to compensate potential imbalance (i.e. a mismatch between energy supply and demand). Grid reliability can also be ensured from the demand side, through deferable loads or dynamic pricing strategies. These markets operate at much shorter temporal scales, from seconds to a few minutes.

There are also contingency reserve markets for standby generators that have the role of mitigating any potential generator tripping. These reserves are secured across multiple time scales.

1.1.2. Impacts of a high PV penetration in power systems

Over recent years, renewable generation has been growing worldwide. Wind and solar photovoltaics (PV) increasing capacity raise concerns due to their non-dispatchable and variable nature; wind or solar power plant managers cannot freely control their output as it is dependent on a natural variable resource (wind speed or solar irradiance)[1].

Wind deployment is already steadfast, and power grids are already capacitated in handling its variability. However, until now, PV power was almost negligible from the grid operation perspective, as the installed capacity was considerably low. Only recently did the technology become cost-competitive. A substantial increase in installed capacity is expected, not only for large PV plants but also for distributed small-scale systems, creating new energy paradigms such as self-consumption and prosumerism. Thus, PV integration entails new challenges for grid operation, mostly at low voltage levels where power generation was previously inexistent.

At the feeder level, aggregated PV has been shown to be substantially more variable than aggregated load for Ota City, Japan (Lave et al., 2016). According to the authors, the load is considerably more decorrelated in space than PV, making the latter less responsive to spatial smoothing (Figure 1.1). Moreover, for high renewable penetration scenarios simulated for the western region of the United States of America (USA), solar (PV and concentrated solar power) has been shown to output more variable generation, and corresponding net load, than wind (Lew et al., 2012a).

[1] However, PV plant operators can reduce their power output by adjusting their point of operation. This control has been demonstrated to be fast enough so that utility PV can provide frequency services (Morjaria et al., 2017).

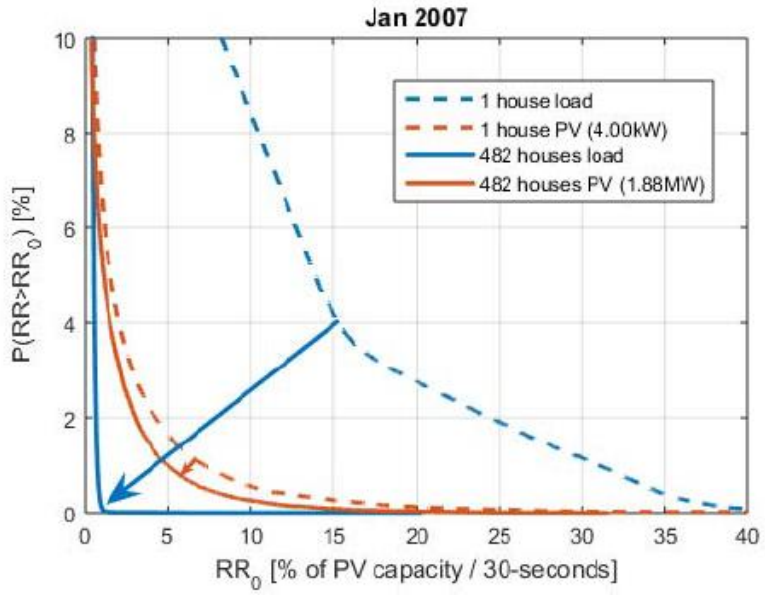

Figure 1.1 – Ramp rate distributions for load and PV ramps for a single house and the aggregate of 482 houses in the month of January 2007 in Ota City, Japan. The arrows illustrate the reduction in variability due to aggregation (Lave et al., 2016).

Solar variability is partly due to the apparent movement of the sun in the sky but, mostly, weather-driven (i.e. the passing of clouds). The first case can lead to power variations up to 10-13 % for a 15-min time period, whereas passing clouds can cause power ramps exceeding 60 % in a matter of seconds (Mills et al., 2011). Example distributions of shading frequency, shading strength and its duration are presented in Lappalainen and Valkealahti (2015) and Tomson (2010).

The underestimation of these fast cloud-driven ramps results in a sub-optimal operation and management of a power grid. Ela et al. (2013) studied the Arizona Public Service balancing area (Arizona, USA) simulating various power sources and regulating reserves with a 1-min resolution. The authors indicate that the integration of almost 3 GW of PV would lead to a 31.2 % increase in the average imbalance, compromising grid reliability.

Several solutions exist to mitigate the imbalance caused by PV: varying the power output and regularly turning on and off fossil-fueled generators (Lew et al., 2012b); inducing or directly shifting the load from the consumption side (Palensky and Dietrich, 2011); curtailing solar generation (Lew et al., 2013); increasing the capacity of operating reserves (Brinkman et al., 2012; Ibanez et al., 2012), storage included (Palizban and Kauhaniemi, 2016); and adapting power grid management, supported by faster energy markets and larger balancing areas (Milligan and Kirby, 2010).

Most of these options either imply higher grid operation costs or lower generation revenues. Some are also unable to tackle the fast variability of solar, which requires a high ramping ability. Additionally, fossil-fueled generators (e.g. gas, which plays a relevant role in current operating reserves) are incompatible with 100 % renewable generation scenarios.

Ela et al. (2013) identified PV variability (i.e. the actual variation in power generation) and its forecast uncertainty (i.e. the deviation between the forecasted and actual generation) as the drivers for the 31.2 % imbalance for a 3 GW scenario in Arizona. The second factor has been shown to be the most relevant, as the imbalance could be reduced to 10.6 % if generator scheduling were based on perfect day-ahead and short-term hourly forecasts.

Forecasting plays a crucial role in management tools and protocols of the power system operator (Bessa et al., 2014) and it is considered as one of the most cost-effective and easily implementable tools to handle PV variability (Zack, 2017). It can also leverage other energy services such as storage scheduling (Litjens et al., 2018) and demand response (Masa-Bote et al., 2014). Thus, it is only natural that the development of accurate forecasting models was defined as one of the milestones for achieving a grid-friendly PV penetration (International Energy Agency, 2014).

1.1.3. High-resolution forecasting and spatially distributed solar data

In the day-ahead energy markets, load and, for some generators, generation bids rely on weather forecasts provided by numerical weather prediction (NWP) models. Typically, these forecasts have a coarse spatial and temporal resolution and are slow to ingest new data (most operational NWP models are typically run once every 6 to 12 h)[2]. On the other hand, the complementary intra-day energy and grid services provision markets, with several bidding periods per day, allow power plant operators to correct their earlier day-ahead bids, and grid operators to dispatch operating reserves. Thus, these more dynamic markets play a major role in achieving a cost-effective grid operation.

For shorter horizons and more detailed resolutions, solar forecasters can leverage satellite or ground-based data to generate more accurate forecasts. Newer satellites can offer temporal and spatial resolutions down to 5 min and 0.5-1 km, respectively, playing a role in short-term

[2] It is important to mention the existence of recent high-resolution and region-specific NWP models that offer spatial resolutions below 1 km and with rapid-update cycles (in USA and at a pilot level in Europe). This is further elaborated in section 2.2.1.

forecasting (Engerer et al., 2018; Miller et al., 2018). For even more detailed resolutions, two traditional approaches exist: autoregressive (AR) and advection models.

AR approaches (such as persistence) using in-situ measurements can accurately forecast the variability's deterministic component (i.e. how the sun changes position with time) and performs rather well in stable weather (either very sunny or very cloudy) conditions. However, AR models lack spatial information, and thus cannot predict the arrival of incoming clouds.

Advective models, based on whole-sky images that are shifted in space according to a cloud motion vector (CMV) and then converted to irradiance (or PV generation), have shown very promising results in predicting cloud advection at the seconds to minutes time scale. Nonetheless, they still face some challenges in terms of image processing (Siddiqui et al., 2019).

Figure 1.2 highlights how different forecasting horizons address different markets and grid operation activities, and are addressed by different models.

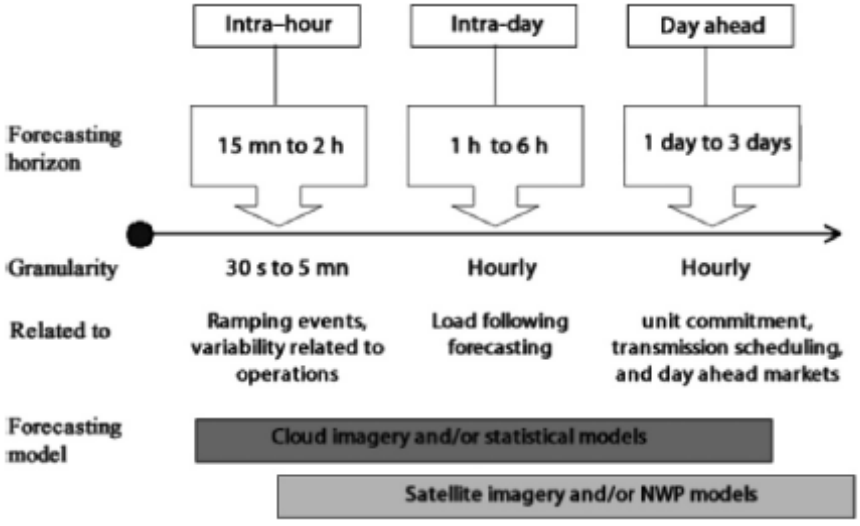

Figure 1.2 – Relation between horizons, models and activities (Diagne et al., 2013).

The expected mass deployment of PV systems will surely result in a growing need for accurate forecasts at the regional and household levels, as well as generate high-resolution data that can feed forecasting models. This thesis focuses on spatio-temporal approaches, an emerging forecasting approach to detect local weather patterns from the exploration of spatially-distributed solar (i.e. irradiance and/or PV generation) time series.

1.2. Thesis research questions

The purpose of this thesis is to provide a better understanding of spatio-temporal solar forecasting, addressing the main research question: what are the potential and limitations of spatio-temporal solar forecasting?

After a thorough literature review, summarized in Chapter 2, three more specific topics pertaining to this question arose, each leading to a more specific research question:

a) Most works in the literature explore individual data sets with specific spatial and temporal resolutions. Not only do the forecasting models tend to lack a spatio-temporal analysis, but none implements the same model and analysis to different spatial and temporal scales at the same time.

Thus, Chapter 3 addresses the question, for which spatial and temporal scales is spatio-temporal forecasting effective?

b) Cloud advection is driven by wind speed and direction at cloud height. However, statistical spatio-temporal forecasting models seldom use these variables. Moreover, those that do report modest improvements.

Thus, Chapter 4 addresses the question, what is the relevance of wind information for spatio-temporal forecasting using statistical models and how can it be integrated into such models?

c) Although PV data is expected to be the main driver for spatio-temporal methods in the near future, most works on very-short term forecasting explore irradiance time series and ignore the challenges of when such data is used.

Thus, Chapter 5 discusses this issue and addresses one particular challenge, how does the varying geometry from PV systems impact spatio-temporal forecasting?

1.3. Thesis outline

This thesis is composed of six chapters. In Chapter 1, the topics of power systems operation, the growing need for solar forecasts and the relevance of spatio-temporal forecasting models are introduced. In Chapter 2, an overview of the current status of solar forecasting research is presented, followed by a discussion on the concept of "spatio-temporal" forecasting. Each of the following three chapters addresses a specific research question. In Chapter 3, a linear

spatio-temporal model is trained using three different data sets, with different spatial coverages and and time resolutions. The resulting forecasts' skill, as well as the coefficients of the linear model, are subjected to a spatial analysis. Moreover, a metric is proposed to better understand how the dynamics of the forecasting model change with the considered temporal resolution and the availability of upwind information. In Chapter 4, the need for spatio-temporal approaches to integrate cloud-height wind information is addressed. Using one of the data sets from the previous chapter, the model is shown to underperform for certain wind directions. Thus, a regime-based approach that is able to detect varying cloud advection patterns is proposed. In Chapter 5, the challenges of spatio-temporal solar forecasting driven by PV data are discussed. Moreover, the impact of a variable PV system geometry (tilt and azimuth angle) is assessed for very-short term forecasting. Since spatially distributed high-resolution PV data are scarce, an irradiance data set was converted to tilted PV power using state-of-the-art irradiance decomposition and transposition techniques. To conclude, Chapter 6 provides an overview of the work presented throughout this thesis and identifies future research lines on spatio-temporal solar forecasting.

This thesis is built on several articles published in international scientific journals and conference proceedings that occurred over the course of this PhD period. A chronologic list of these publications follows:

- **Amaro e Silva, R.**, Brito, M.C., 2017. Understanding Spatio-temporal Solar Forecasting. Oral presentation at the 7th Solar Integration Workshop, Berlin, Germany;
- **Amaro e Silva, R.**, Brito, M.C., 2018. Impact of network layout and time resolution on spatio-temporal solar forecasting. Solar Energy 163, 329–337. doi:10.1016/j.solener.2018.01.095;
- **Amaro e Silva, R.**, Brito, M.C., 2019. Spatio-temporal PV forecasting sensitivity to modules' tilt and orientation. Applied Energy 255, 113807. doi:10.1016/j.apenergy.2019.113807;
- **Amaro e Silva, R.**, Haupt, S.E., Brito, M.C., 2019. A regime-based approach for integrating wind information in spatio-temporal solar forecasting models. Journal of Renewable and Sustainable Energy 11, 056102. doi:10.1063/1.5098763.

Other scientific papers and proceedings not included in this thesis:

- Aido, M., **Silva**, **R.,** Brito, M., 2016. Solar radiation in schools, Poster at the International European Citizen Science Association Conference, Berlin, Germany;
- Freitas, S., Cristóvão, A., **Silva, R.**, Brito, M.C., 2016. Obstruction surveying methods for PV applications in urban environments. Poster presentation at the 32nd EUPVSEC, Munich, Germany;
- Passos, P., **Silva, R.**, 2016. Players' relative position to characterize affordances landscape in football. Oral presentation at the 6th TGfU Conference, Cologne, Germany;
- Silva, J.A., **Amaro e Silva, R.**, Peral, A., del Cañizo, C., 2018. A One Step Method to Produce Boron Emitters. Phys. Status Solidi A. doi:10.1002/pssa.201701076;
- **Amaro e Silva, R.**, Teixeira da Silva, L.C.C., Brito, M.C., 2018. Support Vector Regression for spatio-temporal PV forecasting. Oral presentation at the 35th EUPVSEC, Brussels, Belgium;
- **Amaro e Silva, R.**, Monteiro Baptista, J., Brito, M.C., 2018. Data-driven estimation of expected photovoltaic generation. Sol. Energy 166, 116–122. doi:10.1016/j.solener.2018.03.039;
- **Amaro e Silva, R.**, 2018. Predicción espacio temporal de la radiación solar a corto plazo. Oral presentation at the *Jornada: Predicción del recurso solar: retos y oportunidades para el futuro* from the PROSOL Project, Madrid, Spain;
- Brito, M.C., **Amaro e Silva, R.**, 2019. A sinusoidal model to assess PV generation from daily irradiation data. Journal of Renewable and Sustainable Energy 11, 053502. doi:10.1063/1.5115354;
- Casaleiro, A., **Amaro e Silva, R.**, Serra, J.M., 2019. V2G potential for grid-services provision and the relevance of a technical characterization. Oral presentation at the 3rd E-Mobility Power System Integration Symposium, Dublin, Ireland;
- Goméz-Jordana, L.I., Milho, J., Ric, A., **Silva, R.**, Passos, P., 2019. Landscapes of passing opportunities in Football – where they are and for how long are available? Poster presentation at the Barça Sports Analytics Summit, Barcelona, Spain.

2. Spatio-temporal solar forecasting

2.1. Solar forecasting overview

Solar forecasting is a maturing research field with several review and benchmark works already published (a compilation is listed in Table 2.1).

In a recent literature review work, Yang et al. (2018) analyzed the top 1000 solar forecasting publications on Google Scholar using a text mining approach. The authors critically analyze the various forecasting methods, highlighting recent advances in the field. Moreover, the most published researchers and journals with most solar forecasting publications are identified. From a different perspective, Sweeney et al. (2019) discuss the future of renewable energy forecasting, covering topics such as new forecasting products; possible changes in forecasting business models and new requirements and challenges for scenarios with high penetration of renewable energy.

Most literature on solar forecasting is focused in the forecasting algorithm alone. However, there are works that describe complete operational forecasting systems, such as the Sun4Cast project (Haupt et al., 2018); or explore complementary elements of such a system, as its database infrastructure (Pedro et al., 2018b).

On a different note, Yang (2019a) discusses the existence of a “solar forecasting bubble”, with a considerable number of works published but with a large variance in publication quality. Thus, the author proposes a guideline to solar forecasting research practice focusing on reproducibility and applicability in an operational context, among other qualities.

Table 2.1 – List of published review and benchmark scientific papers on solar forecasting

Type of article	Reference	Focus	Forecast
Review	(Espinar et al., 2010)	General overview	Irradiance
	(Diagne et al., 2013)	General overview	
	(Inman et al., 2013)	General overview	Irradiance & PV
	(Chaturvedi and Isha, 2016)	General overview	PV
	(Raza et al., 2016)	General overview	PV
	(Sobri et al., 2018)	General overview	PV
	(Yadav and Chandel, 2014)	Artificial Neural Networks	Irradiance
	(Antonanzas et al., 2016)	General overview	PV
	(Voyant et al., 2017)	Machine learning	Irradiance
	(Yang et al., 2018)	General overview	Irradiance & PV
	(Sweeney et al., 2019)	General overview	Solar and wind
	(de Freitas Viscondi and Alves-Souza, 2019)	Big data	PV
Benchmark	(Pedro and Coimbra, 2012)	Forecasting without exogenous variables	PV
	(Bartholomy et al., 2014)	State of the art commercial approaches	Irradiance & PV
	(Zamo et al., 2014a)	Deterministic forecasting	PV
	(Zamo et al., 2014b)	Probabilistic forecasting	PV
	(Lauret et al., 2015)	Machine learning	Irradiance
	(Dobbs et al., 2017)	Machine learning	Irradiance
	(Pedro et al., 2018a)	Machine learning	Irradiance
	(Yang, 2019b)	Probabilistic	Irradiance

2.2. Concept

Historically, the "spatio-temporal solar forecasting" term has been used to describe approaches that explore spatially distributed solar time series. This expression has been used in works using ground-measured irradiance (Boland, 2015), satellite irradiance estimates (André et al., 2019; Dambreville et al., 2014) or PV power data (Agoua et al., 2018; Bessa et al., 2015a).

A recent review in Yang et al. (2018) addresses spatio-temporal forecasting as

> *"A particularly promising application […] using data from a monitoring network of appropriate size. […] As clouds propagate over the monitoring network, data collected by the neighboring sensors can be used as predictors for the forecast location."*

However, throughout that same document, the authors also use the expressions "advection approaches using ground sensors" and "sensor network-based". Moreover, in Yang (2019a) the authors distinguish between NWP, image-based and spatio-temporal methods, which seems to indicate that spatio-temporal forecasting is exclusive to when ground data is used.

Previous review works are also unclear on the definition and classification of approaches exploring spatially distributed solar time series. Although the literature on this topic was scarce in 2013, Inman et al. (2013) classified as "local sensing" approaches those which were based on sky imagers, wireless sensor networks or pyranometer arrays. Later, Antonanzas et al. (2016) distinguish approaches based on their inputs and only mention "neighboring PV plants" in terms of spatially distributed ground data.

Other works use either less explicit designations, such as "lagged exogenous variables" (Zagouras et al., 2015), "distributed information" (Bessa et al., 2015b) and "peer-to-peer forecasting" (Elsinga and van Sark, 2017); or more descriptive titles with "using measurements from a network of irradiance sensors" (Lonij et al., 2013) or "based on an irradiance network, cloud motion" (Lorenzo et al., 2015)..

The purpose of this chapter is to highlight that, in fact, most forecasting approaches share a spatio-temporal nature, and how a clear definition for this new trend of research exploring spatially distributed solar time series is lacking.

Throughout this chapter, different forecasting approaches are presented and discussed from a spatio-temporal perspective. It is also important to note that this classification often differs from author to author.

2.2.1. Numerical Weather Prediction

NWP models represent the physical state and dynamic behavior of the atmosphere by running numerical approximations to the atmospheric equations of motion and parameterizing unresolved processes.

An NWP model has both a temporal and a spatial component, as the dynamic equations on which it is sustained are resolved both in time and space. Thus, it is standard to first define a physical domain, discretized in a three-dimensional grid that extends vertically from the Earth's surface, for which the model is numerically solved (Inman et al., 2013), as illustrated in Figure 2.1.

Figure 2.1 – Basis of all NWP models. First, a domain is defined. Secondly, the domain is spatially discretized at the desired resolution. Finally, the NWP predicts desired information by solving equations of motion and thermodynamics (Inman et al., 2013).

Before any NWP model is ran, the atmosphere's initial state is characterized through the assimilation of data from various sources: satellites, ground stations, regular atmospheric soundings, radars, among others. Modelling the atmosphere requires knowledge of the boundary conditions. Thus, a global model for the whole globe is the first stage of any NWP model. This global coverage is done at the expense of the models' temporal and spatial resolution (3 h and 9 to 20 km, respectively) (Sweeney et al., 2019). An NWP forecast is regularly refreshed (usually every 6 or 12 h), with an up to date initial state which considers more recent data. An example of a global model is the Global Forecast System (GFS) produced by the National Centers for Environmental Prediction (NCEP).

To produce more detailed forecasts, i.e. mesoscale NWP models, smaller domains are defined and the outputs from global models are used as boundary conditions. An example of a

mesoscale model is the Weather Research and Forecasting (WRF) community model, first developed by the National Center for Atmospheric Research (NCAR). Mesoscale models were commonly characterized by temporal and spatial resolutions of 1 h and below 20 km (Lorenz and Heinemann, 2012). This increased detail can lead to a lower forecasting accuracy, since smaller-scale processes are more challenging to model. However, it has been argued that, despite this, the model outputs represent better solar variability (Lorenz et al., 2016).

NWP forecasts can also be bias-corrected using model output statistics (MOS) (Glahn and Lowry, 1972), such as in Mathiesen and Kleissl (2011), or post-processed with more advanced techniques, e.g. artificial neural networks (Lauret et al., 2014) or analog ensemble (Alessandrini et al., 2015).

The caveat of NWP models was their typically coarse discretization, mainly due to computational and data constraints. This made NWP unable to model smaller-scale processes, such as cloud dynamics, having to resort to parametrization schemes which model the larger-scale impact of such processes in a coarser grid space (Larson, 2013). However, this is no longer true for many current mesoscale models. Not only there has been a general improvement in terms of NWP spatial resolution, but there are now models with rapid update cycles, also known as rapid refresh, where recent data is assimilated more frequently. This leads to considerable accuracy gains for shorter-term horizons; the trade-off is a shorter maximum forecast horizon. An example of a rapid refresh model is the High-Resolution Rapid Refresh (HRRR) model (Benjamin et al., 2016) from the National Oceanic and Atmospheric Administration (NOAA), ingesting new 15-min radar data every hour in order to produce forecasts with 3-km spatial resolution and a maximum horizon of 36 h.

With solar forecasting becoming ever more relevant, there have been advances in speficically tuning mesoscale models for irradiance forecasting (Mathiesen et al., 2013b).WRF-Solar is one of such models (Jimenez et al., 2016; Jiménez et al., 2016), with considerable accuracy improvements derived from, among other factors, leveraging the outputs from the HRRR model (Haupt et al., 2018; Lee et al., 2017). Nevertheless, NWP models deliver spatial and temporal resolutions that are not adequate for very short-term solar forecasting.

2.2.2. Image-based methods

2.2.2.1. Satellite imagery

Satellite imagery has been used for solar irradiance estimation for decades (Tarpley, 1979). Cloud transmissivity can be inferred by assuming it to be inversely related to the reflected sunlight, which is detected by the satellites' visible channels (Hammer et al., 2003).

Thus, irradiance forecasting based on satellite imagery (Hammer et al., 1999) consists of two steps: i) shifting a transmissivity map in space according to a cloud motion vector (CMV), which characterizes cloud advection (e.g. in Figure 2.2) ; ii) converting transmissivity into irradiance using a clear-sky model. CMVs can be calculated using simple cross-correlations (Leese et al., 1971) or more complex methods (Cros et al., 2014). Other works use satellite images as inputs to machine learning models (Jang et al., 2016; Linares-Rodriguez et al., 2013). In (Hammer et al., 2015) and (Cros et al., 2017) the use of nighttime satellite data is proposed to enable short-term forecasting before sunrise.

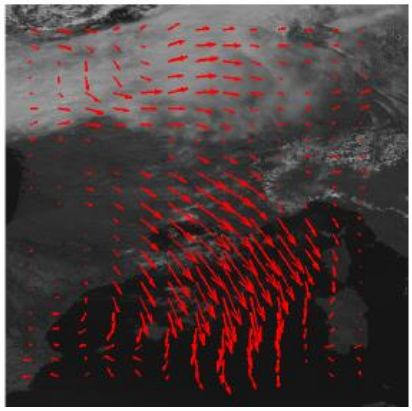

Figure 2.2 – Image from the Meteosat-9 satellite and corresponding CMVs. Adapted from Cros et al. (2014).

The spatial and temporal resolution of satellite-derived products has been improving with the recent launching of newer satellites such as the GOES-R series and Himawari-8. Currently, 0.5 to 1 km and 5 to 10 min resolutions are available, depending on the considered satellite (Engerer et al., 2018; Miller et al., 2018). Parodi et al. (2018) also give an overview of the STEAM (SaTellite Earth observation for Atmospheric Modelling) project which will explore the integration of data from the Copernicus Sentinels 1, 2 and 3 satellites into high-resolution NWP models.

The spatial and temporal resolution of satellite-derived products has been improving with the recent launching of newer satellites, such as the GOES-R series and Himawari-8. Currently, 0.5 to 1 km and 5 to 10 min resolutions are available, depending on the considered satellite

(Engerer et al., 2018; Miller et al., 2018). Parodi et al. (2018) gives an overview of the STEAM (SaTellite Earth observation for Atmospheric Modelling) project which will explore the integration of data from the Copernicus Sentinels 1, 2 and 3 satellites into high-resolution NWP models.

However, it is important to consider some operational constraints: the spatial resolution of satellite images change with the distance from nadir (Figure 2.3 shows that, depending on the location, the Meteosat satellite resolution can degrade up to a factor of 4); or the satellite product timeliness, i.e. the latency between data acquisition on-board the satellite and it being processed into a product for the end-user, e.g. the MSG Downward Surface Shortwave Flux product provided by LSA-SAF has a 1-h timeliness (LSA-SAF, 2009). These constraints limit its use for modeling of individual clouds and short-term operational forecasting.

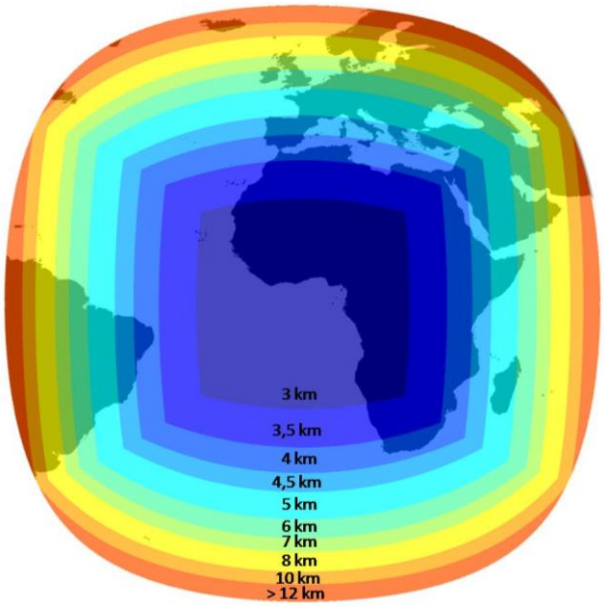

Figure 2.3 – Meteosat spatial resolution (MINES ParisTech, 2009).

2.2.2.2. Sky camera imagery

Consisting of a digital color camera coupled with a fisheye lens, sky cameras capture high-resolution whole-sky images in three different color channels: red, green and blue. Inspired by the satellite experience, solar forecasting based on sky-camera imagery consists in four main steps: i) identification of cloud pixels; ii) estimation of a CMV; iii) forecast the sky condition, moving the latest image according to that same CMV (i.e. it is common to assume its persistence for short time periods); iv) estimate the overall irradiance attenuation at a given target location. Figure 2.4 illustrates the first two steps:

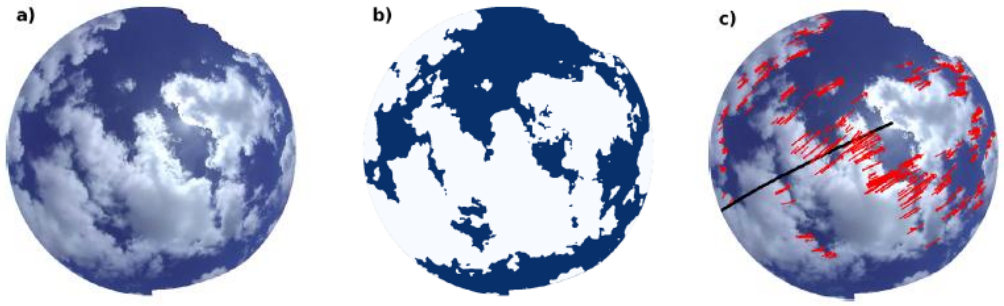

Figure 2.4 – Raw sky camera image (left); processed image for cloud identification (center); and inferred CMV (right). Adapted from Schmidt et al. (2017).

Cloud identification is based on the effect of Rayleigh and Mie scattering and its different impact on the sky camera RGB color channels (Ghonima et al., 2012). It is common to correct the images using a clear-sky library (CSL) in order to remove the effect of non-cloud related factors: the position of the sun (Yang et al., 2016) and the sky turbidity due to aerosol concentration (Kuhn et al., 2017). Common metrics are the red-blue ratio (RBR) and red-blue deviation (RBD), although more complex features can be extracted (Heinle et al., 2010) and, based on predefined thresholds, an algorithm decides if a pixel belongs to a cloud or not (Li et al., 2011). This can then be extended for cloud type (Heinle et al., 2010) and cloud thickness (Ghonima et al., 2012) classification.

The CMV can be inferred using cross-correlation values, inspired by the satellite imagery processing, or more complex criteria such as the variational optical flow proposed in Chow et al. (2015).

Several approaches have been proposed to forecast irradiance from sky camera imagery. In Chow et al. (2011), after assessing the type of clouds that formed during a defined time period and the corresponding local irradiance records, it was assumed that if a cloud pixel intersects the vector connecting the target site and the sun, the irradiance is equal to 40 % of its clear-sky value. In Fu and Cheng (2013), a set of statistical features extracted from the images was used as input for a regression model. Pedro and Coimbra (2015) proposed a nearest-neighbor approach that combines irradiance measurements and image features. Kurtz and Kleissl (2017) suggest that, by deriving global horizontal irradiance (GHI) values directly from the camera, cloud optical depth can be estimated and the advected image can then be converted to a GHI forecast.

According to West et al. (2014), sky cameras are ideal for forecasting microscale weather (up to 1 km spatial scale) with a maximum forecast horizon of 20 min and forecast frequency in

the order of a few tens of seconds. Additionally, sky cameras are also being considered for nighttime operation (Shields et al., 2013) enabling forecasting before sunrise.

Kuhn et al. (2017) quantify the spatial coverage of a sky camera as 10 and 60 km for a cloud height of 2 and 10 km, respectively. On the other hand, Bernecker et al. (2014) and Kuhn et al. (2017) proposed forecasting models based on sky camera imagery which could not surpass persistence for horizons below 1 and 10 min, respectively.

Sky cameras, however, present several limitations: a lower accuracy in estimating the solar resource (Kurtz and Kleissl, 2017); the difficulty in accurately detecting optically thin clouds (Schmidt et al., 2017); the sensitivity to the environment, as soiling, dew or raindrops distort the acquired image (illustrated in Figure 2.5); assessing images from a single camera leads to perspective issues (Kurtz et al., 2018) and makes it challenging to model overlapping cloud layers (Schmidt et al., 2017), moreover, when they are moving in different directions.

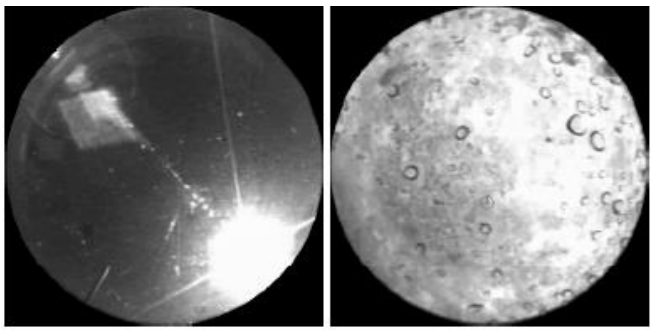

Figure 2.5 – Example of spectral reflectance (left) and camera occlusion from dew or rain (right) on a sky camera. Adapted from Siddiqui et al. (2019).

Recent research proposes the use of multiple distributed sky cameras to leverage 3D cloud modeling (Nouri et al., 2019; Peng et al., 2015) in order to circumvent the limitations of perspective and overlapping clouds at the expense of more hardware and computational resources. However, the authors in Yang et al. (2018) argue that, in the long term, these limitations combined with the appearance of more detailed satellite data, will make sky cameras obsolete for most solar forecasting applications.

It should be noted that defining a CMV from imagery data, either from a satellite of a sky camera is, in its essence, a spatio-temporal method. It relies on a temporal sequence of images which contain spatially resolved information.

2.2.3. Autoregressive time series methods

Autoregressive time series forecasting methods are the only ones lacking a spatial component, as only in-situ past measurements are considered. However, they are good benchmarks to compare spatio-temporal approaches with and quantify the added value from exploring spatially distributed data. Furthermore, they are also useful when only the target variable data is available.

2.2.3.1. Persistence and solar data detrending

Persistence approaches assume that the target variable's latest observation represents future occurrences. For example, for GHI forecasting:

$$\widehat{GHI}(t+h) = GHI(t) \quad [W.m^{-2}] \tag{2.1}$$

where $\widehat{GHI}(t+h)$ is the GHI forecast for an horizon h and is equal to the GHI measurement at time t.

This approach is particularly effective when forecasting in the seconds to a few minutes' time scale (Miller et al., 2013), particularly in stable weather conditions, either clear-sky or cloudy. Although persistence is sensitive to cloud-driven irradiance ramps, missing both up and down ramps, these events are infrequent; for example, Tomson and Tamm (2006) reported that in Tõravere, Estonia, irradiance ramps larger than 50 W.m^{-2}, considering a 1-min time scale, represented less than 10 % of the records from a four-year long data set.

However, since clear-sky irradiance features intrinsic daily, annual and even multi-annual seasonality, the accuracy of the basic persistence rapidly degrades as the forecast horizon increases. It is thus common to detrend solar time series, isolating weather-induced variability from the predictable seasonality. The two most frequent options are either calculating the clearness index (K_t), which quantifies the attenuation in irradiance caused by the whole atmosphere:

$$K_t = \frac{GHI}{TOA_h} \tag{2.2}$$

where TOA_h is the irradiance at the top of the atmosphere for a horizontal surface.

A second alternative is the clear-sky index (K_c), which only considers the attenuation caused by clouds:

$$K_c = \frac{GHI}{GHI_c} \tag{2.3}$$

where GHI_c is the irradiance in clear-sky conditions for a horizontal surface.

While the former approach is based on astronomical and geometrical considerations (Blanc and Wald, 2012), the second implies some physical (Ineichen and Perez, 2002; Lefèvre et al., 2013) or statistical (Baig et al., 1991; Lonij et al., 2012)[3] modeling.

Detrended (or smart) persistence is then converted back to the irradiance or PV power while considering the future position of the sun. For example, for the GHI clearness index persistence:

$$\widehat{K_t}(t+h) = K_t(t) \tag{2.4}$$

$$\widehat{GHI}(t+h) = K_t(t) \times TOA_h(t+h) \quad [W.m^{-2}] \tag{2.5}$$

Where $\widehat{K_t}(t+h)$ is the K_t forecast for an horizon h.

It should be noted that although these equations refer to GHI persistence, the approach is equally applicable to forecasting irradiance or PV power for any tilt/azimuth. GHI measurements only need to be replaced with PV data, and the clear-sky model adapted to PV, either by converting the clear-sky irradiance using an electrical model (Engerer and Mills, 2014) or directly estimating from PV data (Bacher et al., 2009; Lonij et al., 2012).

Due to its simple implementation and light data needs, it is standard to benchmark solar forecasting models against persistence. Recent works have proposed more refined versions of persistence, such as stochastic persistence (Voyant and Notton, 2018) and physics-based smart persistence (Kumler et al., 2019). The former considers averaged past GHI and clear-sky values, decreasing the forecast variance and, therefore, its overall error. The latter decomposes the forecasting of GHI into the computation of top of the atmosphere irradiance and solar zenith angle, and the forecasting of cloud albedo and cloud fraction (these last two variables are inferred directly from the GHI).

[3] Although the Lonij model is proposed for PV, it is extendable to solar irradiance.

2.2.3.2. Linear and non-linear methods

Other works expand on this by using past data as inputs for statistical autoregressive forecasting models. Some linear regression variations exist: the autoregressive (AR) model, which only takes into account past values from the forecasted variable; the autoregressive moving average (ARMA) model, which accounts for past forecast deviations (Ferrari et al., 2013); the autoregressive integrated moving average (ARIMA) model, which removes any trend from the target variable through differentiation (Masa-Bote et al., 2014); or the seasonal autoregressive integrated moving average (SARIMA), which can model seasonal trends present in the data (Kushwaha and Pindoriya, 2018).

There are also non-linear algorithms such as the k-Nearest-Neighbors regression (Wolff et al., 2016b) or machine learning approaches, for example artificial neural networks (Mellit et al., 2014), random forests (Benali et al., 2019) or support vector regression (Wolff et al., 2016a). Solar forecasting literature has been increasingly addressing these models (c.f. Table 2.1), due to their ability to model complex patterns and absence of user-defined assumptions regarding the relationship between variables. In Pedro and Coimbra (2012) autoregressive PV power forecasting models are benchmarked, with the non-linear approaches faring considerably better.

In Reikard et al. (2017), ARIMA models were shown to benefit from having time-varying parameters (i.e. the model coefficients are frequently updated, training in each iteration with the most recent data), surpassing fixed coefficient and NWP methods for horizons below 1 h. However, autoregressive models in general perform best for rather stable weather conditions, and are intrinsically unable to anticipate sudden changes in irradiance. Their main limitation, as noted in Yang et al. (2018), is the absence of information distributed in space, which is essential to accurately model and forecast cloud and weather systems dynamics.

2.2.4. Spatio-temporal time series methods

Some forecasting models complement the autoregressive component with meteorological inputs, be it measurements from an on-site, or nearby, weather station, and/or weather forecasts. Such approach is designated as an autoregressive model using exogenous variables (ARX). Almonacid et al. (2014) forecast PV generation for the next hour based on irradiance and air temperature forecasts. Kemmoku et al. (1999) forecast day-ahead daily insolation by considering the mean atmospheric pressure forecast and previous insolation and temperature measurements. Both works reported gains in forecast accuracy since: i) air temperature

impacts the module temperature and, thus, the PV conversion efficiency; ii) changes in atmospheric pressure can be related to large weather systems, with consequences to the next day overall insolation. However, none of these approaches improve the short-term forecasting of incoming clouds.

On the other hand, high-frequency solar data measured by a ground sensor can accurately identify the presence of a cloud. Works from Bosch et al. (2013) and Bosch and Kleissl (2013) showed that cloud motion vectors can be inferred from spatially distributed solar ground data. Figure 2.6 illustrates the different sensor setups used, with the former exploring a pyranometer array and the latter a set of reference cells from a PV power plant.

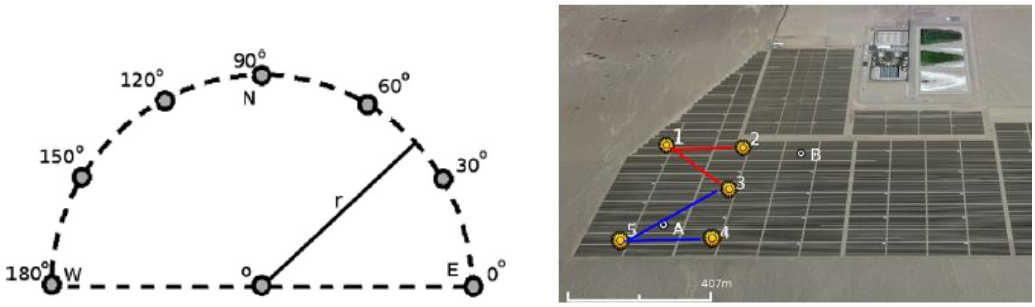

Figure 2.6 – On the left, a pyranometer array from Bosch et al. (2013)*; on the right: a network of reference cells from* Bosch and Kleissl (2013)*. Cloud advection patterns could be derived from each of the data sets.*

Spatio-temporal time series models use solar data from both the target location as well as neighboring sites for their forecasting. This is particularly interesting as an array of sensors can leverage data sets with very high spatial and temporal resolutions. Several data sets have already been reported in the literature (Kuszamaul et al., 2010; Macke et al., 2017; Sengupta and Andreas, 2010). Figure 2.7 illustrates one of these arrays, with 50 pyranometers deployed across a 2.8 km^2 region in Melpitz, Germany, measuring GHI with 1 Hz sampling rate.

Several works have explored spatially distributed solar time series and proposed spatio-temporal approaches for very short- and short-term forecasting. First, the cloudiness condition is mapped by detrending the solar time series (i.e. converting to clearness or clear-sky index). Then, to generate a forecast, the information is either advected (Lorenzo et al., 2015), as illustrated in Figure 2.8, based on cloud-height wind derived from an NWP model or from the ground data itself; or ingested by a statistical model (Yang et al., 2015).

Figure 2.7 – Pyranometer network deployed in Melpitz, Germany, for the HOPE project. Adapted from Macke et al. (2017)

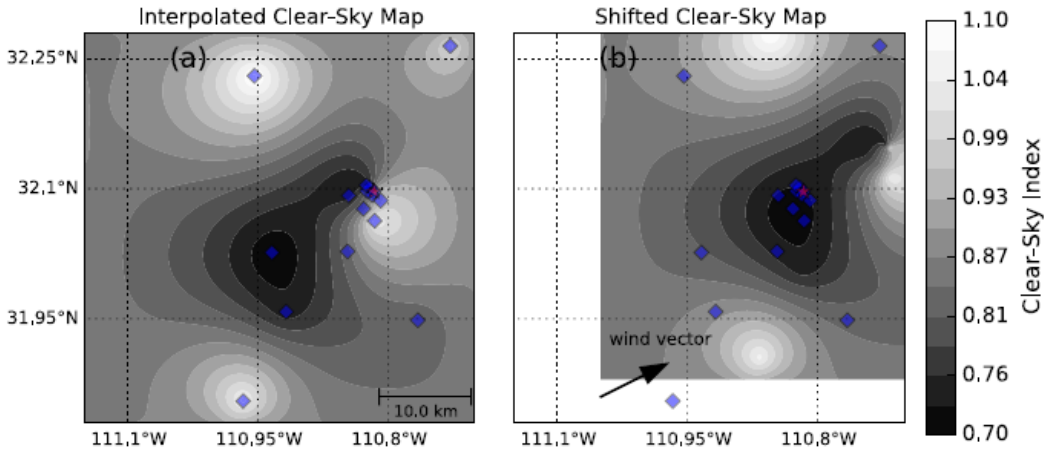

Figure 2.8 – From Lorenzo et al. (2015)*, an interpolated clear-sky index map (left), which is shifted according to a wind vector in order to generate a forecast (right).*

These spatio-temporal methods have also been tested using satellite (Dambreville et al., 2014; Licciardi et al., 2015) and NWP (Andrade and Bessa, 2017; Gagne et al., 2017) irradiance outputs.

Others build models directly from PV data (Elsinga and van Sark, 2017; Lipperheide et al., 2015). Exploring solar power data offers several advantages and poses some challenges. PV measurements have been shown to work as a reliable irradiance proxy (Bertrand et al., 2018; Marion and Smith, 2017) although not as accurate, nor calibrated, as a costly pyranometer. Moreover, the expected mass deployment of these systems will surely result in dense but well scattered spatially distributed data sets, such as from a PV plant with measurements from its

various inverters (Lipperheide et al., 2015), or from an ensemble of small-scale residential systems deployed in dense urban areas (Bright et al., 2019).

In this context, PV forecasting will become a *big data* issue, having to handle large data sets with a considerable number of input variables. To tackle this, Bacher et al. (2009) proposed an online forecasting framework where a model is trained with a small training set, but the coefficients are regularly updated, as new data comes up. According to Bessa et al. (2015a), this solution reduces data storage needs and allows models to capture changing dynamics, such as soiling and obstruction shadowing. Yang (2018) proposed a different approach based on the least absolute shrinkage and selection operator (LASSO) model coupled with an input selection algorithm, reducing the computational demand while making the forecast model more resilient to the amount of data/inputs used. In Licciardi et al. (2015), a non-linear dimensionality reduction approach is proposed using an autoencoder artificial neural network.

Data privacy will also need to be taken into account, as the residential PV systems may be owned by different private parties. Bessa et al. (2018) identify the potential constraints imposed by current legislation on personal data privacy and protection and propose an IT architecture for home energy management systems. Berdugo et al. (2011) propose a forecasting method which takes this privacy issue into account by exchanging a very small amount of information between sites and keeping local measurements private.

Overall, spatio-temporal time series methods show great potential to address solar forecasting for various time scales and can leverage different types of solar data. Therefore, they will be further analyzed in the following chapters.

3. Testing various spatio-temporal scales[4]

3.1. Introduction

As discussed in the previous chapter, it is common to relate a particular data source with a corresponding forecast horizon range, based on its spatial and temporal resolution. Inman et al. (2013) proposed that wireless sensor networks (i.e. approaches which explore spatially distributed ground-measured solar data) are attributed with a 1 m to 1 km spatial resolution and a corresponding 20 s – 3 min horizon range (Figure 3.1).

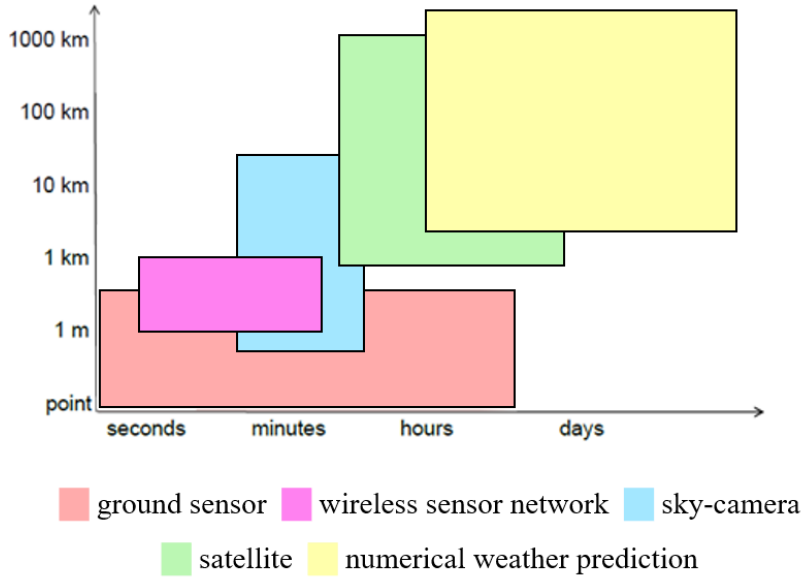

Figure 3.1 – Classification of forecasting approaches based on spatial and temporal resolution. Adapted from (Lorenz, 2014) *and* (Inman et al., 2013).

In 2017, several works reported accuracy improvements from using neighboring solar data, proposing either advective approaches (Lonij et al., 2013; Lorenzo et al., 2015) or linear (Bessa et al., 2015a; Yang et al., 2015) and non-linear (Gutierrez-Corea et al., 2016; Vaz et al., 2016) statistical modelling. Here, both irradiance and PV data sets are shown to provide relevant information for forecast horizons in the seconds to a few hours timescales. However, very few works seek to contextualize the forecasting model dynamics and the obtained results with both the sensor network layout and the local wind patterns, as done in Yang et al. (2015).

Thus, at the time, a lack of a spatio-temporal interpretation of results was identified in the literature. This was even more noticeable for data sets with larger temporal and spatial scales.

[4] Adapted from (Amaro e Silva and Brito, 2017) and (Amaro e Silva and C. Brito, 2018).

Moreover, while different works used different data sets, very few of them tested the exact same forecasting procedure for more than one set.

This chapter aims to test a regression forecasting method for three different data sets, covering different spatial and temporal scales. Locations characterized by prevailing wind patterns were selected to ensure strong spatio-temporal correlations in the data.

3.2. Case studies

3.2.1. Oahu and very short-term forecasting

The *Oahu Solar Measurement Grid*[5] from the National Renewable Energy Laboratory (NREL) (Sengupta and Andreas, 2010) consisted of sixteen LI-COR *LI-200* pyranometers deployed over a 1×1 km^2 region in Oahu, Hawaii (USA) between April 2010 and October 2011. Each measured GHI with a 1 Hz sampling rate and a measurement uncertainty typically less than ±3 % for incidence angles higher than 60° (LI-COR, 2015). Their spatial distribution, characterized by a minimum, median and maximum distance between sites of 86 m, 381 m and 1130 m, respectively, is illustrated in Figure 3.2.

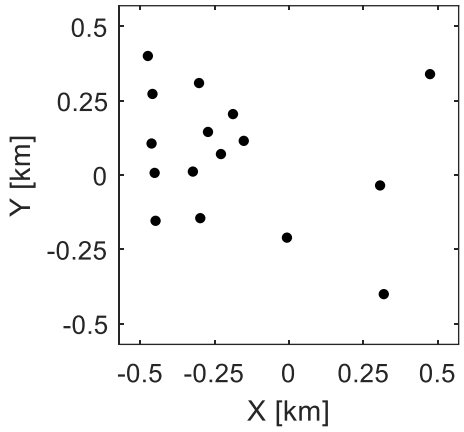

Figure 3.2 –Spatial distribution of the pyranometers from the Oahu Solar Measurement Grid.

Oahu's local climate is characterized by prevailing surface winds from the northeast (Hinkelman, 2013), which influenced the sensors' placement, and broken clouds, mostly small cumulus, resulting in a highly variable solar resource (Lave et al., 2015).

[5] https://midcdmz.nrel.gov/apps/sitehome.pl?site=OAHUGRID

This data set has been extensively used in forecasting and solar variability papers by various authors and in various reference journals (Aryaputera et al., 2015; Hinkelman, 2013; Lave et al., 2015; Yang et al., 2015).

3.2.2. Oklahoma and short-term forecasting

Sensor networks with proper spatial coverage for the time scale of minutes and hours are scarce. Thus, GHI data were obtained from the National Solar Radiation DataBase[6] (NSRDB) (Sengupta et al., 2018), a database modelled using multi-channel measurements from geostationary satellites. This database offers 30-min averaged records with a 4×4 km^2 spatial resolution.

Records from a 17×15 grid covering part of the state of Oklahoma (Figure 3.3), USA, for the 2013-2015 period were considered. This region was chosen because weather systems progress predominantly in a west-east fashion (Hocker and Basara, 2008a, 2008b). The minimum, median and maximum spacing between all pairs of grid cells are, respectively, 3.6, 34 and 85 km.

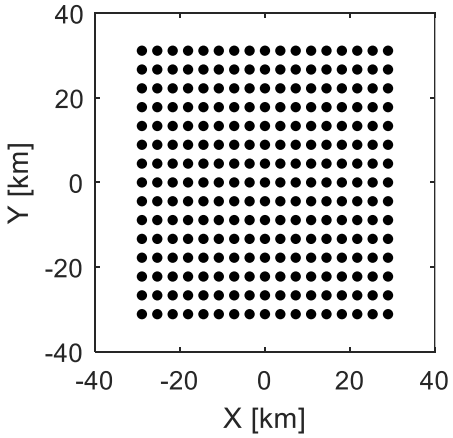

Figure 3.3 – Satellite grid from the NSRDB, covering part of the Oklahoma state (USA) and centered at the (97.54W, 35.41N) coordinates.

3.2.3. Oklahoma and day-ahead forecasting

Across the state of Oklahoma, there is also an operating network of meteorological stations, namely the Oklahoma Mesonet[7] (Figure 3.4). Daily accumulated irradiance records, among other weather variables, from 98 stations covering the 1994-2007 period were made freely available for an international solar forecasting competition organized by the American

[6] https://nsrdb.nrel.gov
[7] http://mesonet.org/

Meteorological Society (McGovern et al., 2015). This data set has already been explored in other works for solar forecasting purposes (Gagne et al., 2017; Martin et al., 2016).

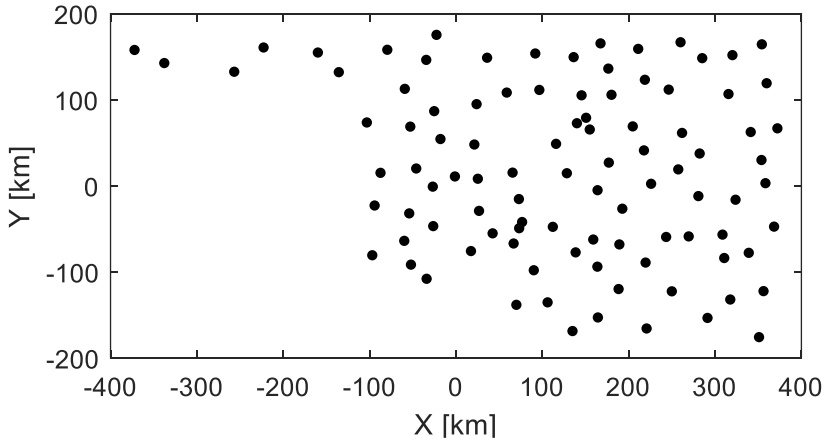

Figure 3.4 - Spatial distribution of the pyranometers from the Oklahoma Mesonet, centered at the (98.76W, 35.41N) coordinates.

The minimum, median and maximum spacing between all pairs of stations is 8, 221 and 796 km. At such temporal and spatial scales, the problem moves from forecasting cloud advection to synoptic-scale weather system dynamics.

3.3. Data pre-processing

As suggested in Yang et al. (2015), the irradiance records from the Oahu data set were averaged to 10-s intervals to remove undesired noise. Then, to detrend the irradiance data sets to clearness index, the irradiance at the top of the atmosphere was calculated for each location using the SG2 algorithm (Blanc and Wald, 2012). This fast and accurate solution simplifies the rather complex formulation for the sun declination using approximations with truncated Fourier series for a restricted time coverage ranging 1980 to 2030. From that, the solar elevation and azimuth angles are calculated using common astronomical equations."

The SG2 model was used for a single point in both the Oahu and NSRDB data sets, considering the mean latitude and longitude of all the sites of each set. Due to its considerably larger spatial coverage, SG2 was ran for each site of the Mesonet data set. Oahu extraterrestrial irradiance was determined using SG2 with 10-s averages; for the two Oklahoma data sets, 1-min values were used and then either averaged or summed, to correspond to the characteristics of the irradiance data.

For the data sets with sub-daily resolutions, only records corresponding to solar elevation values higher than 5° were considered for model training and testing, since lower angles tend to affect measurement quality (King et al., 1997).

3.4. Forecasting model

The forecasting model tested in this and the following two chapters is a linear ARX model, where the inputs are spatially distributed lagged K_t records. It is a multivariate linear regression with two main terms corresponding to the past information from the forecast target itself (i.e. an AR term) and from neighboring sites (i.e. the eXogenous variables), respectively:

$$\widehat{K_{t_k}}(t+h) = \sum_{l} c_{k,l} \times K_{t_k}(t-l) + \sum_{q \neq k} \sum_{l} c_{q,l} \times K_{t_q}(t-l) + b \qquad (3.1)$$

where $\widehat{K_{t_k}}(t+h)$ is the forecast for the k^{th} sensor and horizon h, l is the considered time lag, k and q are indexes pertaining to the target and neighboring sites, respectively, c corresponds to regression-defined corresponding to a given site and lag, and b is the bias term. The ordinary least squares method, which seeks to minimize the sum of the squares of the errors, is used to estimate the regression coefficients.

$\widehat{K_{t_k}}$ is then reconverted to $\widehat{GHI_k}$, by multiplying with the corresponding top of atmosphere irradiance value. Then, the model is assessed using the metrics described in the next section.

Although more complex models may lead to better results (e.g. machine learning approaches), this method was chosen because it allows a physical spatio-temporal analysis and interpretation of the model dynamics.

For the sake of simplicity, for each set, the results shown only consider the first lagged term from all its sites. Further lags and expansion of the number of sites were tested, with a small increase in accuracy but no relevant change to the identified patterns.

The definition of the train and test sets (c.f. table 3.1) aimed to consider complete years for testing, as to assess the model evenly for all seasons, and a train-test distribution as close as possible to 80-20%. Although the first criterion leads to a 37-63 % distribution for the Oahu data set, its high-resolution ensures sufficient data points to properly train the linear ARX model (as showin in section 3.6).

Table 3.1 – Considered training and test periods for each of the three tested data sets.

	Oahu	Oklahoma (NSRDB)	Oklahoma (Mesonet)
Period covered	April 2010 – October 2011	2013-2015	1994-2007
Training set	April 2010 – October 2010 (7 months)	2012-2014 (3 years)	1994-2005 (11 years)
Test set	November 2010 – October 2011 (1 year)	2015 (1 year)	2006-2007 (3 years)

3.5. Assessment metrics

Extensive work on solar forecasting assessment metrics has been reported in Zhang et al. (2015), Voyant et al. (2015) and Vallance et al. (2017). Such metrics either focus on the forecast deviations (e.g. the mean bias error (MBE), mean absolute error (MAE), root mean squared error (RMSE) or the forecast values distribution (e.g. Komolgorov-Smirnov test integral (KSI), skewness or excess kurtosis).

Although there is a clear added value in exploiting the various metrics, since they provide different and complementary information, the results presented throughout this thesis are focused mostly on RMSE and RMSE-based forecast skill (FS), with smart persistence as a reference model, as defined in the following equations:

$$RMSE = \sqrt{\frac{1}{N}\sum\left(\widehat{GHI} - GHI\right)^2} \quad [W.m^{-2}] \tag{3.2}$$

$$FS = 1 - \frac{RMSE_{ARX}}{RMSE_{Persistence}} \times 100 \quad [\%] \tag{3.3}$$

Four main reasons justify this choice: i) according to Yang et al. (2018), RMSE and smart persistence are, respectively, the most common performance criterion and baseline model found in the literature; ii) RMSE gives more relevance to higher magnitude deviations (i.e. due to changes in cloud cover); iii) as persistence excels at forecasting for stable weather conditions, the FS quantifies the accuracy improvement for weather-induced variability; and iv) FS allows comparing results from different time scales and even locations (Marquez and Coimbra, 2012), while RMSE, for example, does not.

3.6. Results

3.6.1. Forecast skill as a function of forecast horizon

For each data set, the most detailed time resolution was considered. Smart persistence and linear ARX forecasting models were built for each site for various horizons (Table 3.2).

Table 3.2 – List of data sets considered in this chapter and their spatio-temporal resolutions. The Oahu data set is also considered in the following two chapters.

Data set	Spatial scale	Time resolution	Forecast horizon range
Oahu	Up to hundreds of meters	10 s	[10 s; 15 min]
Oklahoma (NSRDB)	Up to tens of kilometers	30 min	[30 min; 5 h]
Oklahoma (Mesonet)	Up to hundreds of kilometers	Daily	[1 day; 1 week]

Figure 3.5 presents the corresponding forecast skill distributions from the various sites and for each forecast horizon. For every set, site and horizon, integrating spatially distributed solar data led to better forecasts, with all of the trained models achieving positive skill values.

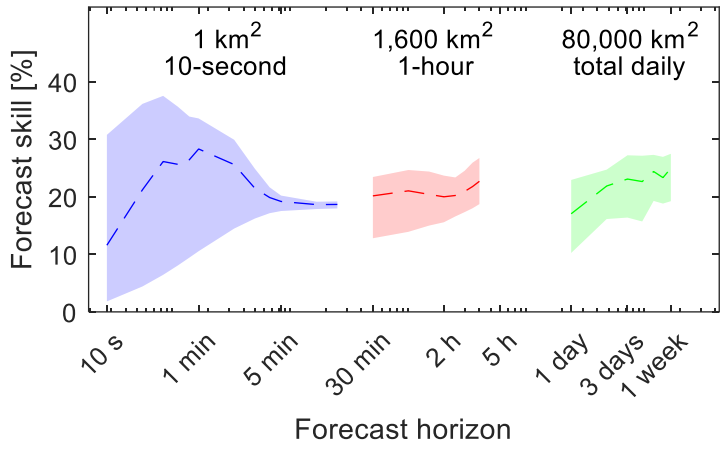

Figure 3.5 - Forecast skill distribution for the various sites of each data set (Oahu in blue, NSRDB in red and Mesonet in green. and for several horizons. The dashed line indicates the median value for each horizon. Above each distribution the spatial coverage and time resolution of each data set are indicated.

The Oahu data set achieves considerably high skill values up to 1 min ahead forecasts (up to 37.6 %) and the maximum value corresponds to a 30 s horizon. It is plausible to infer that that time interval better corresponds to the amount of time a cloud typically takes from covering one of the sites and its downwind neighbor (which is related to the distance between them). However, for these short horizons, there is also a large spread between the worst and best performing sites (up to 31.7 %). For longer horizons, the forecast skill converges around 18.5 %, with a substantially smaller spread.

These positive results qualitatively agree with Yang et al. (2015) and are particularly relevant, since only sky camera solutions tend to be considered for very-short term horizons. Moreover, several works have shown negative results from sky cameras for horizons below several tens of seconds (Bernecker et al., 2014; Schmidt et al., 2016).

The Oklahoma data sets achieve, depending on the forecast horizon, 23-27 % skill values for the best-performing sites with a 7-13 % spread. Their forecast skill profiles appear to be more homogeneous, and the achieved skills seem to increase with the forecast horizon. Similar trends have been reported in the literature (Agoua et al., 2018; Lorenzo et al., 2015). As suggested in Lorenzo et al. (2015), this behavior is attributed to the increase in RMSE with the forecast horizon for the smart persistence, whereas it remains considerably stable for the linear ARX model.

For these coarser resolutions, spatio-temporal approaches could be an interesting complement to more traditional satellite and NWP-driven models.

3.6.2. Spatio-temporal analysis

For all data sets, the mapping of local forecasting skills reveals the correlation with the local wind patterns (Figure 3.6), with higher skill values corresponding to sites with upwind (i.e. northeastern for Oahu and western for Oklahoma) data available.

Exploring how the different neighbors contribute to the linear ARX model and how those contributions change with the forecast horizon provides important information for understanding Figure 3.6. To make reading coefficient maps more intuitive, a relevance metric was calculated as:

$$Relevance_k = \frac{\sum_l |c_{k,l}|}{\sum_l \sum_{q \neq k} |c_{q,l}|} \times 100 \quad [\%] \tag{3.4}$$

This metric quantifies the relative contribution of each site to a given model. A site with 100 % relevance single-handedly drives the regression, whereas 0 % implies that the site does not influence the forecast.

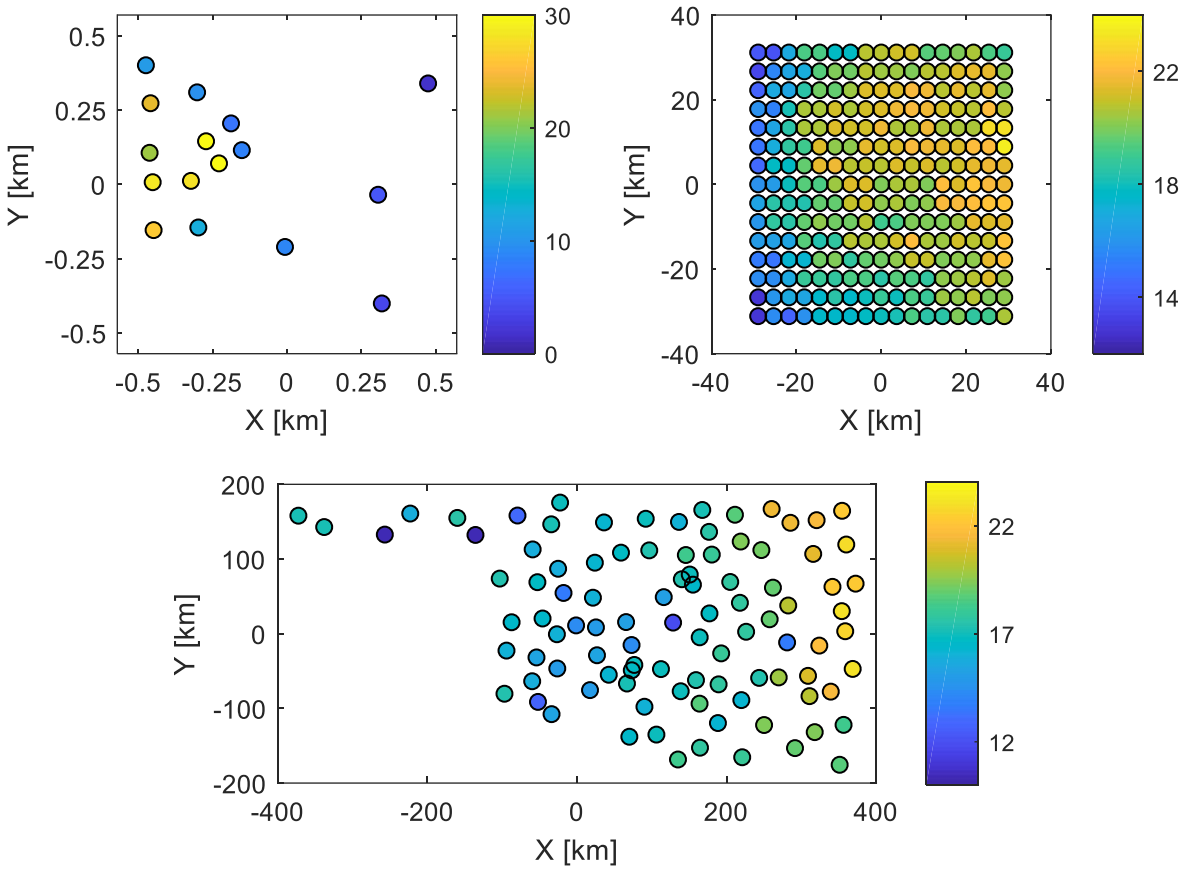

Figure 3.6 – Forecast skill for three different data sets: 10s-ahead for Oahu (top left); 30 min-ahead for Oklahoma NSRDB (top right) and 24 h-ahead for Oklahoma Mesonet (bottom).

The most interesting results were found for the Oahu data set, where cloud dynamics are more present due to the data's high sampling rate. For shorter forecast horizons (up to 1 min ahead), sites achieving high forecast skill values benefitted considerably from nearby upwind (i.e. northeastern) neighbors (Figure 3.7, left); on the other hand, sites with lower skill values rely mostly on their own past (Figure 3.7, right), as they have no adequately placed upwind neighbors.

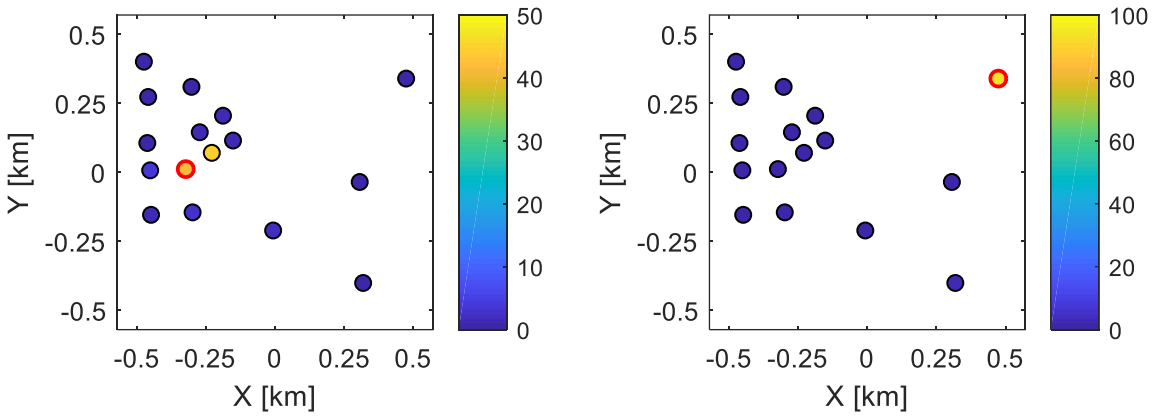

Figure 3.7 – Site relevance (calculated using eq. (3.4) for the ARX model trained with 10-s data. The model was assessed for two different target sites, highlighted in red in each plot, and for 10 s ahead.

As the forecasting horizon increases, the relevance maps evolve quite differently. Sites without upwind neighbors shift from an autoregressive to a spatial averaging framework, benefitting equally from every site (Figure 3.8, first row). This is driven by the K_t autocorrelation decreasing with the forecast horizon (i.e. the farther ahead the forecast is, the less the variable resembles itself), reducing the effectiveness of persistence. For sensors with upwind neighbors, the more distant upwind sites become more relevant (e.g. Figure 3.8, second row for the first two columns). This is to be expected as, for constant wind speed (a reasonable assumption for short time periods), a longer time period implies that a cloud moves a larger distance.

However, when the time scale surpasses the data set's spatial coverage, the model starts lacking relevant information, behaving as for a site with no upwind sensors, i.e. a spatial average approach (Figure 3.8, second row and third column). This justifies the smaller spread in forecast skill for forecasts longer than 5 min ahead found in Figure 3.5, as the model dynamics become less site-sensitive.

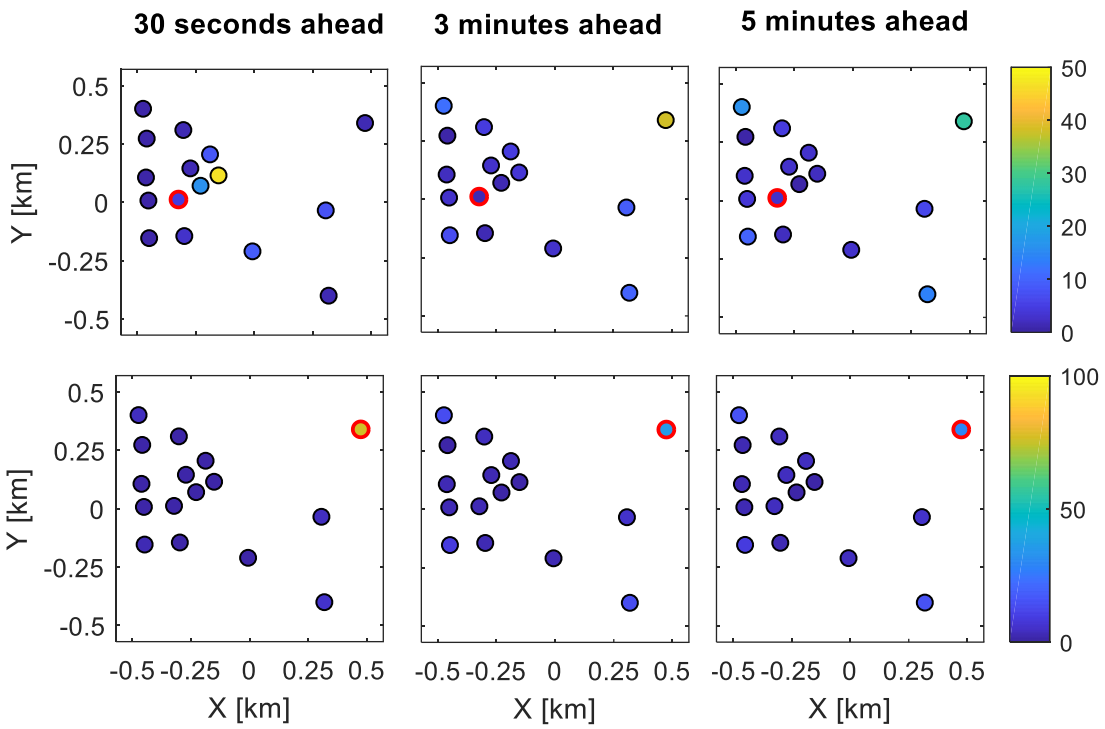

Figure 3.8 – Site relevance (calculated using eq. (3.4) for the ARX model trained with 10-s data. The model was assessed for two different target sites, one per row and highlighted in red, for 30 s (left), 3 min (center) and 5 min (right) ahead.

No clear spatial patterns were found from the regression coefficients for the Oklahoma data sets. Three possible explanations are proposed: i) as the data are averaged in time, the relevant information is more spread out in space (i.e. a longer "trail"); ii) for a coarser temporal and spatial resolution, the scope of the model shifts from forecasting individual clouds to larger weather systems, which should leave larger, and more diluted, "footprints" in the data.

Nonetheless, the forecast skill profiles reflect the impact of increasing the forecast horizon. As an example, Figure 3.9 compares the Oklahoma NSRDB data set for 30 min and 2 h ahead forecasts. It is possible to observe that the higher skill region is considerably thinner for the latter, which can be explained with the decreasing number of sites that have neighbors that are sufficiently distant.

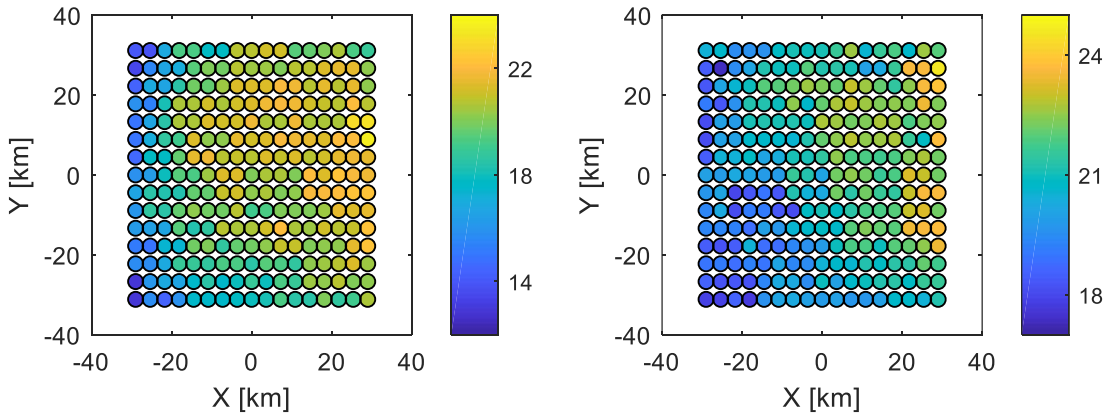

Figure 3.9 – Oklahoma NSRDB data set forecast skill for 30 min and 3 h ahead forecasts The higher skill region is considerably thinner for the latter, plausibly due to the reduced number of sites that have neighbors that are sufficiently distant.

Doing this sort of visual analysis for each data set, target site and forecast horizon, as in Figure 3.8, is not practical. Thus, to assess spatio-temporal dynamics in a more systematic manner, the distance between the forecast target and its relevant inputs could be quantified, namely, through a weighted average distance (WAD):

$$WAD_k = \frac{\sum_{q \neq k} \sum_l \left|c_{q,l}\right| \times dist_{k,q}}{\sum_{q \neq k} \sum_l \left|c_{q,l}\right| + \sum_l \left|c_{k,l}\right|} \quad [m] \tag{3.5}$$

where $\left|c_{q,l}\right|$ and $\left|c_{k,l}\right|$ are the absolute value of the regression-defined coefficients for the l^{th} lagged term for the neighboring site q and target site k, and $dist_{k,q}$ is the distance between the two sites.

The WAD metric aims to identify the spatial scale from which a forecasting model benefits the most. However, to correctly compare different target sites, which are differently distanced from their respective neighbors, and different data sets, the WAD should be normalized by the corresponding simple average neighbor distance:

$$nWAD_k = \frac{WAD_k}{\frac{1}{N}\sum_{q \neq k} dist_{k,q}} \tag{3.6}$$

The results for the three data sets are plotted in Figure 3.10. For the Oahu set, the various sites were aggregated in two subsets: the most peripheral sensors (the four most eastern sites, lacking upwind neighbors) and the remaining sites. This is justified by the two different spatio-temporal dynamics previously identified. For the first subset, shown in light blue, the nWAD consistently increases with the forecast horizon up to a value around 1. This is a consequence of the model dynamics' shifting from autoregressive to a spatial-average (i.e.

when all sites have the same relevance, the WAD equals the average neighbor distance, corresponding to an nWAD equal to 1). On the other hand, for the second subset, in dark blue, the nWAD increases very steeply with the forecast horizon up to 1 min ahead, to a value above 1, decreasing afterwards, to values close to 1. This indicates that up until the inversion point, the farther the horizon the farther the relevant neighbors. Beyond that, the ARX model gives the same relevance to every site (i.e. a spatial average).

For the two Oklahoma data sets, the nWAD is rather homogeneous and around 1, indicating the "smoothing" forecast approach. However, Figure 3.6 shows that for both data sets the linear ARX model is indeed capturing spatial dynamics. Thus, it is possible that the proposed metric is only adequate for a certain range of spatial and/or temporal scales. For example, weather systems at a synoptic scale can have horizontal length scales of the order of 1000 km.

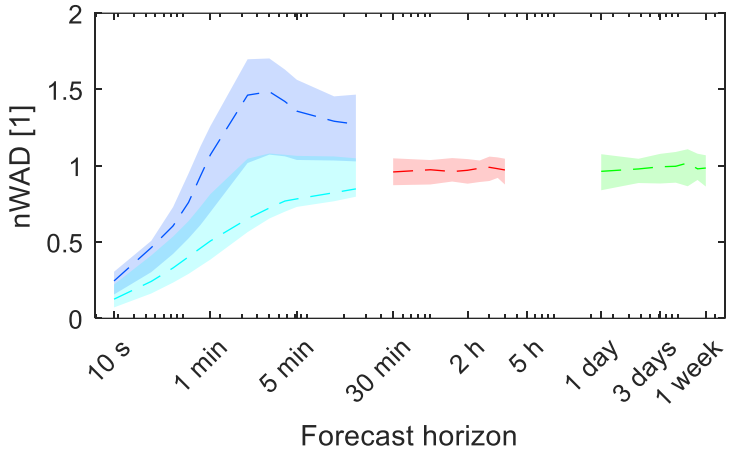

Figure 3.10 – The nWAD metric as a function of the forecast horizon for the Oahu data set (in light blue the peripheral upwind sites and in dark blue the remaining sites), the Oklahoma NSRDB set (in red) and the Oklahoma Mesonet set (in green).

3.7. Final Remarks

This chapter has shown that forecast approaches exploring spatially distributed solar time series can identify local weather patterns for a wide range of spatial and temporal scales (from meters to hundreds of kilometers; and from seconds to days).

By mapping the forecast skill values and the coefficients from the linear ARX model and assessing a new metric for the distance between the forecast target and its relevant inputs, it was possible to identify the spatial weather patterns characteristic of the locations under study.

Three different forecast modes were also identified when testing the model:

i) weather-driven, when upwind information is available for a forecast horizon that is of the order of magnitude of the characteristic time, defined by the distance between sensors divided by cloud speed, greatly surpassing persistence;
ii) persistence-driven, when no adequate upwind information is available for very-short-term horizons (i.e. persistence is particularly accurate) and the model assumes an autoregressive dynamic, with low added value;
iii) spatial-averaging or smoothing, when no adequate upwind information is available for farther horizons (i.e. persistence is less accurate) and more conservative forecasts deliver considerable gains in accuracy.

The need for a good match between the layout of the sensor network and the local weather patterns is highlighted, with spatially distributed data revealing advection patterns from both individual clouds as well as larger (i.e. synoptic scale) weather systems. Additionally, the considerably high forecast skill values achieved for the Oahu data set and a forecasting horizon between 10 s and a few minutes should also be emphasized, as such short time scales are commonly considered exclusive for persistence and sky camera-based algorithms.

4. Integrating wind information in statistical spatio-temporal solar forecasting models[8]

4.1. Introduction

Wind information at cloud height is essential for advective models as it defines the spatial shift that is applied to cloudiness when forecasting future cloud cover. However, this type of data is seldom used in spatio-temporal statistical approaches.

Several solar forecasting works have explored wind information to improve forecast accuracy. Some of them include wind speed and/or direction as inputs (Dolara et al., 2015; Ogliari et al., 2013) while others tested more refined approaches. In Mathiesen et al. (2013a), a bias-correction step was applied to GHI probabilistic forecasts after disaggregating the data by geostrophic wind regimes related to coastal clouds formation. Yang et al. (2015) optimize the number of neighboring sites and time lags considered in a spatio-temporal LASSO model for GHI forecasting based on the local prevailing wind speed and direction. In Agoua et al. (2018), regime-specific spatio-temporal LASSO models were trained based on ground-level wind speed clusters, as this variable impacts PV module temperature and, thus, efficiency. Modest improvements should be expected, as none of these approaches enable solar forecast models to adapt to variable cloud advection dynamics. For example, in Agoua et al. (2018) a maximum 3 % nRMSE improvement was reported for 1 h ahead. In Yang et al. (2015) a maximum nMAE improvement of 5.3 % is reported for ordinary least squares regression. However, the fact this approach only surpasses LASSO regression by 0.68 % indicates that most of the reported improvement was actually derived from addressing a dimensionality issue rather than effectively incorporating cloud advection dynamics.

In this chapter, the impacts of a rigid model that explores information from a single direction in terms of forecast performance are discussed and a regime-based approach is proposed to integrate cloud-height wind information in spatio-temporal solar forecasting.

4.2. The Oahu case study

The Oahu data set in Chapter 3, characterized by prevailing easterly winds as shown in Figure 4.1, showed how local advection patterns shape the spatio-temporal correlations in a data set, with impacts in forecast skill and the linear ARX model's coefficients. However, it is to be

[8] Adapted from (Amaro e Silva et al., 2019).

expected that these results do not characterize the less frequent patterns when the wind deviates from its mainstream direction, much less be transferable for locations where the winds are not as homogeneous.

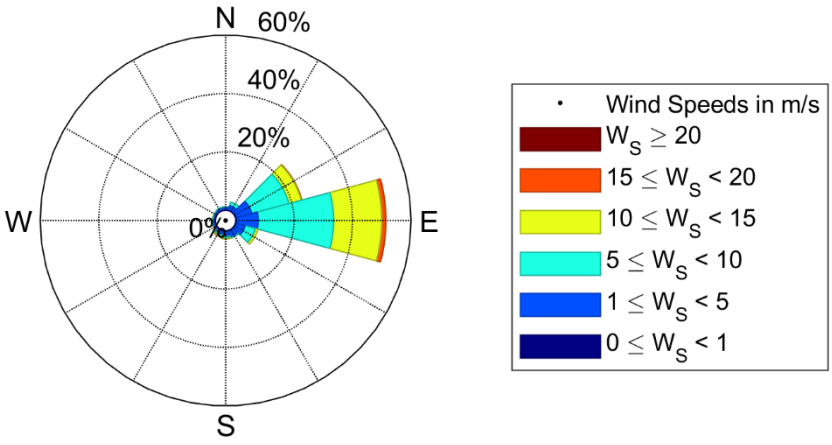

Figure 4.1 – Oahu's wind speed and direction histogram at 850 hPa, with predominance of east and northeast winds faster than 5 m/s.

This can be shown by disaggregating the forecast skill according to the wind direction. Forecasts were clustered in 45-degree bins, centered on the cardinal directions, based on wind data from the ERA5 reanalysis, an open source data set from the European Centre for Medium-Range Weather Forecasts (ECMWF). The 850 hPa pressure level, which usually corresponds to a 1500 m altitude, was chosen because cloud height is unknown and cumulus clouds, the most common in Oahu, tend to form below 2000 m. Despite the coarse temporal and spatial resolutions (1 h and 31 km, respectively), at this pressure level (i.e. above the boundary layer) wind should be substantially more homogenous in time and space due to the reduced effects from surface friction (Stull, 1988).

The method was first tested for the site highlighted in red in Figure 3.7 (left plot), since it has neighboring sensors placed in various directions, making it an interesting target to assess the wind regime-based approach. Figure 4.2 shows the forecast skill values for the different wind directions. Results show that, for a site with adequately placed upwind neighbors, the linear ARX model underperforms for non-prevailing wind directions, being surpassed by the smart persistence. While the model can attain high forecast skills, up to 32.3 %, for prevailing winds, skill values can be as low as -44.7 %. Of course, it is important to note that the forecast skill for each wind regime is as relevant to the overall performance as its frequency of occurrence.

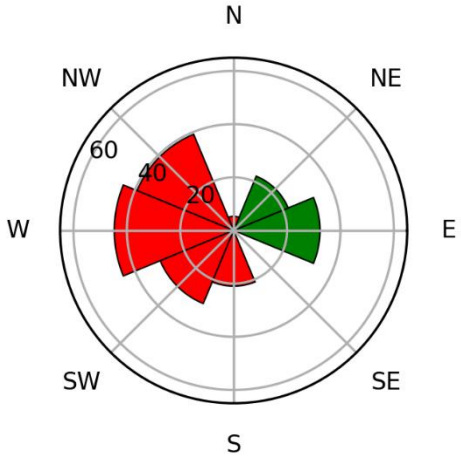

Figure 4.2 - Forecast skill (radial axis) as a function of wind direction for the Oahu site highlighted in red Figure 3.7 (left plot). Positive skills (in green) can be seen for northerly and easterly winds, whereas south and west show considerable negative skill values (in red).

4.3. A proposal for a regime-based forecasting approach

4.3.1. Defining regimes based on wind direction

The training set was divided into eight clusters, based on the 45º wind direction bins, and an individual model was trained for each cluster.

Traditional train-test splitting is not adequate for data from a location with prevailing winds. The scattered and scarce presence of the remaining winds makes it difficult for such configuration to result in properly balanced sets. To circumvent this problem, the data blocks corresponding to each regime were concatenated and then split in 70-30 % proportions (Figure 4.3). Although this setup does not allow calculating annual performance metrics, it should result in a more representative assessment of the proposed regime-based approach.

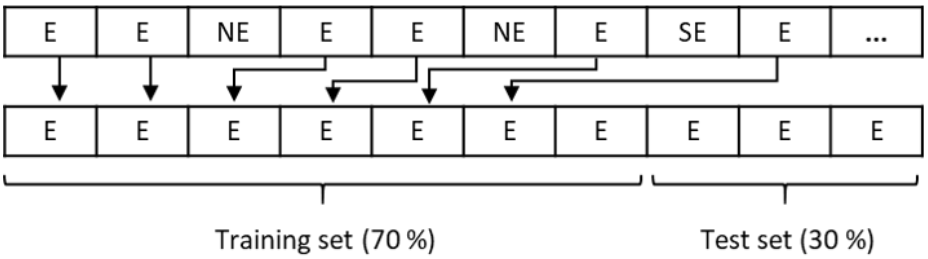

Figure 4.3 – Example of the regime-specific train-test splitting, in this case for easterly winds.

Substantial improvements were achieved for all wind directions, with a slight negative skill for southwesterly winds (Figure 4.4) and an overall forecast skill of 30.4 % (2.4 % improvement over the baseline ARX).

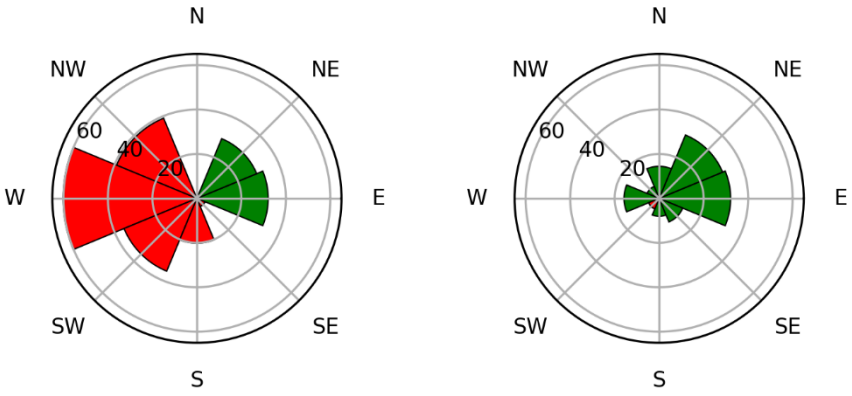

Figure 4.4 - Forecast skill (radial axis) as a function of wind direction for two ARX models: baseline version (left) and one built according to wind direction regimes (right). Positive/negative forecast skill values are shown in green/red. The left plot corresponds to Figure 4.2 but for an alternative train-test configuration.

Nonetheless, it should be highlighted that the model still achieves low skill values for non-prevailing winds. Defining regimes based on wind direction did not lead to the emergence of new spatial patterns and the ARX model behaved similarly to an autoregressive model in most cases (i.e. the regression attributed low coefficients to all neighbors, hence they are of little relevance to the forecast), similar to what is illustrated in Figure 3.7 (right side). This is likely justified by the fact that the spatio-temporal correlations in the data are also affected by wind speed. Thus, a regime framework which does not take this variable into account will not be completely effective in separating different advection patterns.

4.3.2. Defining regimes based on wind direction and speed

The inability to detect new spatial patterns may be related to the broad range of wind speed covered by each wind direction regime weakening the spatio-temporal correlations in the data (i.e. the relevance of a neighboring sensor depends on its distance, the forecasting horizon, and the cloud speed). Hence, further specialized models were trained by partitioning in two each of the wind direction regimes according to a wind speed threshold.

Using a trial and error approach, a 9 m/s threshold led to the best results. The ARX coefficients showed various spatial patterns for speeds above the threshold (Figure 4.5), matching the corresponding wind direction. For most slow wind regimes, the model kept its autoregressive behavior, with the remainder only capturing faint signals.

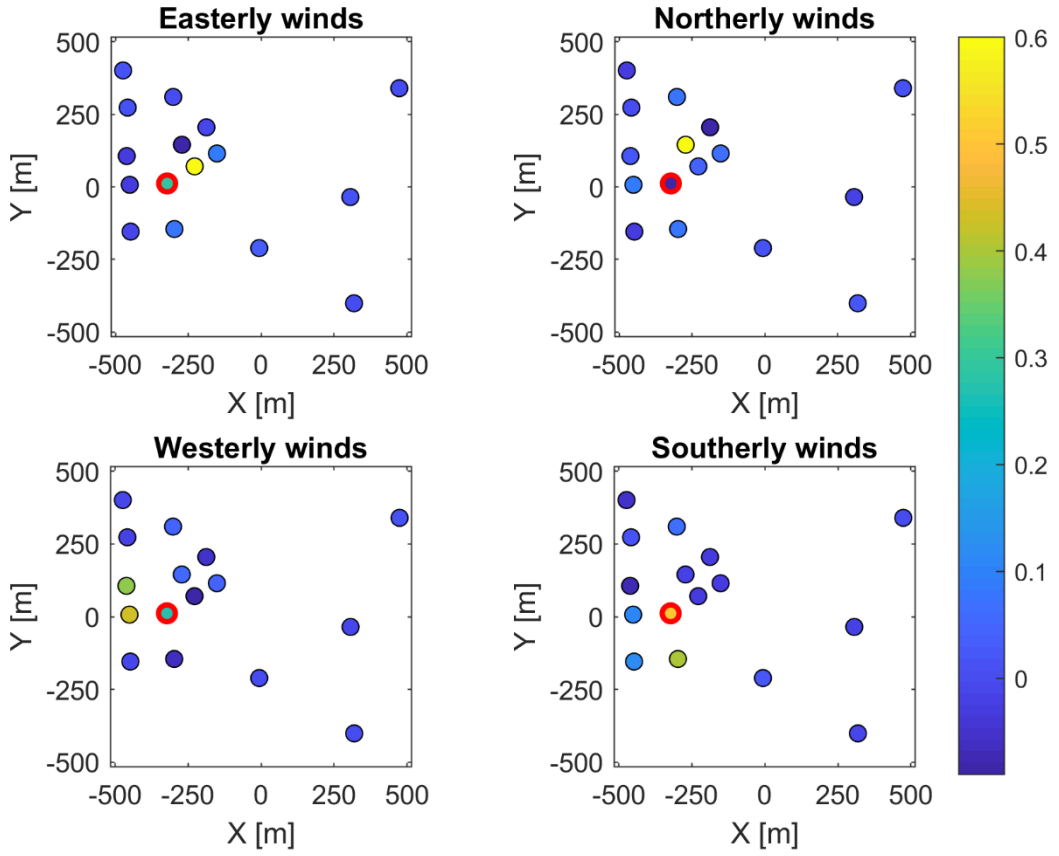

Figure 4.5 - Regression coefficients map for the clustered data with wind speed > 9 m/s. Winds coming from the four cardinal directions are shown, with clear spatial patterns observed. The target sensor is represented by a red circle.

The corresponding forecast skill values are plotted in Figure 4.6. The regimes for faster winds, which detect the different spatial patterns, improved the most (up to 55.2 %, while slow winds up to 27 %). Lower skill values were achieved for southwesterly fast winds (15.7 %), which is explained by the absence of an adequate upwind neighbor. For westerly winds, however, the model greatly underperforms (-164.4 % skill value). Retraining the ARX model with the test set resulted in coefficients which showed no particular spatial pattern. Furthermore, when the model is, instead, retrained using other blocks of the westerly fast winds' subset, sometimes the coefficients correspond to southerly or easterly winds. The complex orography of the island may explain possible limitations of the coarse resolution reanalysis data to fully describe the local wind patterns.

Defining regimes based on both wind direction and speed, except for westerly winds as justified in the previous paragraph, improved the overall forecast skill to 32.3 % (4.3 % higher than the baseline ARX model and 1.9 % higher than the wind-direction regimes framework).

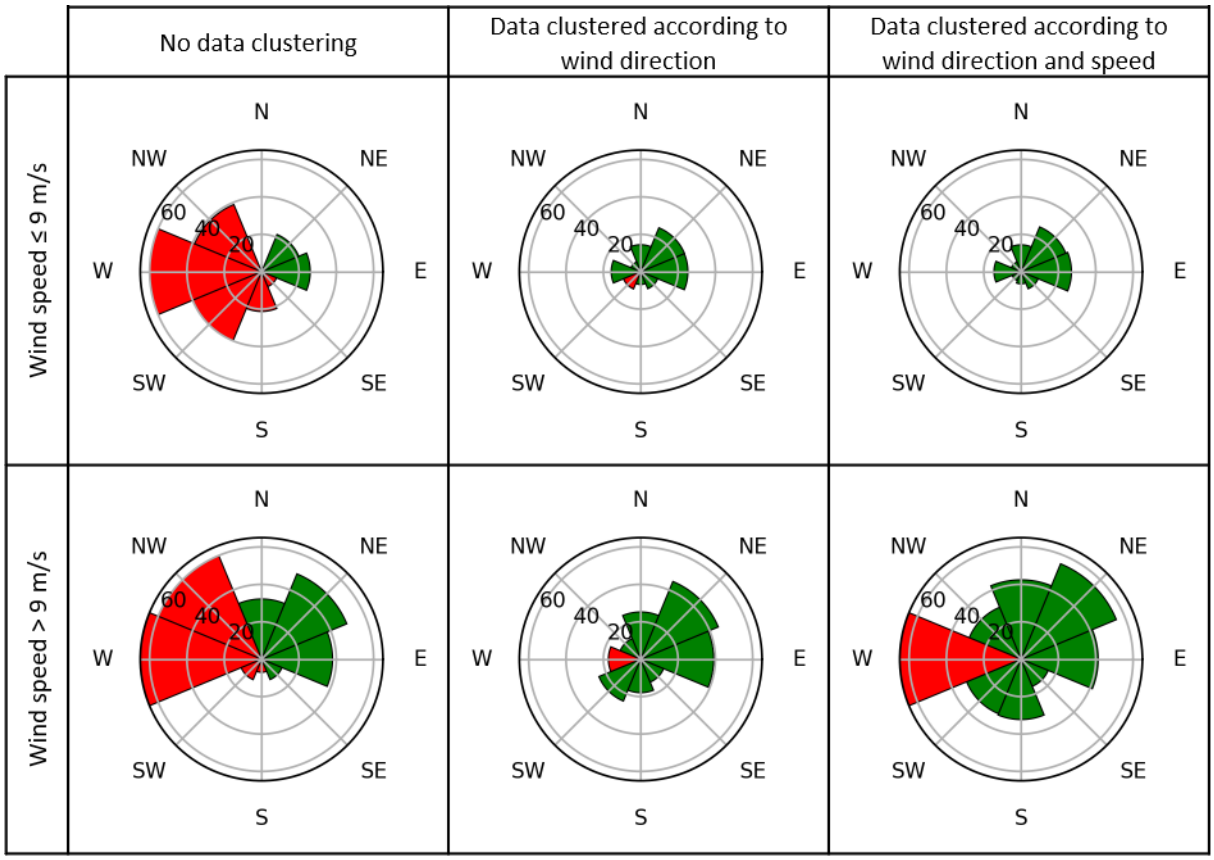

Figure 4.6 – Forecast skill (radial axis) according to wind direction and the implications of the wind regime definition. Negative forecast skill values are shown in red, positive forecast skills shown in green. On the bottom row and for the first and last columns, the negative skill value for westerly winds was saturated for visualization purposes.

4.3.3. Testing the regime-based approach for all sites

The regime-based framework was tested for each individual site in the Oahu data set. When compared to an ARX model without regimes, forecast skill improved between 0.1 % and 6.1 %, with a spatial distribution similar to the one shown in the first plot of Figure 3.6 (i.e. sites with northeastern neighbors improve the most).

The fact that the best performing sites for the baseline ARX also improve the most can be explained by two factors: i) only sites for which the ARX model captures the typical wind patterns do underperform for non-prevailing winds, which the regimes then correct; ii) the prevailing wind direction (i.e. 70 % of the data) also benefits from the regime-based approach.

4.3.4. Impact of wind forecast uncertainty

To implement the proposed approach in an operational context, the reanalysis wind data ought to be replaced by actual forecasts. Forecasts have inherent uncertainty, and their impacts on the regime-based approach forecasting accuracy must be assessed.

ERA5 also provides historical forecasts twice a day, at 6 and 18 UTC, with each run offering hourly values up to 18 h ahead. Because this work focuses on short-term forecasting, the two daily forecast runs were merged so that, at the time of the forecast, the most recent information is considered.

The wind forecasts have a mean absolute deviation below 12° and 0.6 m/s when compared with the reanalysis records. This led to 12.3 % of the data points being allocated to a different regime, mostly in terms of wind direction and towards lower wind speeds.

Using the same setup as in the previous section, the linear ARX model was retrained. Replacing the wind reanalysis by operational forecasts led to a negligible reduction in forecast skill (less than 0.1 %) and showed similar spatial patterns as those in Figure 4.5. This tolerance to forecast uncertainty is driven by the fact that wind variables are not explicit inputs to the model but rather criteria for the regime definition. Therefore, the proposed regime-based approach can be implemented in operational forecasting solutions, such as real-time nowcasting.

4.4. Final remarks

Results from the previous chapter demonstrated that a linear ARX model performs considerably well as a statistical spatio-temporal approach for locations characterized by prevailing winds. However, it was shown here that, despite the positive overall skill values, the model is, in fact, unable to cope with non-prevailing wind patterns.

The proposed regime-based approach proposed allows the model to adapt to changes in local wind patterns by leveraging the integration of wind data. The fact that this approach is compatible with operational wind forecasts shows that it could be implemented in operational forecasting solutions.

The idea that the forecast model performance is highly sensitive to the placement of sensors is reinforced by these results. These results also pave the way for automatic regime selection algorithms for the improvement of solar forecasting, in line with McCandless et al. (2016b)

and McCandless (2016a). As future work, it should be explored how to optimize the wind speed threshold for each individual wind regime and how it relates to the distance between sites. Additionally, even though the benefits of the proposed approach are expected to be generalized for other locations where winds are not as homogeneous, it would be relevant to demonstrate its added value in such contexts.

5. Spatio-temporal approaches sensitivity to modules' tilt and orientation[9]

5.1. Introduction

As discussed in Chapter 1, spatio-temporal solar forecasting is particularly attractive as it may include as input PV power data from neighboring sites, e.g. rooftop PV systems in the same town or different strings on a utility scale solar power plant. Large-scale deployment of PV systems could thus lead to a dense and scattered network providing valuable information for solar forecasting, anticipating the arrival of a cloud or a weather system. Additionally, power data from such systems should better correspond to the spectral, angular and thermal response from PV modules.

However, most results on very-short term spatio-temporal forecasting, including those discussed in this thesis so far, are based on dedicated networks of GHI sensors. This is justified by practical reasons: currently most PV generation loggers are configured to store 15-min averaged data, well above the time scale for which spatio-temporal methods attain the best performance (c.f. Figure 3.5).

Although it might be tempting to directly transfer the added value of these works to PV applications, this would ignore the fact that often (e.g. in urban environments) PV systems are installed with diverse tilts and orientations (Killinger et al., 2018; Leloux et al., 2015). Since the PV output of tilted modules is sensitive to the fraction of beam and diffuse components of irradiance, which depend on the cloud conditions, PV systems with different mountings react differently to a passing cloud. Thus, it seems plausible that the K_t and K_c indexes might not be completely effective in removing the geometry dependency from the raw solar data.

There exist PV forecasting (Elsinga and van Sark, 2017) and performance assessment (Amaro e Silva et al., 2018; Killinger et al., 2016) works showing some concern regarding the influence of plane-of-array (POA) mismatch between target and neighboring PV systems. However, none has sought to evaluate and quantify the consequences of directly using data from a PV system ensemble with varied tilt and orientation angles.

This chapter addresses the impact of using spatially distributed PV data sets with heterogeneous mountings for very-short term spatio-temporal forecasting applications. To circumvent the lack of high-resolution PV data, state-of-the-art decomposition and transposition algorithms are applied to the Oahu data set assuming realistic plane-of-array

[9] Adapted from (Amaro e Silva and Brito, 2019).

distributions. PV reflectivity is also modeled as it is very geometry dependent. This approach is a common practice in resource estimation and PV performance modeling works (Killinger et al., 2016; Marion and Smith, 2017). The modeled data are then ingested into a 10 s-ahead linear ARX model and the forecast skill assessed.

5.2. Simulating a network of differently arranged surfaces

5.2.1. Global irradiance decomposition and transposition

Global irradiance on a given POA (G_{POA}) may be transposed from horizontal irradiance considering its different components:

$$G_{POA} = f_1(BHI) + f_2(DHI) + f_3(GHI) \quad [W.m^{-2}] \tag{5.1}$$

where BHI and DHI stand for beam and diffuse horizontal irradiance, respectively, and f_1, f_2, and f_3 are transposition functions for the beam, diffuse and reflected components.

BHI may be described by simple trigonometric conditions and therefore f_1 is simply:

$$f_1 = BHI \times \frac{\cos\theta}{\cos\varsigma} \quad [W.m^{-2}] \tag{5.2}$$

where θ is the incidence angle and ς the solar zenith angle.

On the other hand, f_2 and f_3 tend to be more complex. The most commonly used approach for DHI transposition (f_2) is the Perez model, due to its simplicity and good performance for different time scales and climates (Perez et al., 1990). It is an anisotropic model which considers the circumsolar and horizon enhancements, based on empirically defined coefficients:

$$f_2 = DHI \times \left[(1 - F_1) \times \frac{1 + \cos\beta}{2} + F_1 \times \frac{a}{b} + F_2 \times \sin\beta\right] \quad [W.m^{-2}] \tag{5.3}$$

where β is the surface tilt, a and b represent the cosine of θ and ς, respectively, and F_1 and F_2 express the degree of circumsolar and horizon anisotropy. It is important to note that F_1 and F_2 depend on parameters empirically derived from the GHI and DHI measurements.

The reflected component was estimated using a 3D isotropic model (Badescu, 2002) and assuming a constant ground albedo (ρ) of 0.2:

$$f_3 = GHI\ \rho\ \frac{1-\cos\beta}{4}\ [W.m^{-2}] \quad (5.4)$$

As only GHI is available, a decomposition model is also necessary to estimate the DHI and BHI required for the transposition. The most detailed temporal resolution for which GHI decomposition has been benchmarked is 1 min (Gueymard and Ruiz-Arias, 2016). From the various models assessed, the *Engerer2* model (Engerer, 2015) was highlighted for its good results and applicability for various climates. This model estimates the diffuse fraction (K_d) based on a logistic function, using as its inputs the K_t, the apparent solar time (AST), the sun elevation angle (α) and two additional variables related to cloud enhancement:

$$K_d = C + \frac{A-C}{1+\exp\left(B_0 + B_1\,K_t + B_2\,AST + B_3\,\alpha + B_4\,\Delta K_{t,c}\right)} + B_5 K_{de} \quad (5.5)$$

where A is the upper asymptote, equal to 1, C is the lower asymptote and B_n are the model coefficients (values listed in the table below).

Table 5.1 – Parameters for the Engerer2 decomposition model. Adapted from Engerer (2015).

Parameter	C	B_0	B_1	B_2	B_3	B_4	B_5
Value	4.2336×10^{-6}	-3.7912	7.5479	-1.0036×10^{-2}	3.1480×10^{-3}	-5.3146	1.7073

ΔK_{tc} quantifies the deviation between the K_t and the corresponding clear-sky condition (K_{tc}):

$$\Delta K_{tc} = K_{tc} - K_t = \frac{GHI_c}{I_{TOA,h}} - \frac{GHI}{I_{TOA,h}}\ \ [W.m^{-2}] \quad (5.6)$$

GHI_c was obtained from the Copernicus Atmosphere Monitoring Service (CAMS) McClear Clear-Sky Irradiation Service[10]. This service is based on a physical clear-sky model (Lefèvre et al., 2013) which considers latitude, longitude, altitude and a user-defined time resolution, as well as the integrated total column content of water vapor, ozone, and aerosol optical depth information.

On the other hand, K_{de} quantifies, when applicable, the fraction of GHI which results from cloud enhancement, assuming a value of zero if no cloud enhancement occurs:

$$K_{de} = MAX\left(0;\ 1 - \frac{GHI_c}{GHI}\right) \quad (5.7)$$

[10] http://www.soda-pro.com/web-services/radiation/cams-mcclear

To calculate the clear-sky index K_c for tilted surfaces, GHI_c was also decomposed, assuming ΔK_{tc} and K_{de} equal to 0, and transposed.

5.2.2. Modelling PV optical losses

To better grasp the implications of using data from PV modules with different tilt and orientation angles, the geometry-dependent PV optical losses due to reflectivity should be considered.

These optical losses (OL) were calculated for the beam irradiance (Martin and Ruiz, 2001) and the remaining components (Marion, 2017). The circumsolar fraction of the diffuse irradiance is assumed to have the same optical losses as the beam irradiance:

$$OL_{beam} = OL_{diffuse,circumsolar} = \frac{1 - e^{-\frac{cos\theta}{a_r}}}{1 - e^{-\frac{1}{a_r}}} \tag{5.8}$$

where a_r is the angular losses coefficient (0.169 for c-Si modules).

For each of the remaining components, these losses are calculated using a polynomial function which depends on the module tilt:

$$OL_{other} = a_0 + a_1\,\beta + a_2\,\beta^2 + a_3\,\beta^3 + a_4\,\beta^4 + a_5\,\beta^5 \tag{5.9}$$

The component-specific empirical coefficients are listed in Table 5.2.

Table 5.2 – Empirical coefficients to calculate the optical losses for each irradiance component. Adapted from Marion (2017).

	Empirical coefficients					
Irradiance component	a_0	a_1	a_2	a_3	a_4	a_5
Diffuse, horizon	9.5453×10^{-1}	3.8205×10^{-4}	1.2345×10^{-5}	-5.5902×10^{-7}	6.7806×10^{-9}	-2.9021×10^{-11}
Diffuse, sky	4.6333×10^{-2}	7.5181×10^{-2}	-2.6741×10^{-3}	4.8924×10^{-5}	-4.4356×10^{-7}	1.5696×10^{-9}
Reflected	1.1497×10^{-3}	6.0806×10^{-2}	-1.8826×10^{-3}	3.2026×10^{-5}	-2.7921×10^{-7}	9.6664×10^{-10}

5.2.3. Limitations

This method features several limitations. GHI data should be more variable than its corresponding PV generation. There is a spatial smoothing effect that increases with the size of a PV system (Mills et al., 2011), as changes in cloud cover will not impact all the modules at the same time. The same reasoning can be applied for the dispersion between PV systems when the aggregated generation is considered.

Additionally, the temperature impact on PV efficiency is not considered. However, the fact that temperature varies at considerably slower rates when compared with cloud-induced ramps, should have a minimal impact on the results.

On the other hand, inverter efficiency can also introduce some non-linearity in the data. The DC/AC conversion efficiency decreases for lower power output in respect to the inverter rated power. Moreover, this equipment can also saturate PV output whenever it surpasses the rated power, an effect designated as inverter clipping.

PV systems installed in urban areas are also likely to be occasionally shadowed by neighboring buildings which will disturb the spatio-temporal correlation between different sites. For short term forecasting, the focus of the present study, this is not expected to be of major concern, as casted shadows are slow-moving (when compared to moving clouds). Its modeling could be addressed by considering the local digital surface model and a shadow-casting algorithm (Freitas et al., 2015) but, since it is very local sensitive, it would lose general applicability.

5.3. Case studies

Throughout this section a network of PV systems in Oahu with the same layout as the irradiance sensors (c.f. Figure 3.2) is considered. Several case studies were designed as to assess the impact of module tilt and azimuth on a spatio-temporal solar forecasting model.

As mentioned in the previous section, the PV power data used here were generated using state of the art GHI decomposition and transposition algorithms, as well as an optical losses model.

5.3.1. The impact of tilted surfaces: one-neighbor model

As an exploratory analysis, the impact of PV system mounting on forecast skill is first assessed only considering the target site and its nearest upwind neighbor as inputs. As shown before (e.g. Figure 3.7), these two sites provide most of the relevant information for the forecasting model.

5.3.1.1. The value of neighboring horizontal irradiance data

To estimate the value of neighboring GHI data, only the target site is transposed. Various setups were tested, with tilt and orientation ranging, in 5-degree intervals, between 0º and 90º and 135º and 215º[11], respectively.

Results show that upwind GHI information improves the model's forecast skill for all mountings (Figure 5.1). However, its added value decreases with the module tilt. There is also an asymmetrical slight reduction as the orientation deviates from the south for tilts higher than 20º, with the model performing worse for west-oriented surfaces. Forecast skill is maximum for horizontal surfaces (28.8 %) ranging between 22.2 and 26.8 % for tilts equal to the Oahu's optimum tilt angle ±10º (i.e. equal to its latitude, maximizing the yearly generation) and 11.5 % and 14.4 % for vertical façades, depending on the orientation.

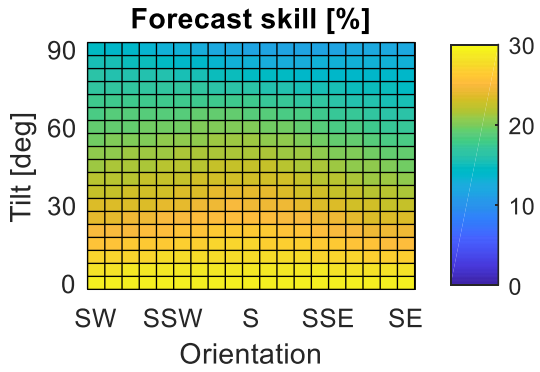

Figure 5.1 – Forecast skill (in %) for a forecast target with a given POA, considering its upwind sensor as a horizontal pyranometer.

The forecast skill is correlated with the overall diffuse fraction for the different tilted surfaces (Figure 5.2). This result supports the idea that surfaces with different mountings perceive changes in cloud cover, in terms of K_t or K_c, differently due to the different direct-diffuse contributions. The east-west asymmetry of the diffuse fraction can be explained by Oahu's local climate, generally cloudier in the morning that in the afternoon.

[11] South orientation corresponds to 180º.

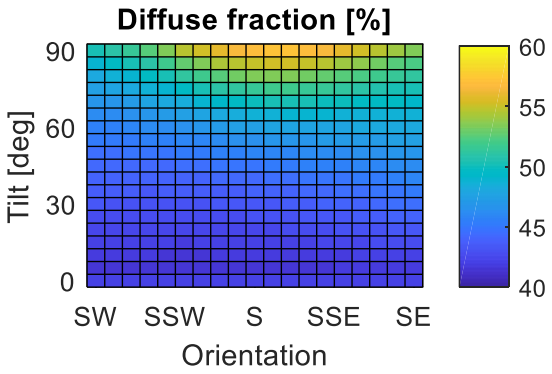

Figure 5.2 – Overall diffuse fraction (in %) for the site highlighted in red in Figure 3.7 assuming various tilts and orientations. The observed pattern appears to be symmetric to the one shown in Figure 5.1.

5.3.1.2. The value of neighboring PV data: optimal and vertical tilts

The same assessment was done assuming that the upwind site was a mounted PV system. Six setups were tested: optimal and vertical tilts; each with south, southwest and southeast orientations.

When the upwind site has an optimum tilt (Figure 5.3, top row), the model achieves forecast skill profiles similar to the one for an upwind horizontal pyranometer (Figure 5.1). There is a slight vertical shift, driven by the surface tilt, and a more noticeable horizontal shift, driven by its orientation. The skill values ranged between 11.6 % and 27.5 %.

A vertically-mounted upwind site, the linear ARX model ranks considerably worse skills when forecasting for a non-vertical PV system (down to 3.2 %), and slightly better results (up to 14.4 %) when both target and upwind sites share a similar POA (Figure 5.3, bottom row). A possible explanation is that vertical surfaces are less sensitive to changes in cloud cover (which has a stronger impact on direct irradiance) and thus are less relevant as upwind information sources.

It is also possible to observe that the forecast skill is considerably more sensitive to the surface's tilt and orientation for vertical installations (Figure 5.3). To quantify this sensitivity, the tilt, and orientation tolerance to which forecast skill relative reductions below 15 % are identified (Table 5.3). One can observe that optimum tilted surfaces are reasonably tolerant to differences in their neighbor's orientation and, to a lesser degree, inclination, but vertical façades can only achieve high forecast skills when there are vertical neighbors with almost the same orientation.

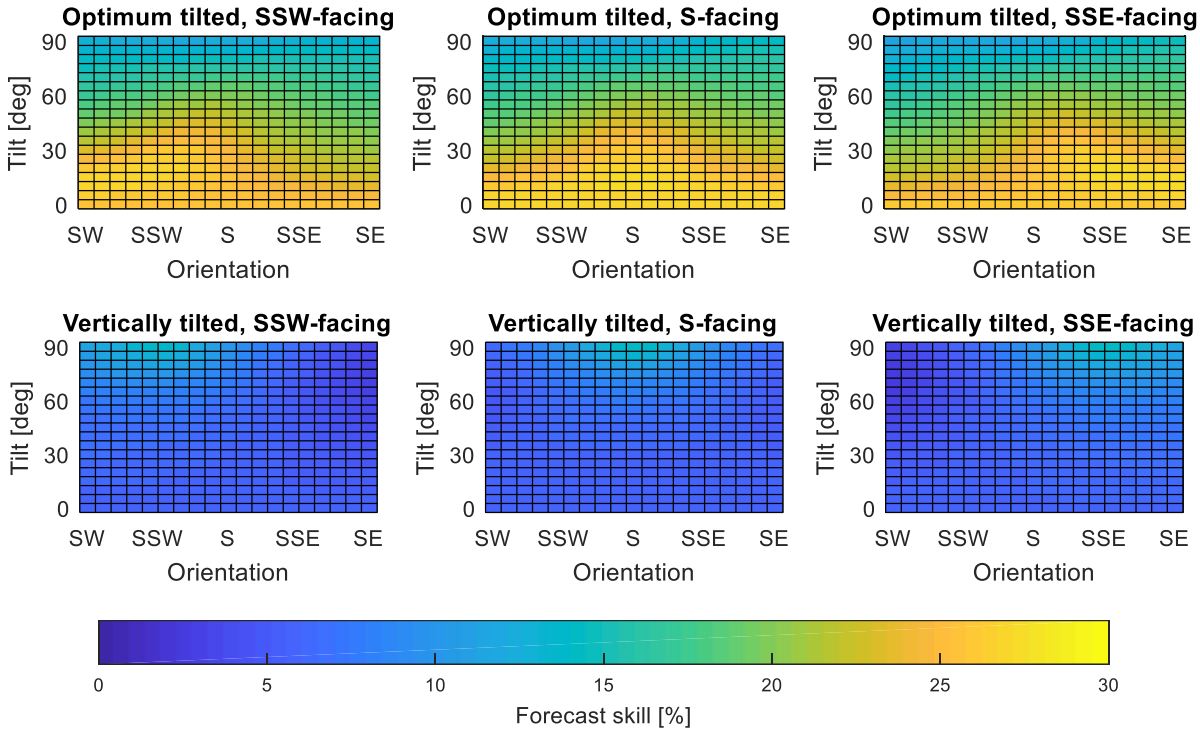

Figure 5.3 – Forecast skill for a PV surface with a given tilt and orientation, assuming for the upwind site either an optimum or vertical tilt, and three possible orientations Façades appear to be more sensitive to POA than non-vertical surfaces.

Table 5.3 - Upwind site tilt and orientation maximum deviation, when compared to the forecast target's mounting, which results in a maximum 15 % relative reduction.

Surface tilt	Maximum relative reduction [%]	Maximum absolute reduction [%]	Tilt tolerance [degrees]	Orientation tolerance [degrees]
Optimum	15	4	±15	±25
Vertical		2	-5	±15

5.3.2. The impact of diverse inclinations: many-neighbors model

For a more comprehensive analysis, a similar procedure was done considering the data from all sixteen sites. Four scenarios were designed assuming a south-facing target site, varying its tilt and the neighboring sites' mounting typology, as listed in Table 5.4.

Table 5.4 – Scenarios tested, where either the forecasting target or the neighboring sites' POA is varied.

Scenario	Forecast target	Neighboring sites
1A	South-facing rooftop surface	Rooftop surfaces with randomized tilt and orientation
1B		Vertical surfaces with randomized orientation
2A	South-facing façade	Rooftop surfaces with randomized tilt and orientation
2B		Vertical surfaces with randomized orientation

For each scenario, 1000 tilt and orientation sets were considered for the fifteen neighboring sites, randomly sampling from a distribution based on more than 30.000 PV systems in Europe (Leloux et al., 2015)[12]. For the scenarios with neighboring PV façades (1B and 2B), the orientation distribution was integrated over the various tilt angles (i.e. as a simplification, it is assumed that the façades share the same orientation as the rooftop PV systems). Moreover, as the results from the previous section pointed out that a mismatch in tilt and orientation was detrimental for the forecast, an "ideal" setup, where all surfaces shared the same POA, was also tested for each scenario.

For each of these 4004 sets, a linear ARX model was trained and assessed. Each model was also trained only considering the nearest upwind information, as in the previous one-neighbor setup. This allows one to quantify the accuracy gains from including a larger sensor network.

The results are shown in Table 5.5 and Figure 5.4. Despite the very diverse mounting sets tested, all scenarios led to positive forecast skills. One can observe that higher skill values could be achieved if all sites shared the same mounting typology, although results were rather sensitive to the varying tilt and orientation (an absolute 10.2 – 11.2 % skill difference between the 100^{th} and 0^{th} quantiles). Moreover, adding more sites to the model showed to have a very positive impact in the worst performing two-site sets and slightly improved the best performing ones (8.1 - 9.3 % versus 2.0 – 4.2 %), reducing the interquartile range (IQR, a measure of statistical dispersion, equal to the difference between 75^{th} and 25^{th} percentiles) by almost half.

Having neighbors with a mounting typology different from the target site has a larger detrimental impact when forecasting for rooftop than for vertical surfaces (-45.3 % versus -24.2 % relative reduction in forecast skill). A possible explanation is that a rooftop systems' ensemble with a range of tilts and orientations may be able to capture the anisotropic behavior of diffuse irradiance, which is the most relevant for vertical installations. The fact that

[12] Using a linear interpolation, a finer angle resolution of 2º, against the original 10º, was considered.

a lower skill value is obtained when all neighboring rooftop systems face the same direction (in this case, south) seems to further support this reasoning.

Table 5.5 – Forecast skill percentiles[13] and interquartile range (IQR) from the 1000 POA sets. Two forecast target tilts, four different scenarios pertaining to the neighbors' POA distribution and two different amounts of sites were considered for the linear ARX model. An "ideal" case, where all the sites share the same POA, was also tested.

Forecast target tilt	**Scenario**	**Forecast Skill [%]**							
		# of sites	**Percentiles**					**IQR**	**Ideal case**
			0th	**25th**	**50th**	**75th**	**100th**		
Optimum	1A	2	8.8	19.1	22.6	25.3	27.2	6.2	27.7
		16	18.1	24.3	26.2	27.9	29.2	3.6	29.6
	1B	2	4.2	5.8	5.9	6.0	6.2	0.2	5.9
		16	9.8	10.7	10.9	11.0	11.4	0.3	16.2
Vertical	2A	2	5.7	9.0	12.3	12.5	13.3	0.5	12.3
		16	9.2	13.3	13.6	13.9	14.7	0.6	9.9
	2B	2	-0.1	6.4	10.3	13.7	14.2	7.2	17.2
		16	8.2	14.0	16.0	17.3	18.4	3.3	19.4

[13] 0^{th} and 100^{th} correspond to minimum and maximum values of the data sets, respectively.

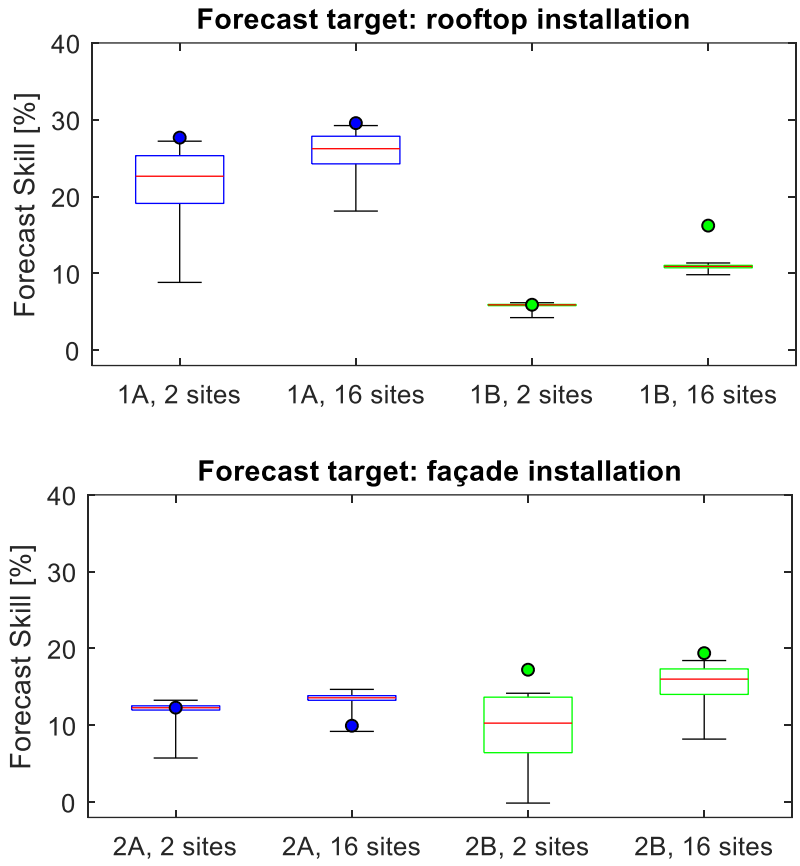

Figure 5.4 – Forecast skill range for the scenarios 1A and 1B (top row) and 2A and 2B (bottom row), when considering either 2 sites (forecast target and closest upwind sensor) or all 16 sites. Blue and green boxplots correspond, respectively, to rooftop and vertical neighbors. The circles represent an "ideal" case where all sites share the same tilt/orientation.

5.4. Final remarks

The results presented in this chapter showed that results from spatially distributed horizontal irradiance data sets for very-short term forecasting cannot be directly translated to an equivalent PV forecasting context as it would systematically overestimate forecasting performance.

By means of an exploratory analysis was possible to quantify the detrimental impact of tilt and orientation mismatching. Surfaces with different tilt and orientation angles react differently to an identical change in cloud cover, supporting the idea that normalizing solar data by a clear-sky reference is not totally effective in removing its geometry dependence.

Thus, the mounting heterogeneity potentially present in PV ensemble data sets has a relevant impact on a model's forecast skill. This is most relevant for very short-term forecasting where irradiance and diffuse fraction may vary abruptly. It ought to be noted that transposition models are yet unable to fully capture the complexity of the diffuse irradiance angular

dependence. The current approach, with a non-isotropic yet largely smooth underlying sky radiance, is likely underestimating the impact of the diverse POA's. Hence, it should be seen as a best-case scenario, for which the tilt/orientation of neighboring and/or target PV systems is already a relevant factor on the forecast skill. In real applications (if/when short term PV data are available) the impact of tilt/orientation is therefore expected to be even more pertinent.

Results also showed that the linear ARX model systematically achieves lower performance values for vertical PV installations. Two possible explanations arise: i) vertical installations are more difficult to forecast since they depend more on the complex diffuse irradiance; ii) that same diffuse irradiance dependency, makes it less sensitive to changes in cloud cover, and thus benefit less from upwind information.

As far as a generalization of these results to other locations and contexts are concerned, spatio-temporal approaches have shown positive results for various case studies, with different climates and spatial/temporal scales. However, it is worth underlining that these models depend on the local wind patterns, the matching between the spatial and temporal scales (i.e. longer horizons require larger spatial coverage) and temporal resolution (i.e. how detailed cloud-induced ramps can be detected).

As future work, machine learing algorithms should be tested to model the non-linearities introduced in the data by the complexity of the environment, inverter and mounting diversity.

6. Conclusions

Over the last few years, power systems have had to handle an increasing penetration of variable generation, mainly from renewable sources such as wind and solar. Optimal grid operation balancing technical and economic conditions is currently supported by various markets where stakeholders bid for both generation and load. There are different markets targeting different time scales; some focus on energy bids for day-ahead in 1-h time intervals while others focus on short time periods of a few seconds. This gives room for the power system to integrate generators with different operational constraints, minimizing the risks of imbalance and scheduling reserves to mitigate the impacts of generation (mostly from renewables) and load variability.

Photovoltaics (PV) is reaching deployment levels that can no longer be dismissed by power grid operators. PV is in general a non-dispatchable and considerably variable energy source, with a very well understood daily and annual cycle. However, the real challenge is its forecasting uncertainty, as short-term solar variability is mostly caused by changes in cloud cover, and clouds can be very small elements that are difficult to accurately model and forecast. It thus becomes clear that the effective operation of a power system requires accurate PV generation forecasts.

Several operational solar forecasting solutions already exist. Numerical Weather Prediction (NWP) methods model the physical weather dynamics through numerical approximations considering the atmosphere's initial condition in a gridded data format. Even though NWP traditionally excel in predicting the longer-term trends of weather, recent advances have led to considerable improvements for short-term forecasting. However, NWP is computationally too demanding for real-time very short-term forecasting, although its outputs can be useful as supplementary inputs for statistical models.

Image-based methods for satellite data, on the other hand, exploit measurements targeting the appropriate wavelength ranges to identify cloud presence and estimate the transmissivity of the atmosphere. Then, from a temporal sequence of images, a cloud motion vector is inferred, and a forecast is made by shifting the latest image according to that same vector and the intended time horizon. Satellites offer an interesting spatial coverage, a spatial resolution on the order of a few squared kilometers, a temporal resolution of 15 min and a delay in data delivery which is compatible with short-term forecasts. Moreover, the new satellites that are being put in orbit, have superior resolutions that can make a difference for solar forecasting.

However, their level of detail will still be insufficient for either very short-term horizons or very localized forecasts.

An emerging trend is to use sky imagers for very-short term forecasts. Sustained in the same principles of satellite methods (i.e. inferring sky transmittance and shifting an image based on a cloud motion vector), sky imagers provide information with considerably more detail but in three color channels (red, green, blue). These methods are commonly tested for forecast horizons in the seconds to minutes time scales. Since perspective issues have been reported when using a single camera, some authors have proposed using a set of spatially distributed cameras, at the expense of additional hardware and computational resources. However, despite the interesting results in cloud and cloud motion vector identification, a lower accuracy in estimating the solar resource, difficulties in identifying thin clouds and sensitivity to the environment (i.e. soiling, dew or raindrops) pose considerable challenges. In a recent literature review, the authors commented that these limitations combined with the development of better satellites would make sky cameras obsolete for most solar forecasting applications.

Time series methods are based on past measurements from the target variable itself, be it PV power or irradiance. The simplest and most common approach is persistence, where the variable is assumed to persist in time. More sophisticated solutions are to use this data as input for statistical autoregressive models. Several of these models exist, with different characteristics and designations, such as AR, ARMA, ARIMA or SARIMA. While these models perform considerably well for stable weather conditions, and for a vast range of time scales, the lack of spatial information makes these models unable to anticipate the arrival of incoming clouds.

A larger PV deployment will necessarily imply the need for better solar forecasting, but, it will also result in larger amounts of solar power data (which has been shown to be a good proxy for the solar resource). These data enable direct forecasting of PV generation, bypassing the need for conversion models which are themselves uncertain. Moreover, they intrisincally capture various relevant factors such as module temperature and sun apparent position. Thus, if the additional data can leverage new and more accurate forecasting approaches, synergies can be created between PV deployment and solar forecasting.

Following this idea, over the last few years, there has been a surge of solar forecasting research on spatio-temporal methods exploring spatially distributed solar data sets. Two main

approaches have been identified: i) mapping sky transmittance from the data and converting it to a forecast by shifting the map according to a cloud motion vector, or; ii) ingesting the data directly into (linear or non-linear) statistical models.

Several works have explored PV data directly and reported interesting results. However, since distributed PV generation data is usually either unavailable (due to privacy protection issues) and/or with a coarse time resolution (typically 15-min averages), most spatio-temporal forecasting works focusing on very-short term horizons explore data sets acquired from irradiance measurement campaigns that took place in locations such as Oahu and Lanai, in Hawaii, or Jülich and Melpitz, in Germany, where dense irradiance sensor networks were temporarily deployed. More recently, possibly inspired by the promising results using ground data, these same methods have been tested with satellite estimates and even NWP irradiance outputs.

This thesis has highlighted the need to revisit the definition and classification of approaches that explore spatially distributed solar time series. Although the "spatio-temporal solar forecasting" expression dominates the current literature, and has been used in this thesis, in fact, most solar forecasting approaches have both a spatial and temporal component.

The main contributions of this thesis explore spatially distributed solar forecasting trying to answer the research question: what are the potential and limitations of spatio-temporal solar forecasting? This question is then addressed in three chapters, each focusing on a particular topic.

The first addressed issue is the identification of the spatial and temporal scales where spatio-temporal forecasting is more effective. Despite there being some literature exploring this forecasting approach, with data sets covering regions from 1 to 100,000 km^2, most of them lack a spatio-temporal analysis of their model dynamics and achieved results. Thus, a linear multivariate model was trained using three data sets with different spatial and temporal scales (10 s/1 km^2, 30 min/1,600 km^2, and daily/320,000 km^2). All three data sets correspond to locations that are known for having prevailing wind patterns, as to ensure strong spatio-temporal correlations. As a result, the spatially distributed irradiance data led to better forecasts for all data sets and corresponding sites. When forecast performance is mapped, gradients symmetrical to the local wind patterns can be observed, as the model benefits from upwind information.

Three forecasting modes were identified, where neighboring data either:

i) effectively help to predict the arrival of incoming clouds;
ii) produce more conservative, smoother, forecasts which still surpass persistence or;
iii) add little value to the forecasts' accuracy.

The assignment of each of these modes to a given model is highly dependent on the local weather patterns, the forecast horizon of interest and the layout of the network of sensors.

The spatio-temporal approach performed best for the downwind sites in every data set, i.e. for different temporal and spatial scales . For such sites, these approaches can both tackle incoming individual clouds, larger cloud packs or even synoptic-scale weather systems. Moreover, for the first data set, where the spatio-temporal correlations are stronger and more localized, it was clear that as the horizon increased, the relevant information shifted to farther neighboring sites, showing the importance of having a dense but spread out network.

The following chapter addressed the relevance of wind information for spatio-temporal forecasting and how it could be integrated into these models. Although cloud advection patterns are driven by cloud-height wind speed and direction, the use of such variables was mostly unexplored. The forecast skill of a high-performing site (i.e. with adequate upwind information) was disaggregated according to the wind direction at a pressure level where cumulus clouds (common in Oahu, Hawaii) tend to form. Despite the overall positive forecast skill, the linear ARX model is, in fact, surpassed by persistence in most non-prevailing wind directions. By training different forecasting models, each specialized in a specific wind regime, the model became able to detect different advection patterns and achieved positive skill values for all wind directions and speeds, showing the potential for combining spatio-temporal and wind data for accurate short and very short-term solar forecasting.

Then, the question of how the varying geometry in PV systems impact spatio-temporal forecasting was addressed. Several works have highlighted the potential of high-resolution ground data from dense sensor networks for very-short term forecasting. However, none have used actual PV data, as it is commonly stored in 15-min averages. To assess the impact of varying PV module mounting, the irradiance data set which was used throughout this thesis was transposed using state-of-the-art decomposition and transposition algorithms for different tilt and orientation angle realistic distributions (according to reported distributions for PV systems across Europe). High forecast skill values were achieved, although lower than for a network of horizontal irradiance sensors. Results showed to be correlated with the difference

in tilt and orientation between target and neighboring PV modules. This is easy to understand if one considers that cloud cover impacts differently the direct and diffuse components of solar irradiance. Thus, the relative ramping caused by a given change in cloud cover depends on the surface, as it affects the contribution from each of the irradiance components.

Results also indicate that normalizing solar data by a clear-sky reference is not totally effective in removing its geometry dependence. Forecasting skill is more sensitive to neighboring geometry whenever the PV systems share a similar mounting (either rooftop or façade). Rooftop systems benefit the most from neighbors that share the same tilt and orientation. However, PV façades with neighboring rooftop systems seem to benefit from tilt/orientation diversity, possibly because it captures better the anisotropic nature of diffuse irradiance. This larger dependence on diffuse irradiance also seems to make forecasting for PV façades more challenging, with overall lower forecast skill values achieved.

To conclude, future research directions were identified. Spatio-temporal forecasting should be tested for locations where wind patterns are more evenly distributed. A direct implementation of this approach is expected to perform poorly, due to the presumably weaker spatio-temporal correlations present in the data. However, this is exactly the context which should favor a regime-based approach driven by wind data, as it should be able to isolate and properly model the different cloud advection patterns. To test this, large eddy simulation (LES) could be used to generate synthetic data, as cloud formation and cloud-height wind patterns can be user-defined.

Spatio-temporal analysis using high frequency real PV data sets clearly requires further work. The results presented in this thesis have shown that reaching out to PV power data for spatio-temporal forecasting raises difficulties usually disregarded in well-maintained dedicated irradiance measurement networks such as PV module tilt and orientation, the impacts of PV module temperature on its conversion efficiency, the effect of inverter efficiency and inverter clipping, shadowing from neighboring obstacles, PV module and/or inverter degradation; malfunction, etc.. These complex effects could perhaps be better addressed by machine learning methods, such as artificial neural netoworks, which excel at identifying relevant non-linear patterns in data, albeit at the cost of some additional computational expense and loss of understanding of the underlying physical mechanisms.

Spatio-temporal solar forecasting based on PV generation data from distributed PV systems across an urban area, a large solar power plant, a region or even at continental scale, will be

an essential contribution to the wider deployment of PV promoting its grid integration whilst feeding itself on the massive amount of data that high levels of penetration of PV will yield.

References

Agoua, X.G., Girard, R., Kariniotakis, G., 2018. Short-Term Spatio-Temporal Forecasting of Photovoltaic Power Production. IEEE Trans. Sustain. Energy 9, 538–546. https://doi.org/10.1109/TSTE.2017.2747765

Alessandrini, S., Delle Monache, L., Sperati, S., Cervone, G., 2015. An analog ensemble for short-term probabilistic solar power forecast. Appl. Energy 157, 95–110. https://doi.org/10.1016/j.apenergy.2015.08.011

Almonacid, F., Pérez-Higueras, P.J., Fernández, E.F., Hontoria, L., 2014. A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator. Energy Convers. Manag. 85, 389–398. https://doi.org/10.1016/j.enconman.2014.05.090

Amaro e Silva, R., Brito, M.C., 2019. Spatio-temporal PV forecasting sensitivity to modules' tilt and orientation. Appl. Energy 255, 113807. https://doi.org/10.1016/j.apenergy.2019.113807

Amaro e Silva, R., Brito, M.C., 2017. Understanding Spatio-temporal Solar Forecasting. 7th Sol. Integr. Work. Berlin, Ger.

Amaro e Silva, R., C. Brito, M., 2018. Impact of network layout and time resolution on spatio-temporal solar forecasting. Sol. Energy 163, 329–337. https://doi.org/10.1016/j.solener.2018.01.095

Amaro e Silva, R., Haupt, S.E., Brito, M.C., 2019. A regime-based approach for integrating wind information in spatio-temporal solar forecasting models. J. Renew. Sustain. Energy 11, 056102. https://doi.org/10.1063/1.5098763

Amaro e Silva, R., Monteiro Baptista, J., C. Brito, M., 2018. Data-driven estimation of expected photovoltaic generation. Sol. Energy 166, 116–122. https://doi.org/10.1016/j.solener.2018.03.039

Andrade, J.R., Bessa, R.J., 2017. Improving Renewable Energy Forecasting with a Grid of Numerical Weather Predictions. IEEE Trans. Sustain. Energy 3029, 1–1. https://doi.org/10.1109/TSTE.2017.2694340

André, M., Perez, R., Soubdhan, T., Schlemmer, J., Calif, R., Monjoly, S., 2019. Preliminary assessment of two spatio-temporal forecasting technics for hourly satellite-derived irradiance in a complex meteorological context. Sol. Energy 177, 703–712. https://doi.org/10.1016/j.solener.2018.11.010

Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-pison, F.J., Antonanzas-torres, F., 2016. Review of photovoltaic power forecasting. Sol. Energy 136, 78–111. https://doi.org/10.1016/j.solener.2016.06.069

Aryaputera, A.W., Yang, D., Zhao, L., Walsh, W.M., 2015. Very short-term irradiance forecasting at unobserved locations using spatio-temporal kriging. Sol. Energy 122, 1266–1278. https://doi.org/10.1016/j.solener.2015.10.023

Bacher, P., Madsen, H., Nielsen, H.A., 2009. Online short-term solar power forecasting. Sol. Energy 83, 1772–1783. https://doi.org/10.1016/j.solener.2009.05.016

Badescu, V., 2002. 3D isotropic approximation for solar diffuse irradiance on tilted surfaces. Renew. Energy 26, 221–233. https://doi.org/10.1016/S0960-1481(01)00123-9

Baig, A., Akhter, P., Mufti, A., 1991. A novel approach to estimate the clear day global radiation. Renew. Energy 1, 119–123. https://doi.org/10.1016/0960-1481(91)90112-3

Bartholomy, O., Vargas, T., Simone, M., Hansen, C., Fitchett, S., Pohl, A., 2014. Benchmarking Solar Power and Irradiance Forecasting Accuracy at Sacramento Municipal Utility District 63–68.

Benali, L., Notton, G., Fouilloy, A., Voyant, C., Dizene, R., 2019. Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. Renew. Energy 132, 871–884. https://doi.org/10.1016/j.renene.2018.08.044

Benjamin, S.G., Weygandt, S.S., Brown, J.M., Hu, M., Alexander, C.R., Smirnova, T.G., Olson, J.B., James, E.P., Dowell, D.C., Grell, G.A., Lin, H., Peckham, S.E., Smith, T.L., Moninger, W.R., Kenyon, J.S., Manikin, G.S., 2016. A North American hourly assimilation and model forecast cycle: The rapid refresh. Mon. Weather Rev. 144, 1669–1694. https://doi.org/10.1175/MWR-D-15-0242.1

Berdugo, V., Chaussin, C., Dubus, L., Hebrail, G., Leboucher, V., 2011. Analog Method for Collaborative Very-Short-Term Forecasting of Power Generation from Photovoltaic Systems. Next Gener. Data Min. Summit Ubiquitous Knowl- edge Discov. Energy Manag. Smart Grids Intell. Mach. to-Machine Telemat.

Bernecker, D., Riess, C., Angelopoulou, E., Hornegger, J., 2014. Continuous short-term irradiance forecasts using sky images. Sol. Energy 110, 303–315. https://doi.org/10.1016/j.solener.2014.09.005

Bertrand, C., Housmans, C., Leloux, J., Journée, M., 2018. Solar irradiation from the energy production of residential PV systems. Renew. Energy 125, 306–318. https://doi.org/10.1016/j.renene.2018.02.036

Bessa, R.J., Moreira, C., Silva, B., Matos, M., 2014. Handling renewable energy variability and uncertainty in power systems operation. Wiley Interdiscip. Rev. Energy Environ. 3, 156–178. https://doi.org/10.1002/wene.76

Bessa, R.J., Rua, D., Abreu, C., Machado, P., Andrade, J.R., Pinto, R., Gonçalves, C., Reis, M., 2018. Data economy for prosumers in a smart grid ecosystem. e-Energy 2018 - Proc. 9th ACM Int. Conf. Futur. Energy Syst. 622–630. https://doi.org/10.1145/3208903.3210282

Bessa, R.J., Trindade, A., Miranda, V., 2015a. Spatial-temporal solar power forecasting for smart grids. IEEE Trans. Ind. Informatics 11, 232–241. https://doi.org/10.1109/TII.2014.2365703

Bessa, R.J., Trindade, A., Silva, C.S.P., Miranda, V., 2015b. Probabilistic solar power forecasting in smart grids using distributed information. Int. J. Electr. Power Energy Syst. 72, 16–23. https://doi.org/10.1016/j.ijepes.2015.02.006

Blanc, P., Wald, L., 2012. The SG2 algorithm for a fast and accurate computation of the position of the Sun for multi-decadal time period. Sol. Energy 86, 3072–3083. https://doi.org/10.1016/j.solener.2012.07.018

Boland, J., 2015. Spatial-temporal forecasting of solar radiation. Renew. Energy 75, 607–616. https://doi.org/10.1016/j.renene.2014.10.035

Bosch, J.L., Kleissl, J., 2013. Cloud motion vectors from a network of ground sensors in a solar power plant. Sol. Energy 95, 13–20. https://doi.org/10.1016/j.solener.2013.05.027

Bosch, J.L., Zheng, Y., Kleissl, J., 2013. Deriving cloud velocity from an array of solar radiation measurements. Sol. Energy 87, 196–203. https://doi.org/10.1016/j.solener.2012.10.020

Bright, J.M., Killinger, S., Engerer, N.A., 2019. Data article: Distributed PV power data for three cities in Australia. J. Renew. Sustain. Energy 11, 035504. https://doi.org/10.1063/1.5094059

Brinkman, G., Lew, D., Hummon, M., Ibanez, E., Ela, E., Hodge, B., 2012. Impacts of Reserve Methodology on Production Cost in High-penetration Renewable Scenarios, in: 2nd Solar Integration Workshop.

Chaturvedi, D.K., Isha, 2016. Solar Power Forecasting: A Review. Int. J. Comput. Appl. 145, 28–50. https://doi.org/10.5120/ijca2016910728

Chow, C.W., Belongie, S., Kleissl, J., 2015. Cloud motion and stability estimation for intra-hour solar forecasting. Sol. Energy 115, 645–655. https://doi.org/10.1016/j.solener.2015.03.030

Chow, C.W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., Washom, B., 2011. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. Sol. Energy 85, 2881–2893. https://doi.org/10.1016/j.solener.2011.08.025

Cros, S., Buessler, E., Guillou, F. Le, Turpin, M., 2017. Country Scale Solar Irradiance Forecasting for PV Power Trading, in: 7th Solar Integration Workshop. pp. 745–772.

Cros, S., Liandrat, O., Sebastien, N., Schmutz, N., 2014. Extracting cloud motion vectors from satellite images for solar power forecasting. Int. Geosci. Remote Sens. Symp. 4123–4126. https://doi.org/10.1109/IGARSS.2014.6947394

Dambreville, R., Blanc, P., Chanussot, J., Boldo, D., 2014. Very short term forecasting of the global horizontal irradiance using a spatio-temporal autoregressive model. Renew. Energy 72, 291–300. https://doi.org/10.1016/j.renene.2014.07.012

de Freitas Viscondi, G., Alves-Souza, S.N., 2019. A Systematic Literature Review on big data for solar photovoltaic electricity generation forecasting. Sustain. Energy Technol. Assessments 31, 54–63. https://doi.org/10.1016/j.seta.2018.11.008

Diagne, M., David, M., Lauret, P., Boland, J., Schmutz, N., 2013. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. Renew. Sustain. Energy Rev. 27, 65–76. https://doi.org/10.1016/j.rser.2013.06.042

Dobbs, A., Elgindy, T., Hodge, B.-M., Florita, A., 2017. Short-Term Solar Forecasting Performance of Popular Machine Learning Algorithms. Int. Work. Integr. Sol. Power into Power Syst. (Solar Integr. Work. 6.

Dolara, A., Leva, S., Manzolini, G., 2015. Comparison of different physical models for PV power output prediction. Sol. Energy 119, 83–99. https://doi.org/10.1016/j.solener.2015.06.017

Ela, E., Diakov, V., Ibanez, E., Heaney, M., 2013. Impacts of variability and uncertainty in solar photovoltaic generation at multiple timescales. Natl. Renew. Energy Lab. https://doi.org/NREL/TP-5500-58274

Elsinga, B., van Sark, W.G.J.H.M., 2017. Short-term peer-to-peer solar forecasting in a network of photovoltaic systems. Appl. Energy 206, 1464–1483. https://doi.org/10.1016/j.apenergy.2017.09.115

Engerer, N.A., 2015. Minute resolution estimates of the diffuse fraction of global irradiance for southeastern Australia. Sol. Energy 116, 215–237. https://doi.org/10.1016/j.solener.2015.04.012

Engerer, N.A., Bright, J.M., Killinger, S., 2018. Himawari-8 Enabled Real-Time Distributed PV Simulations for Distribution Networks. 2017 IEEE 44th Photovolt. Spec. Conf. 1405–1410. https://doi.org/10.1109/pvsc.2017.8521518

Engerer, N.A., Mills, F.P., 2014. KPV: A clear-sky index for photovoltaics. Sol. Energy 105, 679–693. https://doi.org/10.1016/j.solener.2014.04.019

Espinar, B., Aznarte, J.L., Girard, R., Moussa, a. M., Kariniotakis, G., 2010. Photovoltaic Forecasting: A state of the art. 5th Eur. PV-Hybrid Mini-Gird Conf. 33, 250–255.

Ferrari, S., Lazzaroni, M., Piuri, V., Cristaldi, L., Faifer, M., 2013. Statistical models approach for solar radiation prediction. Conf. Rec. - IEEE Instrum. Meas. Technol. Conf. 1734–1739. https://doi.org/10.1109/I2MTC.2013.6555712

Freitas, S., Catita, C., Redweik, P., Brito, M.C., 2015. Modelling solar potential in the urban environment: State-of-the-art review. Renew. Sustain. Energy Rev. 41, 915–931. https://doi.org/10.1016/j.rser.2014.08.060

Fu, C.L., Cheng, H.Y., 2013. Predicting solar irradiance with all-sky image features via regression. Sol. Energy 97, 537–550. https://doi.org/10.1016/j.solener.2013.09.016

Gagne, D.J., McGovern, A., Haupt, S.E., Williams, J.K., 2017. Evaluation of statistical learning configurations for gridded solar irradiance forecasting. Sol. Energy 150, 383–393. https://doi.org/10.1016/j.solener.2017.04.031

Ghonima, M.S., Urquhart, B., Chow, C.W., Shields, J.E., Cazorla, A., Kleissl, J., 2012. A method for cloud detection and opacity classification based on ground based sky imagery. Atmos. Meas. Tech. 5, 2881–2892. https://doi.org/10.5194/amt-5-2881-2012

Glahn, H.R., Lowry, D.A., 1972. The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. J. Appl. Meteorol. 11, 1203–1211. https://doi.org/10.1175/1520-0450(1972)011<1203:tuomos>2.0.co;2

Gueymard, C.A., Ruiz-Arias, J.A., 2016. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. Sol. Energy 128, 1–30. https://doi.org/10.1016/j.solener.2015.10.010

Gutierrez-Corea, F.V., Manso-Callejo, M.A., Moreno-Regidor, M.P., Manrique-Sancho, M.T., 2016. Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations. Sol. Energy 134, 119–131. https://doi.org/10.1016/j.solener.2016.04.020

Hammer, A., Heinemann, D., Hoyer, C., Kuhlemann, R., Lorenz, E., Müller, R., Beyer, H.G., 2003. Solar energy assessment using remote sensing technologies. Remote Sens. Environ. 86, 423–432. https://doi.org/10.1016/S0034-4257(03)00083-X

Hammer, A., Heinemann, D., Lorenz, E., Lückehe, B., 1999. Short-term forecasting of solar radiation: a statistical approach using satellite data. Sol. Energy 67, 139–150. https://doi.org/10.1016/S0038-092X(00)00038-4

Hammer, A., Kuhnert, J., Weinreich, K., Lorenz, E., 2015. Short-term forecasting of surface solar irradiance based on Meteosat-SEVIRI data using a nighttime cloud index. Remote Sens. 7, 9070–9090. https://doi.org/10.3390/rs70709070

Haupt, S.E., Kosović, B., Jensen, T., Lazo, J.K., Lee, J.A., Jiménez, P.A., Cowie, J., Wiener, G., Mccandless, T.C., Rogers, M., Miller, S., Sengupta, M., Xie, Y., HinKeLman, L., KaLb, P., Heiser, J.H., 2018. Building the SUN4CAST system. Bull. Am. Meteorol. Soc. 99, 121–135. https://doi.org/10.1175/BAMS-D-16-0221.1

Heinle, A., Macke, A., Srivastav, A., 2010. Automatic cloud classification of whole sky images. Atmos. Meas. Tech. 3, 557–567. https://doi.org/10.5194/amt-3-557-2010

Hinkelman, L.M., 2013. Differences between along-wind and cross-wind solar irradiance variability on small spatial scales. Sol. Energy 88, 192–203. https://doi.org/10.1016/j.solener.2012.11.011

Hocker, J.E., Basara, J.B., 2008a. A 10-year spatial climatology of squall line storms across Oklahoma. Int. J. Climatol. 28, 765–775. https://doi.org/10.1002/joc.1579

Hocker, J.E., Basara, J.B., 2008b. A geographic information systems-based analysis of supercells across Oklahoma from 1994 to 2003. J. Appl. Meteorol. Climatol. 47, 1518–1538. https://doi.org/10.1175/2007JAMC1673.1

Ibanez, E., Brinkman, G., Hummon, M., Lew, D., 2012. A Solar Reserve Methodology for Renewable Energy Integration Studies Based on Sub-Hourly Variability Analysis Preprint NREL\CP-5500-56169. 2nd Sol. Integr. Work. https://doi.org/NREL/CP-5500-56169

Ineichen, P., Perez, R., 2002. A new airmass independent formulation for the Linke turbidity coefficient. Sol. Energy 73, 151–157. https://doi.org/http://dx.doi.org/10.1016/S0038-092X(02)00045-2

Inman, R.H., Pedro, H.T.C., Coimbra, C.F.M., 2013. Solar forecasting methods for renewable energy integration. Prog. Energy Combust. Sci. 39, 535–576. https://doi.org/10.1016/j.pecs.2013.06.002

International Energy Agency, 2014. Technology roadmap: solar photovoltaic energy. https://doi.org/10.1787/9789264088047-en

Jang, H.S., Bae, K.Y., Park, H.S., Sung, D.K., 2016. Solar Power Prediction Based on Satellite Images and Support Vector Machine. IEEE Trans. Sustain. Energy 7, 1255–1263. https://doi.org/10.1109/TSTE.2016.2535466

Jiménez, P.A., Alessandrini, S., Haupt, S.E., Deng, A., Kosovic, B., Lee, J.A., Monache, L.D., 2016. The role of unresolved clouds on short-range global horizontal irradiance predictability. Mon. Weather Rev. 144, 3099–3107. https://doi.org/10.1175/MWR-D-16-0104.1

Jimenez, P.A., Hacker, J.P., Dudhia, J., Haupt, S.E., Ruiz-Arias, J.A., Gueymard, C.A., Thompson, G., Eidhammer, T., Deng, A., 2016. WRF-SOLAR: Description and clear-sky assessment of an augmented NWP model for solar power prediction. Bull. Am. Meteorol. Soc. 97, 1249–1264. https://doi.org/10.1175/BAMS-D-14-00279.1

Kemmoku, Y., Orita, S., Nakagawa, S., Sakakibara, T., 1999. Daily insolation forecasting using a multi-stage neural network. Sol. Energy 66, 193–199.

Killinger, S., Braam, F., Müller, B., Wille-Haussmann, B., McKenna, R., 2016. Projection of power generation between differently-oriented PV systems. Sol. Energy 136, 153–165. https://doi.org/10.1016/j.solener.2016.06.075

Killinger, S., Lingfors, D., Saint-Drenan, Y.-M., Moraitis, P., Sark, W. Van, Taylor, J., Engerer, N.A., Bright, J.M., 2018. On the search for representative characteristics of PV systems: Data collection and analysis of PV system azimuth , tilt , capacity , yield and shading. Sol. Energy 173, 1087–1106. https://doi.org/10.1016/j.solener.2018.08.051

King, D.L., Kratochvil, J.A., Boyson, W.E., 1997. Measuring solar spectral and angle-of-incidence effects on photovoltaic modules and solar irradiance sensors. Conf. Rec. Twenty Sixth IEEE Photovolt. Spec. Conf. - 1997 1113–1116. https://doi.org/10.1109/PVSC.1997.654283

Kuhn, P., Nouri, B., Wilbert, S., Prahl, C., Kozonek, N., Schmidt, T., Yasser, Z., Ramirez, L., Zarzalejo, L., Meyer, A., Vuilleumier, L., Heinemann, D., Blanc, P., Pitz-Paal, R., 2017. Validation of an all-sky imager–based nowcasting system for industrial PV plants. Prog. Photovoltaics Res. Appl. 26, 608–621. https://doi.org/10.1002/pip.2968

Kumler, A., Xie, Y., Zhang, Y., 2019. A Physics-based Smart Persistence model for Intra-hour forecasting of solar radiation (PSPI) using GHI measurements and a cloud retrieval technique. Sol. Energy 177, 494–500. https://doi.org/10.1016/j.solener.2018.11.046

Kurtz, B., Kleissl, J., 2017. Measuring diffuse, direct, and global irradiance using a sky imager. Sol. Energy 141, 311–322. https://doi.org/10.1016/j.solener.2016.11.032

Kurtz, B., Mejia, F., Kleissl, J., 2018. A virtual sky imager testbed for solar energy forecasting. Sol. Energy 158, 753–759. https://doi.org/10.1016/j.solener.2017.10.036

Kushwaha, V., Pindoriya, N.M., 2018. Very short-term solar PV generation forecast using SARIMA model: A case study. 2017 7th Int. Conf. Power Syst. ICPS 2017 430–435. https://doi.org/10.1109/ICPES.2017.8387332

Kuszamaul, S., Ellis, A., Stein, J., Johnson, L., 2010. Lanai High-Density Irradiance Sensor Network for characterizing SOLAR resource variability of MW-scale PV system. Conf. Rec. IEEE Photovolt. Spec. Conf. 283–288. https://doi.org/10.1109/PVSC.2010.5615868

Lappalainen, K., Valkealahti, S., 2015. Recognition and modelling of irradiance transitions caused by moving clouds. Sol. Energy 112, 55–67. https://doi.org/10.1016/j.solener.2014.11.018

Larson, V.E., 2013. Forecasting Solar Irradiance with Numerical Weather Prediction Models, in: Kleissl, J. (Ed.), Solar Energy Forecasting and Resource Assessment. Elsevier Academic Press, pp. 299–318.

Lauret, P., Diagne, M., David, M., 2014. A neural network post-processing approach to improving NWP solar radiation forecasts. Energy Procedia 57, 1044–1052. https://doi.org/10.1016/j.egypro.2014.10.089

Lauret, P., Voyant, C., Soubdhan, T., David, M., Poggi, P., 2015. A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. Sol. Energy 112, 446–457. https://doi.org/10.1016/j.solener.2014.12.014

Lave, M., Quiroz, J., Reno, M.J., Broderick, R.J., 2016. High temporal resolution load variability compared to PV variability. Conf. Rec. IEEE Photovolt. Spec. Conf. 2016-Novem, 1831–1836. https://doi.org/10.1109/PVSC.2016.7749938

Lave, M., Reno, M.J., Broderick, R.J., 2015. Characterizing local high-frequency solar variability and its impact to distribution studies. Sol. Energy 118, 327–337. https://doi.org/10.1016/j.solener.2015.05.028

Lee, J.A., Haupt, S.E., Jiménez, P.A., Rogers, M.A., Miller, S.D., McCandless, T.C., 2017. Solar irradiance nowcasting case studies near sacramento. J. Appl. Meteorol. Climatol. 56, 85–108. https://doi.org/10.1175/JAMC-D-16-0183.1

Leese, J.A., Novak, C.S., Bruce B. Clark, 1971. An Automated Technique for obtaining Cloud Motion From Geosynchronous Satellite Data Using Cross Correlation. J. Appl. Meteorol. 10, 118–132. https://doi.org/10.1175/1520-0450(1972)011<0752:COATFO>2.0.CO;2

Lefèvre, M., Oumbe, A., Blanc, P., Espinar, B., Gschwind, B., Qu, Z., Wald, L., Schroedter-Homscheidt, M. Hoyer-Klick, C., Arola, A., Benedetti, A., Kaiser, J.W., Morcrette, J.-J., 2013. McClear : a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmos. Meas. Tech. 6, 2403–2418. https://doi.org/10.5194/amt-6-2403-2013

Leloux, J., Taylor, J., Moretón, R., Narvarte, L., Trebosc, D., Desportes, A., 2015. Monitoring 30,000 PV Systems in Europe: Performance, Faults, and State of the Art, in: 31st European Photovoltaic Solar Energy Conference and Exhibition2. Hamburg, pp. 1574–1582.

Lew, D., Bird, L., Milligan, M., Speer, B., Wang, X., Carlini, E.M., Estanqueiro, A., Flynn, D., Gomez-lazaro, E., Menemenlis, N., Orths, A., Pineda, I., Smith, J.C., Soder, L., Sorensen, P., 2013. Wind and Solar Curtailment. Int. Work. Large-Scale Integr. Wind Power Into Power Syst.

Lew, D., Brinkman, G., Ibanez, E., Hummon, M., Hodge, B., Heaney, M., King, J., 2012a. Sub-Hourly Impacts of High Solar Penetrations in the Western United States, in: 2nd Annual International Workshop on Integration of Solar Power into Power Systems Conference. Lisbon, Portugal, pp. 1–9.

Lew, D., Brinkman, G., Kumar, N., Besuner, P., Agan, D., Lefton, S., 2012b. Impacts of wind and solar on emissions and wear and tear of fossil-fueled generators. IEEE Power Energy Soc. Gen. Meet. 1–8. https://doi.org/10.1109/PESGM.2012.6343967

LI-COR, 2015. LI-200R Pyranometer Instruction Manual.

Li, Q., Lu, W., Yang, J., 2011. A hybrid thresholding algorithm for cloud detection on ground-based color images. J. Atmos. Ocean. Technol. 28, 1286–1296. https://doi.org/10.1175/JTECH-D-11-00009.1

Licciardi, G.A., Dambreville, R., Chanussot, J., Dubost, S., 2015. Spatiotemporal pattern recognition and nonlinear PCA for global horizontal irradiance forecasting. IEEE Geosci. Remote Sens. Lett. 12, 284–288. https://doi.org/10.1109/LGRS.2014.2335817

Linares-Rodriguez, A., Ruiz-Arias, J.A., Pozo-Vazquez, D., Tovar-Pescador, J., 2013. An artificial neural network ensemble model for estimating global solar radiation from Meteosat satellite images. Energy 61, 636–645. https://doi.org/10.1016/j.energy.2013.09.008

Lipperheide, M., Bosch, J.L., Kleissl, J., 2015. Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant. Sol. Energy 112, 232–238. https://doi.org/10.1016/j.solener.2014.11.013

Litjens, G.B.M.A., Worrell, E., van Sark, W.G.J.H.M., 2018. Assessment of forecasting methods on performance of photovoltaic-battery systems. Appl. Energy 221, 358–373. https://doi.org/10.1016/j.apenergy.2018.03.154

Lonij, V.P., Brooks, A.E., Koch, K., Cronin, A.D., 2012. Analysis of 80 rooftop PV systems in the Tucson, AZ area. Conf. Rec. IEEE Photovolt. Spec. Conf. 549–553. https://doi.org/10.1109/PVSC.2012.6317674

Lonij, V.P.A., Brooks, A.E., Cronin, A.D., Leuthold, M., Koch, K., 2013. Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors. Sol. Energy 97, 58–66. https://doi.org/10.1016/j.solener.2013.08.002

Lorenz, E., 2014. Where we are today with solar power forecasting, in: Final Workshop WIRE. Paris, France, p. 23 pp.

Lorenz, E., Heinemann, D., 2012. Comprehensive Renewable Energy, Elsevier. ed.

Lorenz, E., Kühnert, J., Heinemann, D., Nielsen, K.P., Remund, J., Müller, S.C., 2016. Comparison of global horizontal irradiance forecasts based on numerical weather prediction models with different spatio-temporal resolutions. Prog. Photovoltaics Res. Appl. 24, 1626–1640. https://doi.org/10.1002/pip.2799

Lorenzo, A.T., Holmgren, W.F., Cronin, A.D., 2015. Irradiance forecasts based on an irradiance monitoring network, cloud motion, and spatial averaging. Sol. Energy 122, 1158–1169. https://doi.org/10.1016/j.solener.2015.10.038

LSA-SAF, 2009. Product Requirement Document, Issue 1.11.

Macke, A., Seifert, P., Baars, H., Barthlott, C., Beekmans, C., Behrendt, A., Bohn, B., Brueck, M., Bühl, J., Crewell, S., Damian, T., Deneke, H., Düsing, S., Foth, A., Di Girolamo, P., Hammann, E., Heinze, R., Hirsikko, A., Kalisch, J., Kalthoff, N., Kinne, S., Kohler, M., Löhnert, U., Lakshmi Madhavan, B., Maurer, V., Kumar Muppa, S., Schween, J., Serikov, I., Siebert, H., Simmer, C., Späth, F., Steinke, S., Träumner, K., Trömel, S., Wehner, B., Wieser, A., Wulfmeyer, V., Xie, X., 2017. The HD(CP)2 Observational Prototype Experiment (HOPE) - An overview. Atmos. Chem. Phys. 17, 4887–4914. https://doi.org/10.5194/acp-17-4887-2017

Marion, B., 2017. Numerical method for angle-of-incidence correction factors for diffuse radiation incident photovoltaic modules. Sol. Energy 147, 344–348. https://doi.org/10.1016/j.solener.2017.03.027

Marion, B., Smith, B., 2017. Photovoltaic system derived data for determining the solar resource and for modeling the performance of other photovoltaic systems. Sol. Energy 147, 349–357. https://doi.org/10.1016/j.solener.2017.03.043

Marquez, R., Coimbra, C.F.M., 2012. Proposed Metric for Evaluation of Solar Forecasting Models. J. Sol. Energy Eng. 135, 011016. https://doi.org/10.1115/1.4007496

Martin, N., Ruiz, J.M., 2001. Calculation of the PV modules angular losses under field conditions by means of an analytical model. Sol. Energy Mater. Sol. Cells 70, 25–38. https://doi.org/10.1016/S0927-0248(00)00408-6

Martin, R., Aler, R., Valls, J.M., Galvan, I.M., 2016. Machine learning techniques for daily solar energy prediction and interpolation using numerical weather models. Concurr. Comput. Pract. Exp. 28, 1261–1274. https://doi.org/10.1002/cpe.3631

Masa-Bote, D., Castillo-Cagigal, M., Matallanas, E., Caamaño-Martín, E., Gutiérrez, A., 2014. Improving photovoltaics grid integration through short time forecasting and self-consumption. Appl. Energy 125, 103–113. https://doi.org/10.1016/j.apenergy.2014.03.045

Mathiesen, P., Brown, J.M., Kleissl, J., 2013a. Geostrophic wind dependent probabilistic irradiance forecasts for coastal california. IEEE Trans. Sustain. Energy 4, 510–518. https://doi.org/10.1109/TSTE.2012.2200704

Mathiesen, P., Collier, C., Kleissl, J., 2013b. A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance forecasting. Sol. Energy 92, 47–61. https://doi.org/10.1016/j.solener.2013.02.018

Mathiesen, P., Kleissl, J., 2011. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. Sol. Energy 85, 967–977. https://doi.org/10.1016/j.solener.2011.02.013

McCandless, T.C., Haupt, S.E., Young, G.S., 2016a. A regime-dependent artificial neural network technique for short-range solar irradiance forecasting. Renew. Energy 89, 351–359. https://doi.org/10.1016/j.renene.2015.12.030

McCandless, T.C., Young, G.S., Haupt, S.E., Hinkelman, L.M., 2016b. Regime-Dependent Short-Range Solar Irradiance Forecasting. J. Appl. Meteorol. Climatol. 55, 1599–1613. https://doi.org/10.1175/JAMC-D-15-0354.1

McGovern, A., Gagne, D.J., Basara, J., Hamill, T.M., Margolin, D., 2015. Solar energy prediction : An international contest to initiate interdisciplinary research on compelling meteorological problems. Bull. Am. Meteorol. Soc. 96, 1388–1393. https://doi.org/10.1175/BAMS-D-14-00006.1

Mellit, A., Massi Pavan, A., Lughi, V., 2014. Short-term forecasting of power production in a large-scale photovoltaic plant. Sol. Energy 105, 401–413. https://doi.org/10.1016/j.solener.2014.03.018

Miller, S.D., Heidinger, A.K., Sengupta, M., 2013. Physically Based Satellite Methods, in: Kleissl, J. (Ed.), Solar Energy Forecasting and Resource Assessment. Elsevier Academic Press, pp. 49–53.

Miller, S.D., Rogers, M.A., Haynes, J.M., Sengupta, M., Heidinger, A.K., 2018. Short-term solar irradiance forecasting via satellite/model coupling. Sol. Energy 168, 102–117. https://doi.org/10.1016/j.solener.2017.11.049

Milligan, M., Kirby, B., 2010. Market Characteristics for Efficient Integration of Variable Generation in the Western Interconnection. Golden, CO (United States). https://doi.org/10.2172/986251

Mills, A., Ahlstrom, M., Brower, M., Ellis, A., George, R., Hoff, T., Kroposki, B., Lenox, C., Miller, N., Milligan, M., Stein, J., Wan, Y.H., 2011. Dark shadows: Understanding variability and uncertainty of photovoltaics for integration with the electric power system. IEEE Power Energy Mag. 9, 33–41. https://doi.org/10.1109/MPE.2011.940575

MINES ParisTech, 2009. Meteosat spatial resolution [WWW Document]. URL http://www.soda-pro.com/help/general-knowledge/meteosat (accessed 7.13.19).

Morjaria, M., Chadliev, V., Gevorgian, V., 2017. Essential Reliability Services from Utility-Scale PV Power Plants, in: 7th Solar Integration Workshop. Berlin, pp. 24–25.

Nouri, B., Kuhn, P., Wilbert, S., Hanrieder, N., Prahl, C., Zarzalejo, L., Kazantzidis, A., Blanc, P., Pitz-Paal, R., 2019. Cloud height and tracking accuracy of three all sky imager systems for individual clouds. Sol. Energy 177, 213–228. https://doi.org/10.1016/j.solener.2018.10.079

Ogliari, E., Grimaccia, F., Leva, S., Mussetta, M., 2013. Hybrid predictive models for accurate forecasting in PV systems. Energies 6, 1918–1929. https://doi.org/10.3390/en6041918

Palensky, P., Dietrich, D., 2011. Demand side management: Demand response, intelligent energy systems, and smart loads. IEEE Trans. Ind. Informatics 7, 381–388. https://doi.org/10.1109/TII.2011.2158841

Palizban, O., Kauhaniemi, K., 2016. Energy storage systems in modern grids—Matrix of technologies and applications. J. Energy Storage 6, 248–259. https://doi.org/10.1016/j.est.2016.02.001

Parodi, A., Pulvirenti, L., Lagasio, M., Pierdicca, N., Marzano, F.S., Riva, C., Venuti, G., Pilosu, L., Realini, E., Passera, E., Rommen, B., 2018. Ingestion of Sentinel-derived remote sensing products in numerical weather prediction models: First results of the ESA steam project. Int. Geosci. Remote Sens. Symp. 2018-July, 3901–3904. https://doi.org/10.1109/IGARSS.2018.8518830

Pedro, H.T.C., Coimbra, C.F.M., 2015. Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances. Renew. Energy 80, 770–782. https://doi.org/10.1016/j.renene.2015.02.061

Pedro, H.T.C., Coimbra, C.F.M., 2012. Assessment of forecasting techniques for solar power production with no exogenous inputs. Sol. Energy 86, 2017–2028. https://doi.org/10.1016/j.solener.2012.04.004

Pedro, H.T.C., Coimbra, C.F.M., David, M., Lauret, P., 2018a. Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts. Renew. Energy 123, 191–203. https://doi.org/10.1016/j.renene.2018.02.006

Pedro, H.T.C., Lim, E., Coimbra, C.F.M., 2018b. A database infrastructure to implement real-time solar and wind power generation intra-hour forecasts. Renew. Energy 123, 513–525. https://doi.org/10.1016/j.renene.2018.02.043

Peng, Z., Yu, D., Huang, D., Heiser, J., Yoo, S., Kalb, P., 2015. 3D cloud detection and tracking system for solar forecast using multiple sky imagers. Sol. Energy 118, 496–519. https://doi.org/10.1016/j.solener.2015.05.037

Perez, R., Ineichen, P., Seals, R., Michalsky, J., Stewart, R., 1990. Modeling daylight availability and irradiance components from direct and global irradiance. Sol. Energy 44, 271–289. https://doi.org/10.1016/0038-092X(90)90055-H

Raza, M.Q., Nadarajah, M., Ekanayake, C., 2016. On recent advances in PV output power forecast. Sol. Energy 136, 125–144. https://doi.org/10.1016/j.solener.2016.06.073

Reikard, G., Haupt, S.E., Jensen, T., 2017. Forecasting ground-level irradiance over short horizons: Time series, meteorological, and time-varying parameter models. Renew. Energy 112, 474–485. https://doi.org/10.1016/j.renene.2017.05.019

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., Ratier, A., 2002. An introduction to METEOSAT Second Generation (MSG). Bull. Am. Meteorol. Soc. 83, 977–992. https://doi.org/10.1175/1520-0477(2002)083<0977:AITMSG>2.3.CO;2

Schmidt, T., Calais, M., Roy, E., Burton, A., Heinemann, D., Kilper, T., Carter, C., 2017. Short-term solar forecasting based on sky images to enable higher PV generation in remote electricity networks. Renew. Energy Environ. Sustain. 2, 23. https://doi.org/10.1051/rees/2017028

Schmidt, T., Kalisch, J., Lorenz, E., Heinemann, D., 2016. Evaluating the spatio-temporal performance of sky-imager-based solar irradiance analysis and forecasts. Atmos. Chem. Phys. 16, 3399–3412. https://doi.org/10.5194/acp-16-3399-2016

Sengupta, M., Andreas, A., 2010. Oahu Solar Measurement Grid (1-Year Archive): 1-Second Solar Irradiance; Oahu, Hawaii (Data); NREL Report No. DA-5500-56506.

Sengupta, M., Xie, Y., Lopez, A., Habte, A., Maclaurin, G., Shelby, J., 2018. The National Solar Radiation Data Base (NSRDB). Renew. Sustain. Energy Rev. 89, 51–60. https://doi.org/10.1016/j.rser.2018.03.003

Shields, J.E., Karr, M.E., Johnson, R.W., Burden, A.R., 2013. Day/night whole sky imagers for 24-h cloud and sky assessment: History and overview. Appl. Opt. 52, 1605–1616. https://doi.org/10.1364/AO.52.001605

Siddiqui, T.A., Bharadwaj, S., Kalyanaraman, S., 2019. A deep learning approach to solar-irradiance forecasting in sky-videos. Proc. - 2019 IEEE Winter Conf. Appl. Comput. Vision, WACV 2019 2166–2174. https://doi.org/10.1109/WACV.2019.00234

Sobri, S., Koohi-Kamali, S., Rahim, N.A., 2018. Solar photovoltaic generation forecasting methods: A review. Energy Convers. Manag. 156, 459–497. https://doi.org/10.1016/j.enconman.2017.11.019

Stull, R.B., 1988. An Introduction to Boundary Layer Meteorology. Klewer Academic Publishers, Dordrecht, The Netherlands.

Sweeney, C., Bessa, R.J., Browell, J., Pinson, P., 2019. The future of forecasting for renewable energy. Wiley Interdiscip. Rev. Energy Environ. 1–18. https://doi.org/10.1002/wene.365

Tarpley, J.D., 1979. Estimating incident solar radiation at the surface from geostationary satellite data. J. Appl. Meteorol. https://doi.org/10.1175/1520-0450(1979)018<1172:EISRAT>2.0.CO;2

Tomson, T., 2010. Fast dynamic processes of solar radiation. Sol. Energy 84, 318–323. https://doi.org/10.1016/j.solener.2009.11.013

Tomson, T., Tamm, G., 2006. Short-term variability of solar radiation. Sol. Energy 80, 600–606. https://doi.org/10.1016/j.solener.2005.03.009

Vallance, L., Charbonnier, B., Paul, N., Dubost, S., Blanc, P., 2017. Towards a standardized procedure to assess solar forecast accuracy: A new ramp and time alignment metric. Sol. Energy 150, 408–422. https://doi.org/10.1016/j.solener.2017.04.064

Vaz, A.G.R., Elsinga, B., van Sark, W.G.J.H.M., Brito, M.C., 2016. An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in Utrecht, the Netherlands. Renew. Energy 85, 631–641. https://doi.org/10.1016/j.renene.2015.06.061

Voyant, C., Notton, G., 2018. Solar irradiation nowcasting by stochastic persistence: A new parsimonious, simple and efficient forecasting tool. Renew. Sustain. Energy Rev. 92, 343–352. https://doi.org/10.1016/j.rser.2018.04.116

Voyant, C., Notton, G., Kalogirou, S., Nivet, M.L., Paoli, C., Motte, F., Fouilloy, A., 2017. Machine learning methods for solar radiation forecasting: A review. Renew. Energy 105, 569–582. https://doi.org/10.1016/j.renene.2016.12.095

Voyant, C., Soubdhan, T., Lauret, P., David, M., Muselli, M., 2015. Statistical parameters as a means to a priori assess the accuracy of solar forecasting models. Energy 90, 671–679. https://doi.org/10.1016/j.energy.2015.07.089

West, S.R., Rowe, D., Sayeef, S., Berry, A., 2014. Short-term irradiance forecasting using skycams: Motivation and development. Sol. Energy 110, 188–207. https://doi.org/10.1016/j.solener.2014.08.038

Wolff, B., Kühnert, J., Lorenz, E., Kramer, O., Heinemann, D., 2016a. Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data. Sol. Energy 135, 197–208. https://doi.org/10.1016/j.solener.2016.05.051

Wolff, B., Lorenz, E., Kramer, O., 2016b. Statistical learning for short-term photovoltaic power predictions. Stud. Comput. Intell. 645, 31–45. https://doi.org/10.1007/978-3-319-31858-5_3

Yadav, A.K., Chandel, S.S., 2014. Solar radiation prediction using Artificial Neural Network techniques: A review. Renew. Sustain. Energy Rev. 33, 772–781. https://doi.org/10.1016/j.rser.2013.08.055

Yang, D., 2019a. A guideline to solar forecasting research practice: Reproducible, operational, probabilistic or physically-based, ensemble, and skill (ROPES). J. Renew. Sustain. Energy 11. https://doi.org/10.1063/1.5087462

Yang, D., 2019b. A universal benchmarking method for probabilistic solar irradiance forecasting. Sol. Energy 184, 410–416. https://doi.org/10.1016/j.solener.2019.04.018

Yang, D., 2018. Ultra-fast preselection in lasso-type spatio-temporal solar forecasting problems. Sol. Energy 176, 788–796. https://doi.org/10.1016/j.solener.2018.08.041

Yang, D., Kleissl, J., Gueymard, C.A., Pedro, H.T.C., Coimbra, C.F.M., 2018. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. Sol. Energy 168, 60–101. https://doi.org/10.1016/j.solener.2017.11.023

Yang, D., Ye, Z., Lim, L.H.I., Dong, Z., 2015. Very short term irradiance forecasting using the lasso. Sol. Energy 114, 314–326. https://doi.org/10.1016/j.solener.2015.01.016

Yang, J., Min, Q., Lu, W., Ma, Y., Yao, W., Lu, T., Du, J., Liu, G., 2016. A total sky cloud detection method using real clear sky background. Atmos. Meas. Tech. 9, 587–597. https://doi.org/10.5194/amt-9-587-2016

Zack, J., 2017. Integration of Large-Scale Renewable Energy into Bulk Power Systems, Power Electronics and Power Systems. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-55581-2

Zagouras, A., Pedro, H.T.C., Coimbra, C.F.M., 2015. On the role of lagged exogenous variables and spatio-temporal correlations in improving the accuracy of solar forecasting methods. Renew. Energy 78, 203–218. https://doi.org/10.1016/j.renene.2014.12.071

Zamo, M., Mestre, O., Arbogast, P., Pannekoucke, O., 2014a. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: Deterministic forecast of hourly production. Sol. Energy 105, 792–803. https://doi.org/10.1016/j.solener.2013.12.006

Zamo, M., Mestre, O., Arbogast, P., Pannekoucke, O., 2014b. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production. Part II: Probabilistic forecast of daily production. Sol. Energy 105, 804–816. https://doi.org/10.1016/j.solener.2014.03.026

Zhang, J., Florita, A., Hodge, B.M., Lu, S., Hamann, H.F., Banunarayanan, V., Brockway, A.M., 2015. A suite of metrics for assessing the performance of solar power forecasting. Sol. Energy 111, 157–175. https://doi.org/10.1016/j.solener.2014.10.016