SOLAR IRRADIANCE FORECASTING AND MODELING FOR SMART CITIES WITH HIGH PHOTOVOLTAIC PENETRATION

ALOYSIUS WISHNU ARYAPUTERA

(B.Eng.(*Hons.*), NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF

ELECTRICAL AND COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2016

Declaration

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in this thesis.

This thesis has also not been submitted for any degree in any university previously.

Aloysius Wishnu Aryaputera

2016

Acknowledgments

I would like to acknowledge Prof. Armin Aberle and Dr. Wilfred Walsh for the opportunity to work in the Solar Energy Research Institute of Singapore (SERIS). I would also like to thank Dr. Yang Dazhi, Dr. André Nobre, and Dr. Nicholas Engerer (Australian National University) for the support. Last but not least, I would like to thank all the past and present colleagues in SERIS, especially those in the Solar Energy Systems cluster, and those who have worked on the data management of the SERIS meteorological network.

Abstract

Solar irradiance forecasting and modeling for smart cities with high photovoltaic penetration

Aloysius Wishnu Aryaputera

This thesis focuses on the development of solar irradiance forecasting methods. Spatio-temporal modeling and forecasting of irradiance will be essential requirements for the management of electricity grids in all cities with high photovoltaic (PV) penetration. Temporal variability of the harvestable solar energy on earth remains a great challenge facing widespread deployment of PV. To increase solar energy competitiveness, this work aims to provide tools to mitigate the problems caused by variability. Accurate near future forecasts and predictions of geographical smoothing effects of distributed systems over a certain area enhance the energy manageability, but the precise manner in which this occurs must be quantified. In this thesis, very short-term (≤ 5 minutes ahead) solar irradiance forecasting and modeling of irradiance geographical smoothing are performed using the anisotropic kriging method. Probabilistic forecasts on accumulated irradiance (6 and 12 hours ahead) are also generated using various ensemble methods. Finally, preliminary results of day-ahead irradiance forecast in Singapore using the Weather Research and Forecasting (WRF) model are presented.

List of publications

Journal articles

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

- A. W. Aryaputera, D. Yang, and W. M. Walsh, "Day-ahead solar irradiance forecasting in a tropical environment," *Journal of Solar Energy Engineering*, 137(5):051009, 2015. doi: `http://doi.org/10.1115/1.4030231`

- D. Yang, V. Sharma, Z. Ye, L. I. Lim, L. Zhao, and A. W. Aryaputera, "Forecasting of global horizontal irradiance by exponential smoothing, using decompositions," *Energy*, 81:111-119, 2015. doi: `http://doi.org/10.1016/j.energy.2014.11.082`

Conference presentations

- A. W. Aryaputera, N. Engerer, L. Zhao, and W. M. Walsh, "Output power forecast of a photovoltaic network," *26th International Photovoltaic Science and Engineering Conference*, Singapore, 2016 (poster presentation)

- A. W. Aryaputera, H. Verbois, and W. M. Walsh, "Probabilistic accumulated solar irradiance forecast for Singapore using ensemble techniques," 43rd IEEE Photovoltaic Specialists Conference, Portland, Oregon, US, 2016 (oral presentation)

- A. W. Aryaputera, D. Yang, L. Zhao, and W. M. Walsh, "Very short-term irradiance forecasting at unobserved locations using spatio-temporal kriging with polynomial anisotropy fitting," *25th International Photovoltaic Science and Engineering Conference*, Busan, Korea, 2015 (oral presentation)

- L. Zhao, A. W. Aryaputera, D. Yang, and W. M. Walsh, "Modeling of solar variability and geographic smoothing effect," *PV Asia Scientific Conference*, Singapore, 2015 (oral presentation)

- A. W. Aryaputera, L. Zhao, and W. M. Walsh, "Solar irradiance variation in Singapore," *PV Asia Scientific Conference*, Singapore, 2015 (poster presentation)

Contents

List of Figures

List of Tables

Acronyms

Roman Symbols

D	Kolmogorov-Smirnov test statistic.
d_n	day sequence in a year.
E_t	annual solar radiation.
f	frequency.
F_1	circumsolar brightening coefficient.
F_2	horizon brightening coefficient.
$\boldsymbol{h}$	spatial displacement between two points.
h_1	along wind component of $\boldsymbol{h}$.
h_2	cross wind component of $\boldsymbol{h}$.
$H\left[\cdot\right]$	Heaviside step function.
I_clr	modeled clear sky irradiance.
I_Dif	Diffuse Horizontal Irradiance.
I_Dir	Direct Normal Irradiance.
I_Glo	Global Horizontal Irradiance.
I_o	Extraterrestrial Direct Normal Irradiance.
I_sc	solar constant ($1362\ \mathrm{W/m^2}$).
I_t	global tilted irradiance.
m_lag	maximum time lag included in the kriging computation.
m_air	air mass.

r	Pearson product-moment correlation coefficient between the series of the absolute errors of deterministic forecast and the series of square root of the predictive PDFs standard deviations.
R_{d}	diffuse transposition factor.
r_{e}	earth eccentricity correction factor.
R_{r}	ground reflection transposition factor.
$SN\left(\xi, \mathfrak{w}^2, \alpha\right)$	skew-normal PDF with slant parameter α, scale parameter $\mathfrak{w}$, and location parameter ξ.

Greek Symbols

α	slant parameter of a skew-normal distribution.
β	panel azimuth angle.
δ	Kronecker delta function.
η	nugget effect.
$\Gamma\left(\cdot\right)$	gamma function.
κ	sky brightness.
μ_X	mean of variable X.
∇^x	xth order difference operator.
ω	panel tilt angle.
$\phi\left(\mathcal{B}\right)$	autoregressive operator.
$\Phi\left(\cdot\right)$	CDF of normal distribution $N\left(0,1\right)$.
$\rho\left(X,Y\right)$	fitted correlation function between variables X and Y.
σ_X	standard deviation of variable X.
τ	length of training period for BMA and EMOS.
$\theta\left(\mathcal{B}\right)$	moving average operator.
θ_{inc}	incidence angle of sun rays on a tilted plane.
θ_{z}	solar zenith angle.

Ξ	day angle.
ξ	location parameter of a skew-normal distribution.

Other Symbols

ACM.WVM	WVM method with wavelet correlation function based on Arias-Castro et al. (2014).
AGG.L	a benchmark method to predict the distribution of ramp events over an area by taking the spatial average of irradiance data of the learning stations.
ARIMA.BU	collective power output forecasting method using ARIMA algorithm where the forecast values of the individual PV system are taken as that of the collective one.
ARIMA.IND	collective power output forecasting method using ARIMA algorithm where the collective power output time series is treated as an independent one.
$\mathcal{B}$	backshift operator.
BMA.n	BMA method which adopts normal PDF.
BMA.s	BMA method which adopts skew-normal PDF.
$\mathcal{C}(\boldsymbol{h}; u)$	fitted correlation function based on spatial displacement $\boldsymbol{h}$ and/or time lag u.
$\mathcal{C}_{\mathrm{LGR}}(\boldsymbol{h}; u)$	Lagragian term of the STAT fitted correlation function.
$\mathrm{cor}(\cdot)$	Pearson's (empirical) correlation function.
$\mathcal{C}_{\mathrm{S}}(\boldsymbol{h})$	purely spatial correlation function (based on spatial displacement $\boldsymbol{h}$).
$\mathcal{C}_{\mathrm{T}}(u)$	purely temporal correlation function (based on time lag u).
$\mathcal{C}_{\mathrm{WIND.POLY}}(\boldsymbol{h}; u)$	anisotropic term of the POLY fitted correlation function.
$\mathbb{E}(X)$	expectation value of variable X.

$\mathfrak{e}$	sky clearness index.
$\mathcal{E}$	error component of the ETS method.
EMOS.n	EMOS method which adopts normal PDF.
EMOS.s	EMOS method which adopts skew-normal PDF.
EMP	empirical class of correlation function.
frac$_{\text{PSD}}$	ratio between PSD of aggregated stations and that of single stations.
F.SYM	fully-symmetric-fitted class of correlation function.
$\mathcal{K}_\nu$	second kind modified Bessel function of the order ν.
k_{T}	clear sky index.
$k_{\text{T.daily}}$	daily clear sky index.
lasso.BU	collective power output forecasting method using lasso algorithm where the forecast values of the individual PV system are taken as that of the collective one.
lasso.IND	collective power output forecasting method using lasso algorithm where the collective power output time series is treated as an independent one.
LAVE.WVM	WVM method with wavelet correlation function based on Lave et al. (2013).
OK	Ordinary kriging.
Perez.WVM	WVM method with wavelet correlation function based on Perez et al. (2011).
Pers	smart persistence forecast algorithm.
POLY	polynomial-fitted class of correlation function.
POLY.A	polynomial-line-fitted class of correlation function.
POLY.B	polynomial-surface-fitted class of correlation function.
$\mathfrak{P}$	foreground albedo.

$\mathfrak{S}$	correlation matrix.
$\mathcal{S}$	seasonal component of the ETS method.
STAT	general-stationary-fitted class of correlation function.
$\mathcal{T}$	trend component of the ETS method.
UK	Universal kriging.
UK.SUNY	Universal kriging with incorporated SUNY data.
$\mathfrak{w}$	scale parameter of a skew-normal distribution.

Acronyms/Abbreviations

ACM	Anisotropic Correlation Model.
ACT	Australian Capital Territory.
AIC	Akaike information criterion.
AnEn	Analog Ensemble.
ARIMA	Auto-Regressive Integrated Moving Average.
ARM	Atmospheric Radiation Measurement.
ARW	Advanced Research WRF.
BCE	Bias-Corrected Ensemble.
BFGS	Broyden-Fletcher-Goldfarb-Shanno optimization algorithm.
L-BFGS-B	Limited-memory BFGS algorithm with bound constraints.
BMA	Bayesian Model Averaging.
BoM	Australia's Bureau of Meteorology.
CDF	Cumulative Density Function.
CL	cloudy.
CLIM	Climatology.
CMA	China Meteorological Administration.
CNB	Canberra.
CO	Colorado.

CPTEC	Centro de Previsão Tempo e Estudos Climáticos.
CRPS	Continuous Ranked Probability Score.
CS	clear sky.
DHI	Diffuse Horizontal Irradiance.
HI	Hawaii.
DNI	Direct Normal Irradiance.
DR	Demand Response.
EC	Environment Canada.
ECMWF	European Centre for Medium-Range Weather Forecasts.
EM	Expectation-maximization algorithm.
EMOS	Ensemble Model Output Statistics.
ES	Ensemble size of AnEn method.
ETS	Exponential Smoothing.
FS	Forecast Skill.
GFS	Global Forecast System.
GHI	Global Horizontal Irradiance.
GOES	Geostationary Operational Environmental Satellites.
JMA	Japan Meteorological Agency.
KMA	Korea Meteorological Administration.
KS	Kolmogorov-Smirnov.
lasso	Least Absolute Shrinkage and Selection Operator.
LH	latent heat.
LRSS	Lowry Range Solar Station.
LW	longwave.
MBE	Mean Bias Error.
rMBE	Relative Mean Bias Error.
MM5	Fifth-Generation Penn State/NCAR Mesoscale Model.

MSSL	Market Support Services Licensees.
NAM	North American Model.
NCAR	National Center for Atmospheric Research.
NCEP	National Centers for Environmental Prediction.
NEMS	National Electricity Market of Singapore.
NOAA	National Oceanic and Atmospheric Administration.
NREL	National Renewable Energy Laboratory.
NWP	Numerical Weather Prediction.
NWTC	National Wind Technology Center.
OA	overall.
OC	overcast.
PBL	planetary boundary layer.
PDF	Probability Density Function.
PIT	Probability Integral Transform.
PSD	Power Spectral Density.
PSI	Pollutant Standards Index.
PV	Photovoltaic.
Qv	moisture.
RF	Relative frequency in a PIT histogram.
RMSE	Root Mean Square Error.
rRMSE	Relative RMSE.
SB	Squared bias error of a PIT histogram.
SERIS	Solar Energy Research Institute of Singapore.
SG	Singapore.
SGP-OK	Southern Great Plains.
SGT	Singapore time.
SH	sensible heat.

SolarTAC	Solar Technology Acceleration Center.
SRRL	Solar Radiation Research Laboratory.
SUNY	State University of New York.
SW	shortwave.
T	temperature.
THORPEX	The Observing System Research and Predictability Experiment.
TIGGE	THORPEX Interactive Grand Global Ensemble.
TMY	Typical Meteorological Year.
UKMO	UK Met Office.
UTC	Coordinated Universal Time.
VTIF	Vehicle Testing and Integration Facility.
WRF	Weather Research and Forecasting.
WVM	Wavelet-based Variability Model.

Chapter 1

Introduction

The imperative to reduce the global carbon footprint incurred from the combustion of fossil fuels is driving a massive global deployment of renewable energy, such as wind, solar, biomass, and geothermal. However, such energy sources are inherently time variable (Miller et al., 2013) and this variability has significant consequences for large scale implementation. The temporal variability of renewable energy sources has been studied on various time scales, from seconds, days, months, to years. Intra-day variability is perhaps the most important timescale, and is usually described in terms of "ramp rate" events where the resource exhibits significant increases or decreases in magnitude in less than one hour. Intra-day ramp events present challenges to conventional electric power grid management and can cause disturbances to grid stability (Lave and Kleissl, 2010). On the other hand, longer-scale (inter-day) variability is of interest for generation and transmission scheduling (Martinez-Anido et al., 2016). Both intra- and inter-day variability can be mitigated using a variety of methods, amongst which generation forecasting is a cost-effective method (West et al., 2014). In the field of solar energy, solar forecasting at various timescales is required for grid integration purposes, especially in

locations with high photovolatic (PV) penetration[1] (GE, 2010). When the solar resource can be forecast accurately, spinning energy reserves can be minimized (Miller et al., 2013). Accurate forecasts can help to reduce dependency on large energy storage systems, which are not generally attractive from the economical point of view (McMahan et al., 2013). Mixed technology solutions which employ batteries, spinning reserves, and other short period storage capacity (Nottrott et al., 2013) may be optimal, but these solutions require locally tailored forecasting to facilitate operational grid management (Hanna et al., 2014).

Fig. 1.1 illustrates the effect of higher PV penetration on the severity of energy generation variability. It is generated using the measurement data of 24 irradiance sensors spread throughout the island of Singapore (section 1.5.1.1). To produce this plot, every 1 W/m^2 is directly converted into 1 W. The data used come from four days when sudden storm took place (Nobre et al., 2016). During these moments, adding more measurement data from several sensors increases the severity of the aggregated measurement (Aryaputera et al., 2016a). At other moments, when irradiance variability at several sites are caused mainly by transformations of small-scale clouds, the irradiance variability at one place can be cancelled out by those at other places. This effect is called geographic smoothing.

This thesis focuses on solar irradiance forecasting, with a strong focus on the application of forecasting methods to Singapore. Singapore is a tropical city-state currently highly dependent on fossil fuel power generation. To reduce greenhouse gas emissions and to improve energy security, Singapore is actively pursuing the only significant renewable energy resource available locally, namely solar energy (NCCS, 2016). Studies have indicated that up to 30% of total electricity generation could be provided by inland and floating PV systems (Luther and Reindl, 2013).

---

[1]PV grid penetration is not consistently defined, and what constitutes 'high' penetration is somewhat dependent on the nature of the grid in question, but figures of 15% or more are commonly quoted (Bank et al., 2013).
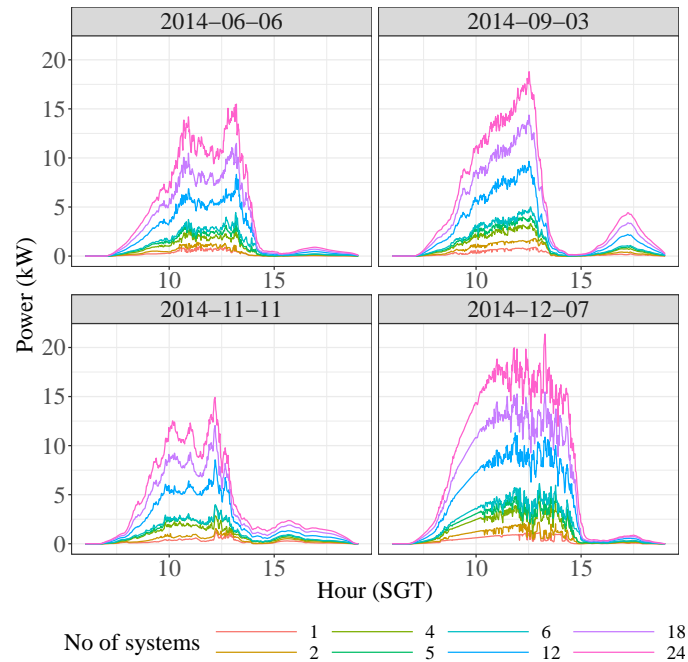
Fig. 1.1: Effect on PV penetration level on the severity of energy generation variability during sudden storm in Singapore.

Grid management issues will be particulalry acute in Singapore, owing to its small physical size, and the isolated nature of its electricity grid. However, the lessons learned in Singapore will be of direct relevance to any other urban area that implements high PV power generation in the coming decades. Given the ongoing decline in PV system installed costs, it is very likely that many Asian and tropical cities will face similar power grid challenges to those currently arising in Singapore.

Irradiance and PV system characteristics are the major factors that determine PV power generation, but temperature also influence the amount of electricity generated, since silicon cells' output decreases as temperature increases. However, since changes in irradiance are almost always much more extreme than those in temperature, irradiance has more influence on PV power generation (Yang, 2014), and thus irradiance becomes the central object of this thesis.

This introduction chapter covers the following:

- Section 1.2 reviews irradiance and forecast terminology and lists forecast time horizons.

- Sections 1.3 and 1.4 describe the general and specific contributions of this thesis.

- Section 1.5 lists the data sources utilized in this thesis.

- Section 1.6 explains the verification methods used in this thesis

- Section 1.7 describes the data processing tools employed.

1.1 Energy market in Singapore and lessons to learn from Germany

In this section, the energy market system in Singapore is briefly explained. Besides, some descriptions on the German electricity production are included. Germany is chosen since it is one of the leading countries which have successfully utilized significant amount of renewable energy for the electricity consumption. In 2015, 38% of the German net electricity consumption is covered by renewable energy. PV covered 7.5% of the net electricity consumption (Wirth, 2017).

1.1.1 The Singapore electricity plan

The National Electricity Market of Singapore (NEMS) was formed in 2003, with a vision: to create the effective supply of electricty with competitive price (Energy Market Authority, 2010). Today, there are two main groups of electricity consumers: the non-contestable and contestable ones. The former group purchases electricity from the SP Services Limited, as the Market Support Services Licensees

(MSSL), at the regulated tariff. The latter one may get electricity from the wholesale market directly, from the wholesale market indirectly through the MSSL, or from a retailer. Retailers may provide several pricing plans and services which are designed for different usage patterns (MyPower, 2017b). Contestable consumers who buy the electricity from the wholesale market directly, are charged based on real time prices which change every half-hour, based on the demand and supply. Currently, only commercial or industrial consumers with an average monthly electricity usage of equal to or larger than $2,000$ kWh are able to be in the contestable group (MyPower, 2017a).

In order to support the competition in the NEMS, the Energy Market Authority has created a Demand Response (DR) programme. This programme encourages the contestable consumers to cut off their electricity demand voluntarily, by giving them a share in the system-wide advantage from their actions (Energy Market Authority, 2016). The reduction of the electricity demand can be performed by running on-site back-up generators, energy storage systems, or using renewable energy. In Singapore, the main renewable energy option is solar energy (Bieri et al., 2016). In this context, short-term irradiance forecasting, as will be described in Section 1.2, may assist the decision makings of contestable consumers with private PV systems in the DR programme.

In the future, when there are more solar power plants in Singapore, from the operator point of view, accurate forecasts with longer horizons, such as day-ahead, will also be useful. It is because they can avoid over-commitment of conventional generators with "slow" downward ramping capabilities (Martinez-Anido et al., 2016).

1.1.2 The German power supply: lessons to learn

In 2016, the power generated from PV in Germany adds up to 38.3 TWh (Wirth, 2017). This comes from 1.5 million power plants whose total capacity reaches 41 GW. Compared to Singapore, Germany has several advantages as follows (Wirth, 2017):

- The PV systems in Germany are spread out throughout the country. Cloud cover changes which happen locally do not affect the total energy production of Germany as a whole.

- The generated solar energy in Germany is relatively predictable. This is in contrast with that in Singapore, which is located in the tropics. Tropical forecasts have been extremely challenging due to the local and mesoscale weather phenomena, and the sparse weather sensor network in the area (Laing and Evans, 2011).

- In Germany, wind and solar energy typically complement each other. Due to the particular climate condition, high wind speed is generally negatively correlated with high solar irradiance. In Singapore, however, there is no significant amount of hydro, wind, and geothermal resources (Bieri et al., 2016).

Therefore, the solar energy manageability in Singapore may be very challenging. To aim for significantly higher portion of renewable energy may require collaborations with the neighbouring countries whose renewable resources are more bountiful.

1.2 Solar irradiance, forecast horizons and irradiance modeling

The solar irradiance components which fall on the earth's surface are illustrated in Fig. 1.2. Irradiance received along the line of sight towards the sun is called "direct" while "diffuse" irradiance is that which comes from multiple scatterings and reflections within the atmosphere to produce an almost isotropic "diffuse" component. The sum of direct normal irradiance (DNI, I_{Dir}) and diffuse horizontal irradiance (DHI, I_{Dif}) on a horizontal plane is called global horizontal irradiance (GHI, I_{Glo}). I_{Ref} in Fig. 1.2 symbolizes the ground-reflected irradiance received by a tilted plane. The summation of $I_{\mathrm{Dir.tilt}}$, $I_{\mathrm{Dif.tilt}}$, and I_{Ref} is called global tilted irradiance (I_{t}), which is the total irradiance received by a tilted plane (Duffie and Beckman, 2013). In this thesis, the word "irradiance" refers to the GHI, unless stated otherwise.
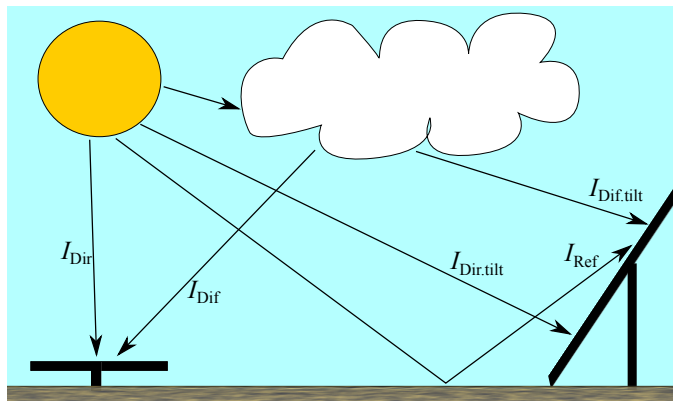
Fig. 1.2: Visualization of diffuse, direct and reflected irradiance on horizontal and tilted surfaces at the earth's surface, see text.

Solar irradiance is forecasted from several minutes up to several days in advance, depending on the application (Diagne et al., 2013; Kostylev and Pavlovski, 2011; Miller et al., 2013):

- Minutes to approximately one hour ahead forecasts are utilized by grid man-

agement operators to deal with high-frequency ramping events due to cloud movements, and also by real time market where electricity is sold on cycles as short as five minutes (Ela et al., 2014). Such high temporal resolution forecasting is, at present, always statistical in nature, and this kind of forecasting may also require high spatial resolution modeling.

- Hour to ∼6 hours ahead forecasts are utilized to deal with load following forecasting[2].

- Single day to ∼3 days ahead forecasts are utilized to deal with unit commitment[3] and day-ahead markets[4].

- Week to months ahead forecasts are utilized to deal with generator planning and asset optimization.

- One or more years ahead forecasts are utilized to deal with resource assessment, which may not be strictly considered as forecasting, but can involve broadly similar techniques. Resource assessment covers at least two aspects, namely the expected amount of solar energy and spatio-temporal variability in different timescales.

1.3   Thesis novelty

Working to provide solutions for better management of solar power variability, this thesis has contributed to the development of forecasting algorithms for three types of forecasts, namely very short-term (several seconds to 5 minutes ahead)

---

[2]Load following is the adjustment made within the day to follow the general pattern of electricity load of the day (Ela et al., 2011).

[3]Unit commitment problem is an optimization problem of generating unit scheduling constrained by the demand and grid operation (Ashraphijuo et al., 1999).

[4]Day-ahead market is an electricity market system where energy is sold one day ahead (Ela et al., 2014).

spatial forecasts, intraday (6 to 12 hours ahead), and day-ahead (± 24 hours ahead). There is a very large variety of forecasting methods available in the literature, but forecasting of solar irradiance in the tropics is a new field. There are very few published studies that are of direct relevance to the questions that must be answered by Singapore' grid management system. In particular, the topic of spatial irradiance forecasting is almost completely new, but such forecasting is required in order to quantify the effects of varying irradiance over a spatially distributed collection of PV generation systems across the grid.

The first contribution made by this thesis is an advancement to the existing methods describing how to accurately forecast irradiance at unobserved locations, by exploiting the spatio-temporal correlations of irradiance measurement of a sensor network. The second contribution utilizes ensemble techniques to produce probabilistic forecasts for accumulated irradiance. These ensemble methods have been widely used in meteorological forecasts, but not, until now, for irradiance forecasts. Lastly, this thesis provides results of day-ahead irradiance forecasting in Singapore, a tropical site, using the Weather and Research Forecasting (WRF) model.

A secondary contribution of this thesis is in the modeling the spatio-temporal variability of solar irradiance and its geographic smoothing. This second contribution is more useful for resource assessment.

1.4 Thesis chapters: overview and their specific contributions

Chapters 2-8 are the core chapters of this thesis. Chapters 2, 5, and 6 deal principally with resource assessment, while chapters 4, 7, and 8 describe the development of intra-hour- and several-hour-ahead forecasts.

1.4.1 Chapter 2: Singapore irradiance characteristics

This chapter presents the spatial and temporal irradiance variations in Singapore for various timescales (less than one hour) and the comparison with those in Colorado, Southern Great Plains (Oklahoma and Kansas), and Oahu (Hawaii). Fluctuations in solar irradiance on these timescales are mainly affected by formation, destruction and passings of clouds. Affected by these phenomena, extreme ramp events can occur, and PV systems may cause problems with the grid stability (Lave et al., 2012). Since fluctuations do not occur uniformly within a large area, the ramps of irradiance at different spatial locations may cancel each other (Hoff and Perez, 2012). This is called the geographic smoothing effect, a phenomenon which will be described further quantitatively in chapter 5. Understanding the spatial behaviour of irradiance for a region of interest (in this case Singapore) is important for meteorological station design and resource assessment, especially when the whole system is intended to be less dependent on base load power and more reliant on variable renewable generation.

Some parts of this chapter have been published in Aryaputera et al. (2015d).

1.4.2 Chapter 3: Short-term forecasts using a network of PV systems

This chapter exhibits preliminary work on 5-minute-ahead power output forecast of various PV systems in a network. Here, the least absolute shrinkage and selection operator (lasso) method is implemented to systematically select the historical observation values of adjacent PV systems to do forecast for a particular system.

The materials in this chapter has been submitted for publication (Aryaputera et al., 2016a).

1.4.3 Chapter 4: Very short-term forecasting using spatio-temporal kriging

Using spatio-temporal statistical methods, this chapter presents irradiance forecasting with very short forecast horizons (less than five minutes) using only irradiance data from a sensor network as input. This chapter shows novelty in this particular topic compared to existing studies since it exhibits forecasting at unobserved locations using leave-one-out cross-validation. The core method is adopted from Gneiting et al. (2007b) although the method is tailored to irradiance forecasting by way of implementation of fitted functions to represent anisotropy in correlation functions.

A sensor network in Oahu, Hawaii, is utilized for validation. The reason for the choice of this area is the consistent wind direction throughout the year (Hinkelman, 2013) and thus it is very suitable for the first step to develop the forecasting model.

The materials in this chapter have been published in Aryaputera et al. (2015a,c).

1.4.4 Chapter 5: Modeling variability and geographic smoothing

Moving a step ahead from chapters 2 and 4, this chapter presents modeling of solar variability and the geographic smoothing effect using spatio-temporal statistics. From the model presented, a quantitative prediction on the distribution of irradiance ramp rates in the presence of geographic smoothing is performed. Similarly to chapter 4, this chapter is inspired by the work of Gneiting et al. (2007b). The results of the proposed methods are compared with those of Lave et al. (2013); Arias-Castro et al. (2014); Perez et al. (2011); Monger et al. (2016).

Most materials in this chapter have been published in Zhao et al. (2015).

1.4.5 Chapter 6: Optimizing solar PV system orientations

This chapter consists of a work on the refinement of satellite image-based solar radiation and optimum photovoltaic orientation maps by incorporating ground measurement data. The kriging method is applied here.

1.4.6 Chapter 7: Numerical weather prediction applied to Singapore irradiance forecasts

This chapter presents a work using the Weather Research and Forecasting (WRF) model to forecast irradiance in Singapore. The WRF model takes in the Global Forecast System (GFS) data as boundary conditions, and produces more spatially- and temporally-refined forecasts. This chapter has been published in Aryaputera et al. (2015b).

1.4.7 Chapter 8: Ensemble techniques for probabilistic forecasts

In this chapter, intra-day accumulated solar irradiance forecasts for Singapore are reported. Here, spatially coarse outputs of some global numerical weather predictions (NWPs), including perturbed variants of each model, are blended using various methods and input combinations in order to produce the best predictive probability density function (PDFs) of the aggregated solar irradiance forecast in Singapore for 6 and 12-hours ahead. A predictive PDF gives more information than a deterministic forecast, and thus is more useful in practice (Sloughter et al., 2010). There are three main probabilistic ensemble methods which are tested here, namely Bayesian model averaging (BMA) (Raftery et al., 2005), ensemble model

output statistics (EMOS) (Gneiting et al., 2005), and analog ensemble (AnEn) (Delle Monache et al., 2013).

This chapter can be considered as the continuation of the work of Thorey et al. (2015), which used similar datasets, but only performed a deterministic forecast. Zamo et al. (2014) showed an analogous work, but only utilizes a set of perturbed forecasts from single NWP.

Some parts of this chapter have been presented in Aryaputera et al. (2016b).

1.5 Data

There are two types of data used in this thesis: observation data and NWP outputs. The former are obtained from six sources: Singapore (SG), Colorado (CO), Southern Great Plains (SGP-OK), Oahu, Hawaii (HI), Canberra (CNB), and the US National Solar Radiation Database (NSRDB), while the latter are gathered from The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE) (Bougeault et al., 2010).

1.5.1 Observation data sources

1.5.1.1 SG irradiance sensor network

The SG irradiance sensor network, maintained by the Solar Energy Research Institute of Singapore (SERIS), consists of 25 stations and it covers regions throughout the island of Singapore ($\sim$ 45 km $\times$ 25 km) (Kubis and Nobre, 2014). Fig. 1.3 shows the map of this network. Only 24 stations, however, are displayed in the figure since one of the stations do not have sufficient data in the period of interest used in this thesis. The data of this network are used in chapters 2, 8, and 7. Each station shown on the map has at least a silicon irradiance sensor which measures the GHI.
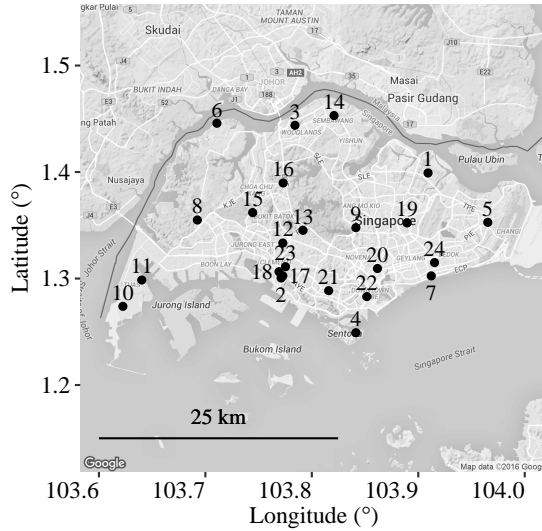
Fig. 1.3: Map of the SG solar irradiance sensor network. Map data source: Google Maps.

1.5.1.2 CO irradiance sensor network

The CO irradiance sensor network consists of five measurement stations maintained by the USA's National Renewable Energy Laboratory (NREL), located in an area of $\sim$ 47 km $\times$ 37 km. Fig. 1.4 shows the map of this network. The five stations are situated at the National Wind Technology Center (NWTC) M2 Tower (Jager and Andreas, 1996), the NREL Solar Radiation Research Laboratory (SRRL) (Andreas and Stoffel, 1981), the Vehicle Testing and Integration Facility (VTIF) (Lustbader and Andreas, 2012), the Solar Technology Acceleration Center (SolarTAC) (Andreas and Wilcox, 2011), and the Lowry Range Solar Station (LRSS) (National Renewable Energy Laboratory, 2014). Table 1.1 shows the details of each station. The data of this network are used in chapter 2.

Table 1.1: List of stations in the CO irradiance sensor network

Station	Longitude (°)	Latitude (°)	Elevation (m)
NWTC	-105.23	39.91	1855
SRRL	-105.18	39.74	1829
VTIF	-105.18	39.74	1793
SolarTAC	-104.62	39.76	1674
LRSS	-104.58	39.61	1860
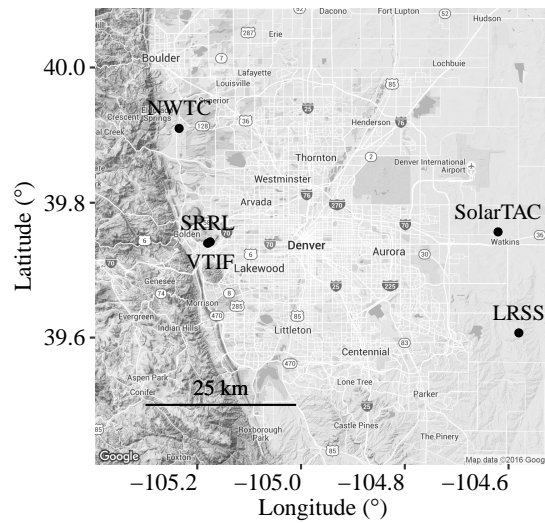
Fig. 1.4: Map of the CO solar irradiance sensor network. Map data source: Google Maps.

1.5.1.3 SGP-OK irradiance sensor network

The SGP-OK irradiance sensor network (Ackerman and Stokes, 2003) is located at the border of Kansas and Oklahoma states in the USA and is part of the United States Department of Energy's Atmospheric Radiation Measurement (ARM) programme. It covers an area of $\sim$ 140 km $\times$ 140 km and consists of 13 irradiance measurement sites where each site has a pyranometer as the GHI measurement device. Fig. 1.5 shows the map of this network and Table 1.2 shows the detail of each station. The data of this network are used in chapter 2.

Table 1.2: List of stations in the SGP-OK irradiance sensor network

Station	Longitude (°)	Latitude (°)	Elevation (m)
C1	-97.49	36.60	318.00
E11	-98.29	36.88	360.00
E12	-96.43	36.84	331.00
E15	-98.28	36.43	418.00
E31	-98.36	37.15	412.10
E32	-97.82	36.82	328.00
E33	-97.08	36.93	357.00
E34	-96.76	37.07	417.00
E35	-97.07	35.86	294.10
E36	-97.51	36.12	336.80
E37	-97.93	36.31	378.90
E38	-98.17	35.88	371.20
E9	-97.27	37.13	386.00
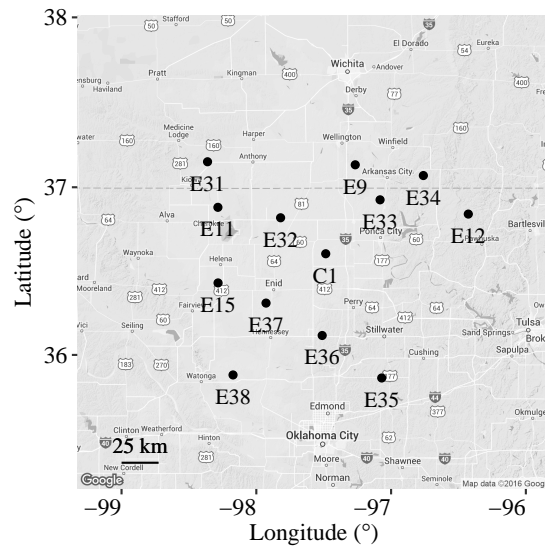
Fig. 1.5: Map of the SGP-OK solar irradiance sensor network. Map data source: Google Maps.

1.5.1.4　HI irradiance sensor network

The HI irradiance sensor network (Sengupta and Andreas, 2010) is located at the vicinity of the Honolulu International Airport (in an area of $\sim$ 1 km $\times$ 1 km) in the island of Oahu, Hawaii, USA and was managed by NREL until October 2011. Each station has a LICOR LI-200 pyranometer as the GHI measurement device which recorded data every second. The data obtained from this network are used in chapters 4 and 5.

This dataset has been used in previous works (Hinkelman, 2013; Arias-Castro et al., 2014; Yang et al., 2015b; Lave et al., 2015). An interesting property of this dataset is the existence of wind with persistent direction, as mentioned in section 1.4.3. The map of this network, including the dominant wind direction, is shown in Fig. 1.6.
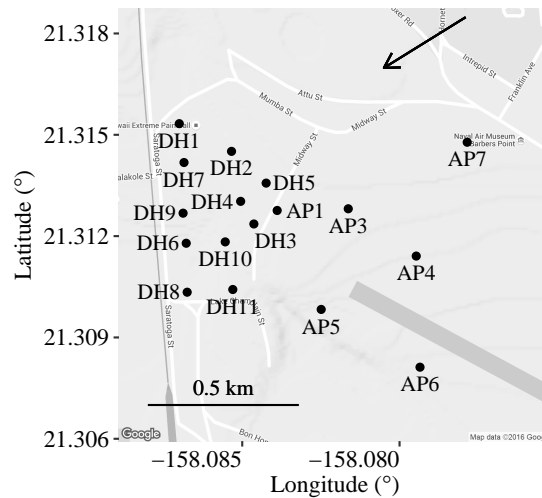
Fig. 1.6: Map of the HI solar irradiance sensor network. Dominant wind direction is shown by the arrow. Map data source: Google Maps.

1.5.1.5 CNB PV system network

The CNB PV system network (Engerer and Hansard, 2015) is a network located in the Australian Capital Territory (ACT), Australia and it covers an area of $\sim$ 19.8 km $\times$ 35.9 km. The map of the network is shown in Fig. 1.7 while the histogram of the array ratings is displayed in Fig. 1.8. Each PV system has its own capacity, tilt angle, and orientation. The dataset contains each system power output with 5-minute temporal resolution. Based on data availability, there are only 93 systems whose data are utilized in this thesis. The dataset is used in chapter 3 of this thesis.

1.5.1.6 NSRDB Typical Meteorological Year (TMY3) data

The third NSRDB Typical Meteorological Year (TMY3) dataset contains year-long (12-month-long) hourly-averaged solar irradiance and various meteorological parameters taken from monitoring stations in 1,020 locations in the continental US, Puerto Rico, Guam, Alaska, and Hawaii (Lubitz, 2011; Wilcox, 2012). The TMY3
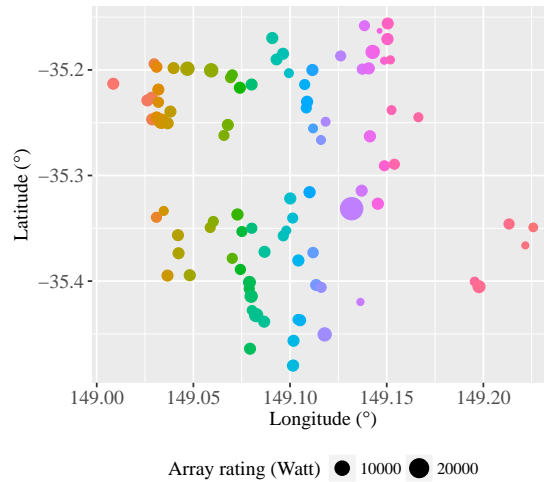
Fig. 1.7: Map of the CNB PV system network. The station colour coding is arranged based on the longitude. The same coding is used in Fig. 3.1.

dataset was derived from 1991 - 2005 data. For each month, data of a particular year which represents the best typical condition of that month were taken into the dataset. The dataset is used in chapter 6 of this thesis.

1.5.1.7 NSRDB State University of New York (SUNY) data

The NSRDB SUNY data were derived from satellite images, specifically those of the Geostationary Operational Environmental Satellites-East (GOES-East) and the GOES-West, using the model of SUNY, Albany (Perez et al., 2002; Wilcox, 2012). The SUNY data comprise of instantaneous (with resolution of one hour) and hourly-averaged Global Horizontal Irradiance (GHI), Direct Normal Irradiance (DNI), and Diffuse Horizontal Irradiance (DHI) from 1 January 1998 to 31 December 2005. Only hourly-averaged data are used in this work. The area covered is US land within 125ºW to 66ºW and 24ºN to 50ºN with 0.1º× 0.1º (10 km × 10 km) resolution. Originally, the GOES satellite image resolution is 1 km × 1 km. However, simplification was made to reduce the computational time of the SUNY model (Wilcox, 2012). Surface irradiation values were derived from calculated cloud in-
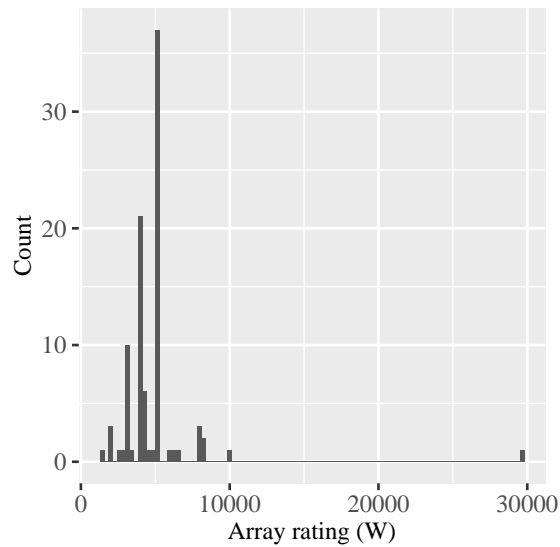
Fig. 1.8: Histogram of the array ratings of the CNB PV system network.

dex together with atmospheric turbidity, snow index, specular reflectans of ground, altitude, and sun-satellite angle (Lave and Kleissl, 2011).

The dataset is used in chapter 6 of this thesis.

1.5.2 TIGGE forecasting outputs

TIGGE (Bougeault et al., 2010) is a project which gathers various forecast outputs of global NWPs: the Australia's Bureau of Meteorology (BoM), China Meteorological Administration (CMA), the Brazil's Centro de Previsão Tempo e Estudos Climáticos (CPTEC), the Environment Canada (EC), the European Centre for Medium-Range Weather Forecasts (ECMWF), the Japan Meteorological Agency (JMA), the Korea Meteorological Administration (KMA), the Météo France, the United States's National Centers for Environmental Prediction (NCEP), and the UK Met Office (UKMO). Each NWP model has one control forecast output and several perturbed forecast members, containing forecasts of multiple meteorological variables, including, in the case of some NWPs, accumulated irradiance parameter. These data are available free-of-charge with a 2-day delay (Thorey et al., 2015).

In chapter 8 of this thesis, only the accumulated irradiance parameter is utilized. Furthermore, only three global models are used, namely the ECMWF, JMA, and KMA due to the data availability and literature review on forecast accuracies (Thorey et al., 2015).

1.6 Error metrics

1.6.1 Deterministic forecasts evaluation

In order to quantify forecast accuracies, standard error metrics are employed. For deterministic forecasts, root mean square error (RMSE) and relative RMSE (rRMSE) are utilized, where

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \left(\hat{I}_n - I_n \right)^2} \tag{1.1}$$

$$\text{rRMSE} = \frac{\text{RMSE}}{\frac{1}{N} \sum_{n=1}^{N} I_n} \times 100\% \tag{1.2}$$

where $\hat{I}_n$ represents the forecast irradiance, I_n is the actual irradiance, and N signifies the number of data in the series. When the RMSE value of one method is to be compared with that of a benchmarking method, forecast skill (FS) is employed

$$\text{FS} = 1 - \frac{\text{RMSE}_{\text{f}}}{\text{RMSE}_{\text{benchmark}}} \tag{1.3}$$

where RMSE_{f} is the RMSE of a forecasting technique and $\text{RMSE}_{\text{benchmark}}$ is that of the benchmarking method. Eq. (1.3) shows that FS has a maximum value of 1 when the forecasting technique is a perfect one. Negative FS value indicates that the benchmarking method performs better than the proposed forecasting technique.

1.6.2 Ramp rate cumulative density functions (CDFs) evaluation

In chapter 5, the simulated ramp rate cumulative density functions (CDFs) are validated against the measured ones using the two-sample Kolmogorov-Smirnov (KS) test (Arnold and Emerson, 2011). The KS statistic (D) signifies the maximum distance of two CDFs being compared and mathematically can be written as follows

$$D_{n,n'} = \sup_{x} \left\{ \left| F_{a,n}\left(x\right) - F_{b,n'}\left(x\right) \right| \right\} \tag{1.4}$$

where F_a and F_b are two discrete CDFs being tested, n and n' are the discretized length of F_a and F_b respectively, and $\sup\{\cdot\}$ represents the supremum function.

1.6.3 Probabilistic forecasts evaluation

For probabilistic forecasts (chapter 8), continuous ranked probability score (CRPS) is utilized to quantify forecast accuracies, and squared bias (SB) (Raftery et al., 2005) is used to measure the calibratedness of a forecasting method

$$\text{CRPS} = \frac{1}{N} \sum_{i=1}^{N} \text{crps}\left(F_i, y_i\right) = \frac{1}{N} \sum_{i=1}^{N} \left[\int_{-\infty}^{\infty} \left[F_i\left(x\right) - H\left[x - y_i\right] \right]^2 dx \right] \tag{1.5}$$

$$\text{SB} = \frac{1}{N_{\text{B}}} \sum_{i=1}^{N_{\text{B}}} \left(\text{RF}_i - 1\right)^2 \tag{1.6}$$

where $F_i\left(\cdot\right)$ and y_i denote the predictive CDF and observation value respectively for the case i, $H\left[\cdot\right]$ represents the Heaviside step function, N_{B} signifies the number of bins in the probability integral transform (PIT) histogram, and RF_i is the i-th bin relative frequency (RF) in the PIT histogram. PIT is the predictive CDF value at the observation value (Raftery et al., 2005). A well-calibrated probabilistic forecasting method typically produces a uniform PIT histogram.

1.7 Simulation and visualization tools

Besides the WRF simulation in chapter 7, results in this thesis are produced using the R statistical software (R Core Team, 2015). R is a free application for statistical computing and graphics, having various abilities for data simulations and visualizations. It has been built by the R core team, together with a wide network of contributors, who are producing contributed packages, expanding the usability of R. These packages are downloadable at: `https://cran.r-project.org/web/packages/`.

The most frequently employed packages in this thesis are as follows. For plotting, two packages are used: `ggplot2` (Wickham, 2009) and `ggmap` (Kahle and Wickham, 2013). For calculation of solar positions, package `insol` (Corripio, 2014) is mainly employed. The other contributed packages used in this thesis are mentioned separately in the corresponding chapters.

Chapter 2

Singapore irradiance characteristics

2.1 Introduction and literature review

Fluctuations of solar irradiance on the earth surface, mainly due to cloud formation and movement and hydrological activities (at short- and medium-term), pose challenges for the stability of electric power systems. Spatial variations may mitigate unwanted temporal fluctuations as aggregating power production from several solar power plants spread over a region takes advantage of the geographical smoothing phenomenon. Various authors have reported studies on this topic. Lave and Kleissl (2010) conducted computational experiments on solar variability based on irradiance data from four sites in Colorado (CO). Hoff and Perez (2012) attempted to model irradiance correlations in terms of distance from satellite images in three locations in the United States (US): Southwest, Southern Great Plains (SGP-OK),

---

This chapter is partly based on: A. W. Aryaputera, L. Zhao, and W. M. Walsh, "Solar irradiance variation in Singapore," *PV Asia Scientific Conference*, Singapore, 2015 (poster presentation).

and Hawaii (HI). Marcos et al. (2012); van Haaren et al. (2014); Klima and Apt (2015) assessed the geographical smoothing effect using photovoltaic (PV) system data in Spain, US and Canada, and Gujarat respectively.

The aforementioned studies show different degrees of smoothing in different areas. Therefore, in this chapter, an empirical study is made of the irradiance variability in Singapore (SG), which may be representative of tropical regions. For comparison purposes, data from irradiance sensor networks in three other regions, namely CO, SGP-OK, and HI (refer to sections 1.5.1.2 - 1.5.1.4), are studied.

This chapter is arranged as follows. Section 2.1 presents short literature review and motivation of the work. Section 2.2 briefly describes the data used in this study. Section 2.3 explains the methods employed in this chapter, especially the frequency domain analysis using power spectral density (PSD). The results are discussed in section 2.4 and finally the chapter is closed with a conclusion in section 2.5.

2.2 Data

The data used in this chapter are taken from, as mentioned in section 2.1, four irradiance sensor networks which are explained in section 1.5.1. Table 2.1 shows more details on the data, including the 1-minute average global horizontal irradiance (GHI) values since they may be compared with ramp rate values. These averages are computed after excluding data with zenith angle $(\theta_z) > 80°$. It should be noted that every dataset has a length of approximately 1 year.

Table 2.1: Network details

Network name	1-min. average GHI (W/m²min)	Period
SG	403	2 January 2014 - 31 December 2014
CO	457	1 January 2013 - 31 December 2013
SGP-OK	427	9 May 2014 - 10 May 2015
HI	504	1 November 2010 - 31 October 2011

In this work, station 22 of the SG network (refer to Fig. 1.3) is ignored due to the occurence of sensor shading during the corresponding period of measurement. Station AP3 of the HI network (refer to Fig. 1.6) is excluded as well due to its incomplete data during the period mentioned in Table 2.1.

2.3   Methods

2.3.1   PSD computation

Temporal variability of solar irradiance can be characterized in the frequency domain by utilizing power spectral density (PSD) (Katzenstein et al., 2010). In this chapter, the PSD values are computed using the function `lsp` in the R package `lomb` (Ruf, 1999; R Core Team, 2015). It works by employing the Lomb periodogram which is appropriate or "safe" for data with uneven time intervals (Katzenstein et al., 2010). Nevertheless, all datasets mentioned in Table 2.1 have data completeness rate of more than 99%.

In order to minimize the noise of the plots in the frequency domain (especially in the high frequency region), the data are temporally partitioned and passed into the periodogram algorithm, then the results are averaged in all frequencies (Klima and Apt, 2015). Furthermore, those PSD values, of all possible combinations of a certain amount of aggregated sensors, are averaged as well. To enable this, the `lsp` function is slightly modified to guarantee a fixed set of scanned frequencies.

The number of data partitions is fixed at 24, resulting in a partition length of around 15 days. Due to the partitioning process, seasonality which takes place longer than 6 - 15 days will not be reflected in the PSD plots.

2.3.2 PSD fitting

For comparability, following Klima and Apt (2015), the PSD values are fitted according to

$$\text{PSD}\,(f) = \frac{A}{1 + Bf^\epsilon} \tag{2.1}$$

where f is the frequency value, A represents the PSD value at low frequencies, B corresponds to the interception at y-axis in log-log domain, and ϵ signifies the PSD slope at high frequencies. The degree of variability is represented by ϵ (Lave and Kleissl, 2010). More negative ϵ means less high frequency component in the irradiance time series.

The fittings of A, B, and ϵ are done in a certain manner which allows comparability between PSD plot of single irradiance sensor with those of aggregated sensors (Klima and Apt, 2015).

- Eq. (2.1) is fitted to the empirical PSD plot of single irradiance sensors in the log-log domain. Fitting is done only in the frequency region slower than 48 hours and faster than 6 hours in order to exclude the impulses occur at frequencies related to solar diurnal cycle. All the fittings in this chapter are done using the limited–memory Broyden–Fletcher–Goldfarb–Shanno algorithm with boundary conditions (L–BFGS–B) (Byrd et al., 1995) integrated in the function `optim` in R.

- Empirical PSD plots of aggregated multiple irradiance sensors are normalized by, first, multiplying all its values by $\frac{A_\text{single}}{A_\text{aggregated}}$. Then, fitting is done for the second time by forcing as much as possible that the fitted curves would cross that of the single sensor at $f = \frac{1}{24\,\text{hours}}$. This is done since we may assume that the variabilities of single and aggregated sensors located in the same geographical region are getting more similar to each other in the long time step (Hoff and Perez, 2012). Here, the frequency of 24 hours is used as a

benchmark although this procedure generally also causes the lower frequencies to have similar PSD values.

An example of a fitting of a PSD plot derived from time series of single irradiance sensors is displayed in Fig. 2.1.
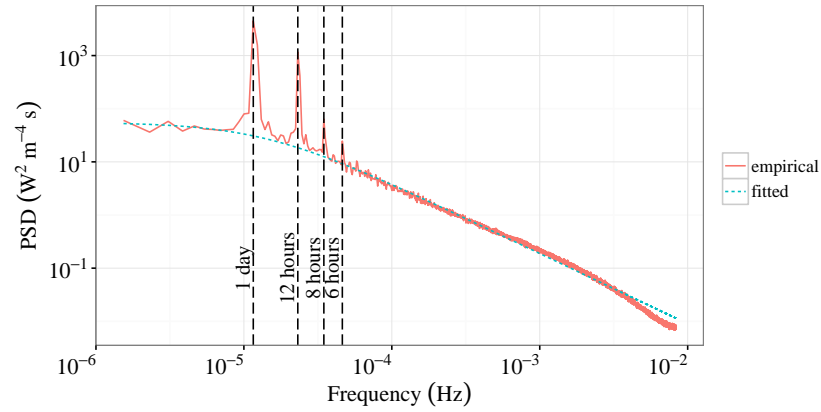
Fig. 2.1: PSD plot of irradiance from time series of single sensors in the SG network.

2.3.3 Computation of correlation and purely spatial correlation fitting

Apart from analysis of the SG data in the frequency domain, the correlations of irradiance ramps between two points in the four networks (SG, CO, SGP-OK, and HI) are investigated as well. Here, ramp means the difference between two consecutive irradiance values. Empirical correlations between two variables are computed using the Pearson's correlation coefficient formula

$$\text{cor}\,(X, Y) = \frac{\mathbb{E}\left[(X - \mu_X)\,(Y - \mu_Y)\right]}{\sigma_X \sigma_Y} \tag{2.2}$$

where $\text{cor}\,(X, Y)$ is the empirical correlation between variables X and Y, $\mathbb{E}$ represents the expectation, μ_X symbolizes the mean of variable X, and σ_X is the standard deviation of variable X.

Afterwards, following Gneiting (2002); Yang (2015), the relationship between empirical correlations and pair-station distances are fitted according to

$$\mathcal{C}\left(\boldsymbol{h}\right) = \frac{\left(1-\eta\right)\left(c||\boldsymbol{h}||\right)^{\nu}\mathcal{K}_{\nu}\left(c||\boldsymbol{h}||\right)}{2^{\nu-1}\Gamma\left(\nu\right)} + \eta\delta_{\boldsymbol{h}=0}, \quad 0 \le \eta \le 1, c > 0, \nu > 0 \qquad (2.3)$$

where $\mathcal{C}\left(\cdot\right)$ is the function of correlation with respect to spatial displacement $\boldsymbol{h}$, η is the nugget effect, $\mathcal{K}_{\nu}$ represents the second kind modified Bessel function of the order ν, $\Gamma\left(\cdot\right)$ is the gamma function, and c and ν are fitted parameters.

2.3.4   Clear sky model and index

Spatial correlation analysis described in section 2.3.3 can also be done on ramps of clear sky index. Clear sky index is a fraction between actual irradiance value and the corresponding modeled clear sky irradiance. Clear sky irradiance model (Yang et al., 2014b) used in this chapter is as follows

$$I_{\mathrm{clr}} = a r_{\mathrm{e}} I_{\mathrm{sc}} \left(\cos\theta_{\mathrm{z}}\right)^{b} e^{\Upsilon\left(90-\theta_{\mathrm{z}}\right)} \qquad (2.4)$$

$$r_{\mathrm{e}} = 1.00011 + 0.034221\cos\Xi + 0.001280\sin\Xi +$$
$$0.000719\cos 2\Xi + 0.000077\sin 2\Xi \qquad (2.5)$$

$$\Xi = \frac{2\pi\left(d_{\mathrm{n}}-1\right)}{365} \qquad (2.6)$$

where I_{clr} is the modeled clear sky irradiance, I_{sc} is the solar constant which is equal to 1362 W/m^2, r_{e} is the earth eccentricity correction factor, Ξ is day angle (in radians), d_{n} is day sequence in the corresponding year, and a, b, and Υ are fitted parameters.

For the SG network, the fitted parameters are taken from Yang et al. (2014b) while for the other three networks, the parameters are fitted using actual clear sky irradiance days, by considering only data with $\theta_{\mathrm{z}} < 80°$. The fitted parameters are written in Table 2.2.

Table 2.2: Clear sky model fitting details

Network	Parameters		
	a	b	Υ
SG (Yang et al., 2014b)	0.82980	1.35850	-0.00135
CO	0.93051	1.19228	-0.00209
SGP-OK	0.87510	1.28812	-0.00135
HI	0.89075	1.25756	-0.00186

2.4 Results and discussion

2.4.1 Geographical smoothing, PDF of ramp events, and general sky conditions

Fig. 2.2 shows the normalized empirical PSD plots of the irradiance sensors in the SG networks together with that of the clear sky model (Yang et al., 2014b). Here, the numbers of combinations of aggregated multiple sensors are capped at 100, and the combinations are chosen at random. The distances between stations are not taken into account. Nevertheless, the correlations between measurements of pair-stations and their relations with the corresponding inter-station distances are investigated in section 2.4.2.
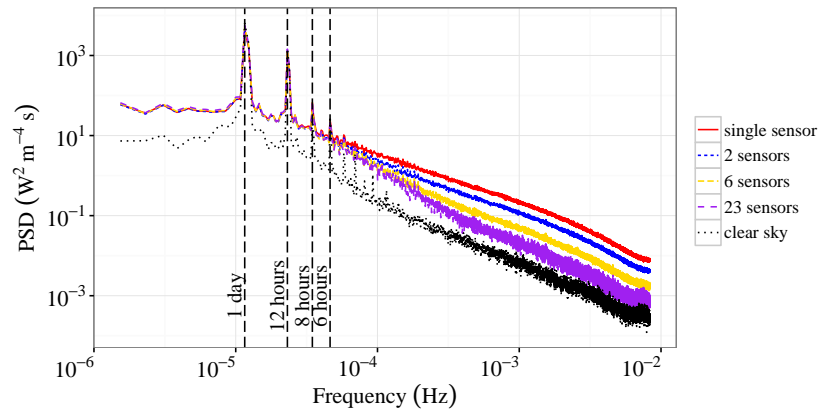
Fig. 2.2: Normalized PSD plots of irradiance from single and aggregated multiple sensors in the SG network. That of clear sky irradiance is included as well. The numbers of combinations of aggregated multiple sensors are capped at 100.

As expected, in Fig. 2.2, the clear sky model produces high PSD values only

at frequencies related to diurnal cycle. For the measured irradiance values, as mentioned in section 2.3.2, the variabilities at the lower frequency region are the same for irradiance values of single and aggregated multiple sensors. Due to geographic smoothing, high frequency components diminish as the number of aggregated sensors increases. As a consequence, the fitted PSD curves (in Fig. 2.3) tend to have steeper gradient in the high frequency region as there are more sensors aggregated together.
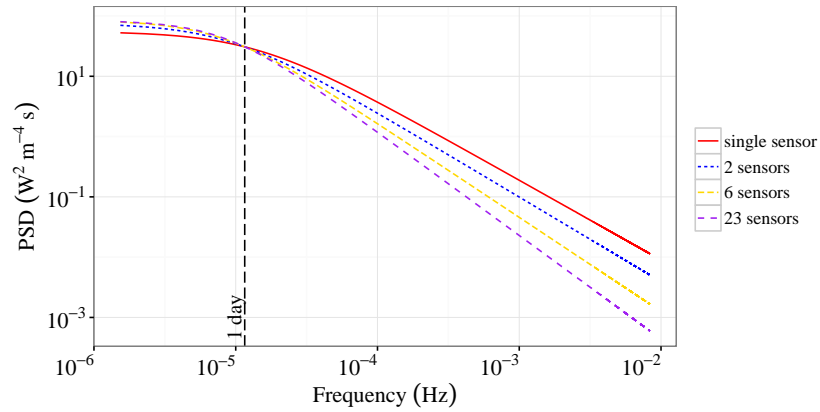
Fig. 2.3: The corresponding fitted plots of those of SG network in Fig. 2.2.
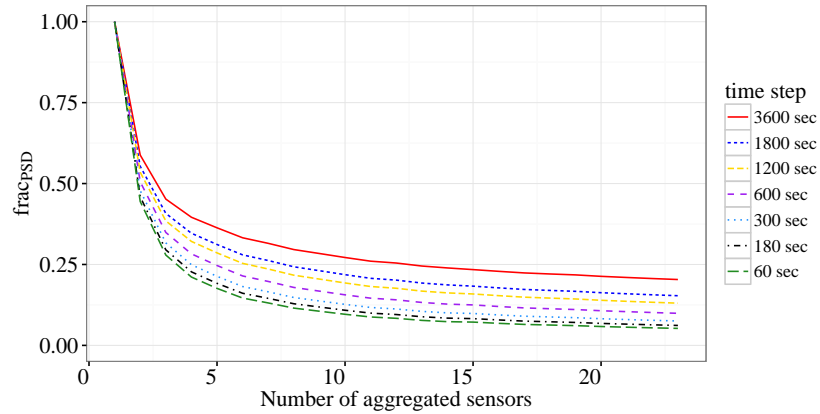
Fig. 2.4: The corresponding frac$_{\text{PSD}}$ of those of SG network in Fig. 2.3.

Following Klima and Apt (2015), the fitted PSD values of aggregated mul-
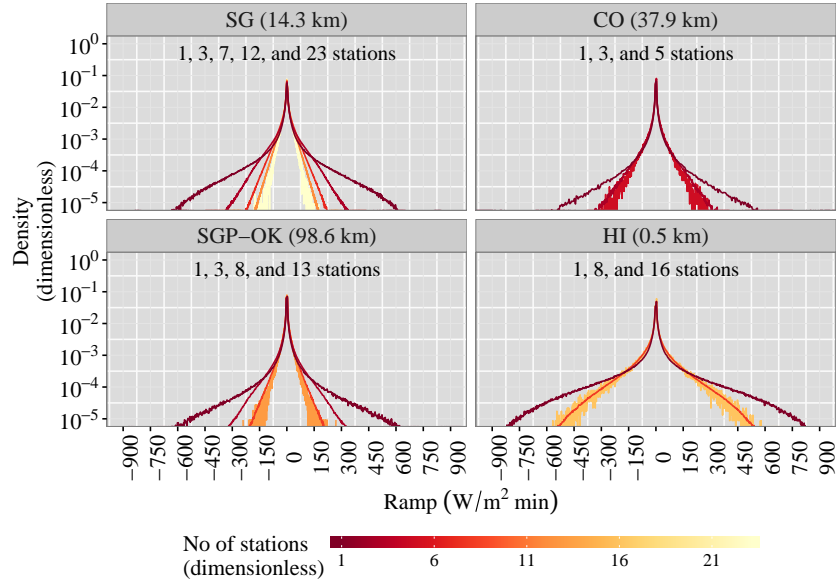
Fig. 2.5: Histogram of the ramp events at the SG, CO, SGP-OK, and HI networks. The numbers of station combinations are capped at 100. Average distances between stations are shown in the parentheses. The numbers of the aggregated stations are included in the plot and represented by the colours of the plot lines. Data with $\theta_z > 80°$ are excluded.

tiple stations can be compared to those of single sensors by computing

$$\text{frac}_{\text{PSD}} = \frac{\text{PSD}_{\text{multiple}}}{\text{PSD}_{\text{single}}} \qquad (2.7)$$

Plotting the frac_{PSD} with respect to the number of interconnected sensors in the SG network produces the graph in Fig. 2.4. It shows that the averaged irradiance values have much less fluctuations compared to that of single sensors when 2-10 sensors are interconnected. As the number of aggregated sensors grows, the increment of the geographical smoothing effects are not as effective as those of fewer interconnected stations.

The geographical smoothing effect can be visualized as well using the probability density function (PDF) of the ramp events as shown in Fig. 2.5. The figure reveals that a point in the SG network generally has more large fluctuations compared to those in the CO and SGP-OK networks. However, when the irradiance values from 23 sensors in the SG network are averaged, the resulting time series
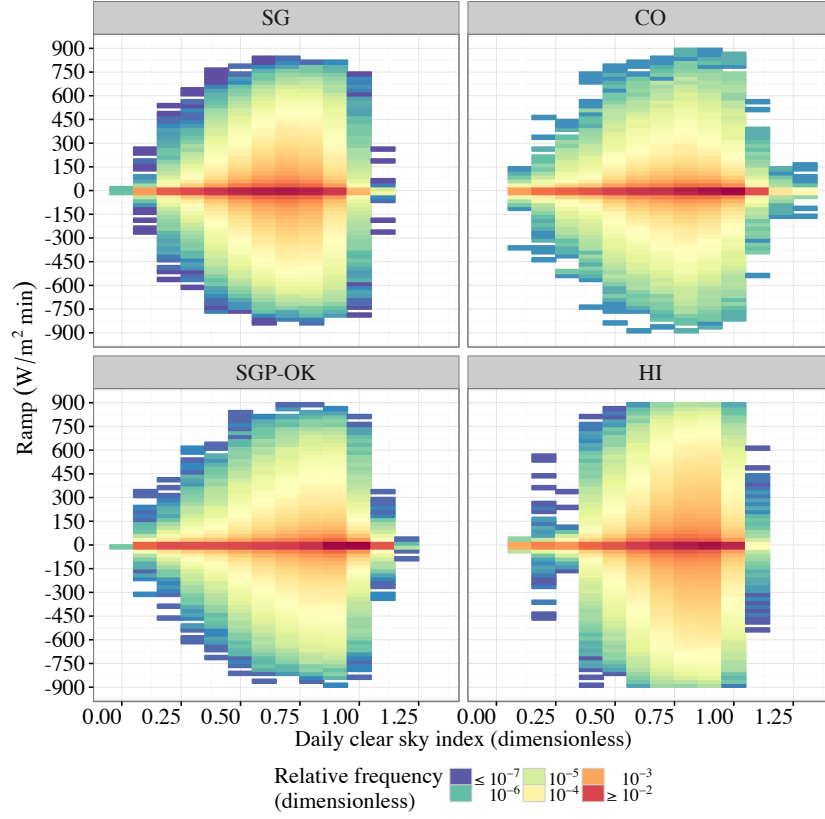
Fig. 2.6: 2-dimensional histogram of the ramp events at single sensors in the SG, HI, CO, and SGP-OK networks broken down according to daily clear sky indeces. Data with $\theta_z > 80°$ are excluded.

has significantly less fluctuations, even less than those of the averaged irradiance values from maximum number of sensors in the CO network. The irradiance values in the HI network seem to fluctuate the most. The fluctuations remain high even after the irradiance values from the 16 HI stations are averaged. This is related to the spatial closeness between stations in the HI network. The correlations between ramp events and inter-station distances will be discussed in section 2.4.2.

In agreement with Fig. 2.4, as the number of aggregated stations keeps increasing, the increment in the geographical smoothing effect saturates. For the SG network, it does not seem to be much difference in the PDF of the ramp events for 12 and 23 aggregated stations. Similar phenomenon can be observed in the other
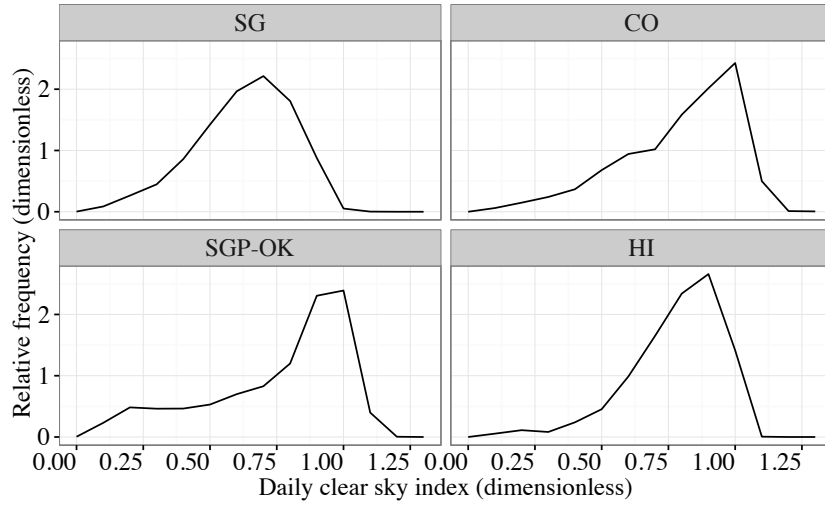
Fig. 2.7: Histogram of the daily clear sky indeces at the four networks. Data with $\theta_z > 80°$ are excluded.

networks as well.

Fig. 2.6 shows 2-dimensional PDF plots of the ramp events of single stations in the four networks broken down by the daily clear sky index. As expected, this plot reveals that the ranges of ramping events typically grow progressively as the daily clear sky indeces increase. Ultimately, they drop as the clear sky indeces get bigger than 1. Clear sky index may grow larger than unity due to the diffuse component of the irradiance (Yang, 2014). In line with Fig. 2.5, there are more large ramp events of single stations in the SG network than those in the CO and SGP-OK networks. In addition, the HI network is shown to possess many large ramp events of single stations for clear sky indeces between 0.6 - 0.9.

Fig. 2.7 visualizes the PDF of the clear sky indeces in the four networks. Stations in the SG network generally have much more cloudy days than the other three networks do. This explains the relatively low 1-minute average irradiance of the SG network in Table 2.1 although this network is located very near to the equator. It is worthwhile to note that during the period of the SG dataset (Table 2.1), based on the historical pollutant standards index (PSI) values, there was
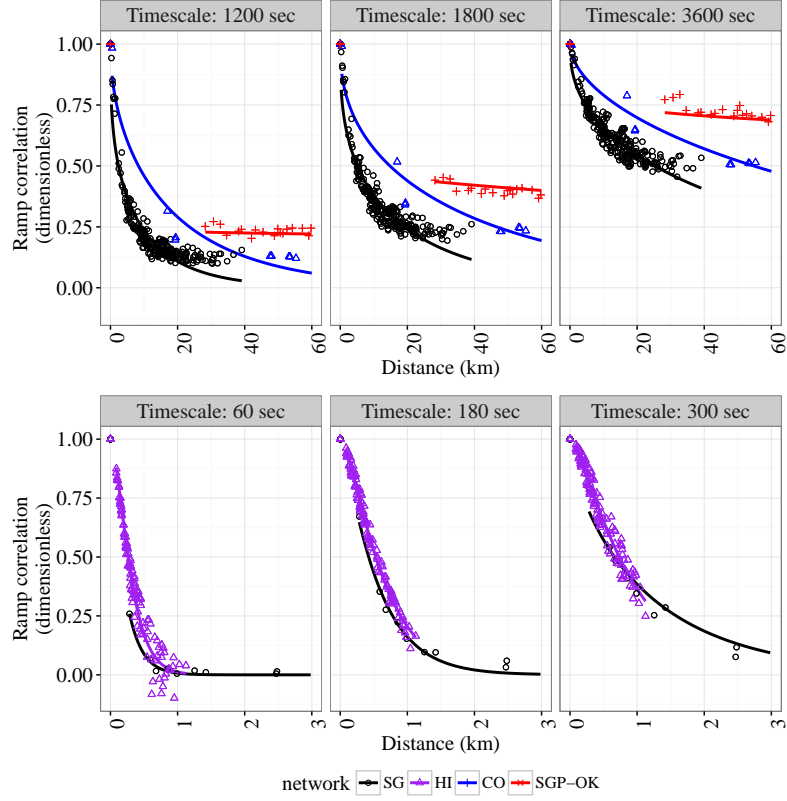
Fig. 2.8: Ramp correlations of station pairs with respect to their distances at the four networks (1-year data). Data with $\theta_z > 80°$ are excluded.

almost no significant smoke haze event, which is normally triggered by forest fires in Indonesia and affects Singapore almost every year (NEA, 2016; Asiaone, 2016). The CO and SGP-OK networks have the most clear sky days while the HI network has the fewest overcast days. This network also has more cloudy days compared to those of the CO and SGP-OK networks.

2.4.2 Spatial variations

Fig. 2.8 shows the purely spatial correlations of ramp events with respect to pair-station distances. The lines in the figure are fitted according to Eq. (2.3). It shows that in general the spatial variations in the SG network is higher than those in
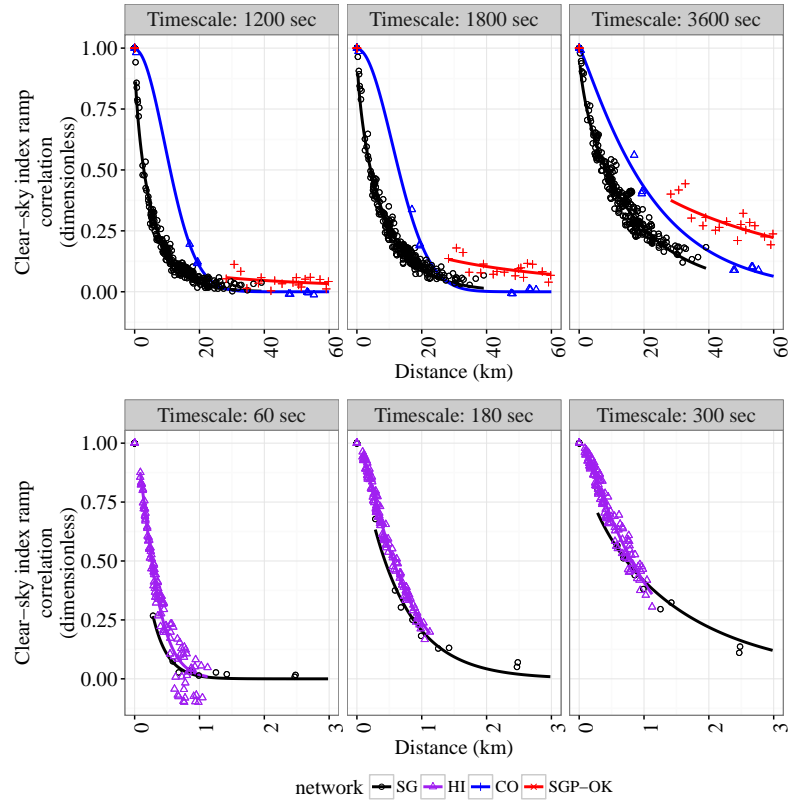
Fig. 2.9: As in Fig. 2.8, but for clear sky index. Data with $\theta_z > 80°$ are excluded.

the CO and SGP-OK networks, but comparable to those in the HI network. In agreement with Hoff and Perez (2012), the ramp correlations tend to get higher as the time step gets longer.

The fact that the SG network has high spatial variations may be related to Fig. 2.7 which shows that the SG network has many cloudy days and hardly any clear sky day. On top of that, weather phenomena in the tropical areas are typically dominated by local and mesoscale events while in midlatitude regions, synoptic phenomena are dominant (UCAR, 2011). Further investigations on other meteorological parameters such as wind speed and cloud types are required (Remund et al., 2015) in order to understand the spatial variations in different climates.

Following Hinkelman (2013), Figs. 2.10 and 2.11 show the ramp correlations

of station pairs in the SG network with respect to their distances according to various orientation degrees (shown on the top of the panels). These degrees show the rotation angles of the Singapore map. The distances of the station pairs are calculated based on the original horizontal axis. For smaller timescales, there is a nonuniformity in terms of spatial correlations for various orientation degrees. This phenomenon may be attributed to the dominant wind during this period, and it may be investigated further.

2.5   Conclusion

This chapter has shown that Singapore has higher spatial irradiance ramp variability compared to that in Colorado and Southern Great Plains. High spatial variability (low inter-station irradiance correlation) causes effective geographical smoothing, which is expected to make PV grid integration more manageable. However, the improvement of the geographical smoothing in Singapore becomes less obvious as more than ± 10 stations are interconnected.
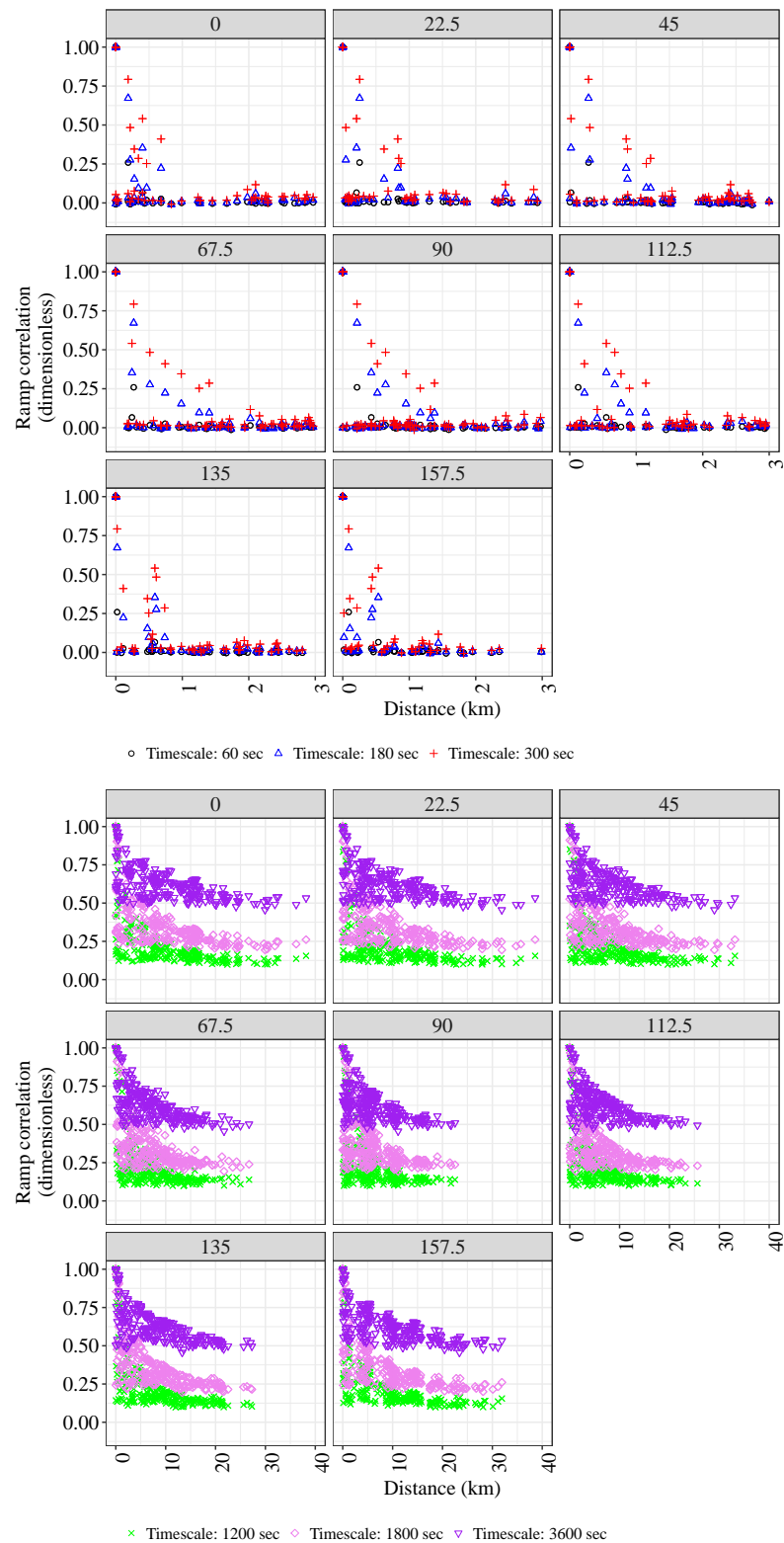
Fig. 2.10: Ramp correlations of station pairs in the SG network with respect to their distances according to various orientation degrees, and for various timescales. The number on the top of each panel shows the corresponding orientation degree.
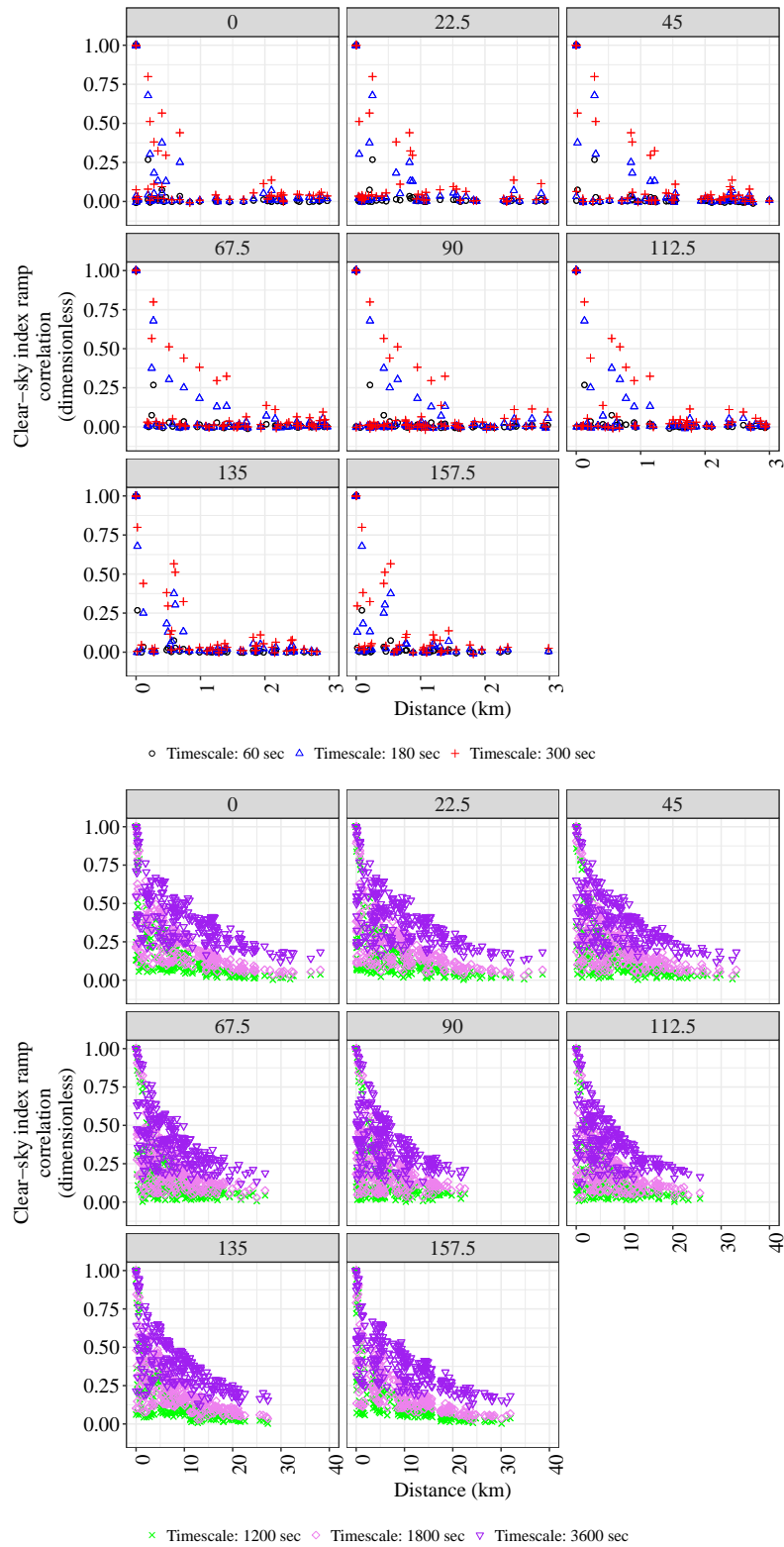
Fig. 2.11: As in Fig. 2.10, but for clear sky index.

Chapter 3

Short-term forecasts using a network of PV systems

3.1 Introduction

Photovoltaic (PV) power output forecasting techniques are critical tool in managing PV variability. These techniques vary widely in their application and methods, according to their spatial and temporal scales. For sub-hourly PV power or irradiance forecasting, most approaches have relied on statistical methods which utilize the recent historical values to make a short-term prediction, doing so for single sites. Recently however, several studies have attempted to expand forecasting to include multiple sites in their predictions. Since many studies now produce solar forecasts, baseline metrics have been established, and it has become common to compare new forecasting methods to pre-established methods. The most prominent baseline is that of smart persistence where it is assumed that the sky condition remains

unchanged between time-steps. One of the most commonly utilized models for single-site forecasting is the auto-regressive integrated moving average (ARIMA). Both of these methods are utilized as baselines for forecast improvements in this study.

For distributed PV forecast applications, sky cameras have been suggested as one promising approach (Yang et al., 2014c). Sky camera images may be utilized to anticipate coming clouds using temporal extrapolation upon cloud detection in the images. Satellite image analysis is another solution deployed for larger areas and longer timescales. However, universal coverage by sky-imagers is impractical, and satellite-derived forecasts often lack sufficient ground validation networks or may be limited in their resolution (in both space and time). In these cases, utilizing the historical data of neighbouring PV systems to forecast the output power of a given PV system is a promising approach.

Methods which utilize distributed networks of the power generation systems themselves to produce solar forecasts mean that it is possible to anticipate fluctuations in power output by sensing similar behaviour in the power output of adjacent PV systems. In this chapter, we apply this approach to a network of 93 PV systems, located in the Australian Capital Territory (ACT), Australia, to produce 5 min forecasts of single sites and collective power output via the least absolute shrinkage and selection operator (lasso) method (Efron et al., 2004; Tibshirani, 2011). The description of the network is written in section 1.5.1.5. The forecasting results are compared with smart persistence and ARIMA.

This work can be considered as a continuation of the previous work on the lasso method (Yang et al., 2015b), but it exchanges a distributed solar irradiance network for a solar PV one. This work is novel in that it is a first step in quantifying the accuracy of the lasso method for various meteorological events in the ACT which feature critical collective ramp events (Wellby and Engerer, 2016).

3.2   Methods and results

Before forecasts are performed, filtering is done to exclude data with solar zenith angle larger than 70°. Furthermore, only stations with data availability more than 95% are used. Missing data of the selected stations are filled using the most recent available data, or in the case of missing data at the start of a time series, arbitrary values are assigned.

ARIMA and lasso forecasts are performed by, first, training the models using the earliest 20% of the data (the "morning" part of the day). Subsequently, a moving training window is applied. This means that forecasts are only produced for the noon and afternoon parts of the day (latest 80% of the data). For each forecasting case, data from the previous day are not included in the learning period especially since the wind direction, which affects the movement of cloud or other particles in the sky, may be different from one day to the other. It brings a disadvantage since the "morning" part of the day cannot be forecast, and thus some days in Wellby and Engerer (2016) whose collective ramp events occur early in the morning, such as those triggered by fog phenomena, are excluded here. This problem may be improved in the future using technologies which enable the tracking of cloud or other particles at night until early in the morning before the sunrise.

We note that our data filtering and limited forecasting time period does not critically impact the value of the proposed power forecasting application: the bulk of PV power generation still falls within the forecast period. We also note that the data selected represent extreme events which can be the most difficult to forecast but which are of particular interest to grid operators.

Prior to the forecasting process, we normalize our data against the clear sky power output in order to remove any diurnal or seasonal tendency, and we transform the data back to un-normalized power output before the forecast accuracies are computed. The normalization process is done according to the method of Engerer

and Mills (2014).

Since we are particularly interested in the critical collective ramp events, two types of ARIMA and lasso methods are used to perform the accumulated power forecasting job (Hyndman et al., 2011). The first types are called independent ARIMA and lasso (ARIMA.IND and lasso.IND), where the collective power output values are forecast independently. The second types are called bottom-up ARIMA and lasso (ARIMA.BU and lasso.BU), where the forecast values of the individual PV system are taken as the forecast values of the collective one. Here the lasso method is implemented using the R package `lars` (Hastie and Efron, 2013) while the R package `forecast` (Hyndman, 2016; Hyndman and Khandakar, 2008) is adopted to realize the ARIMA method.
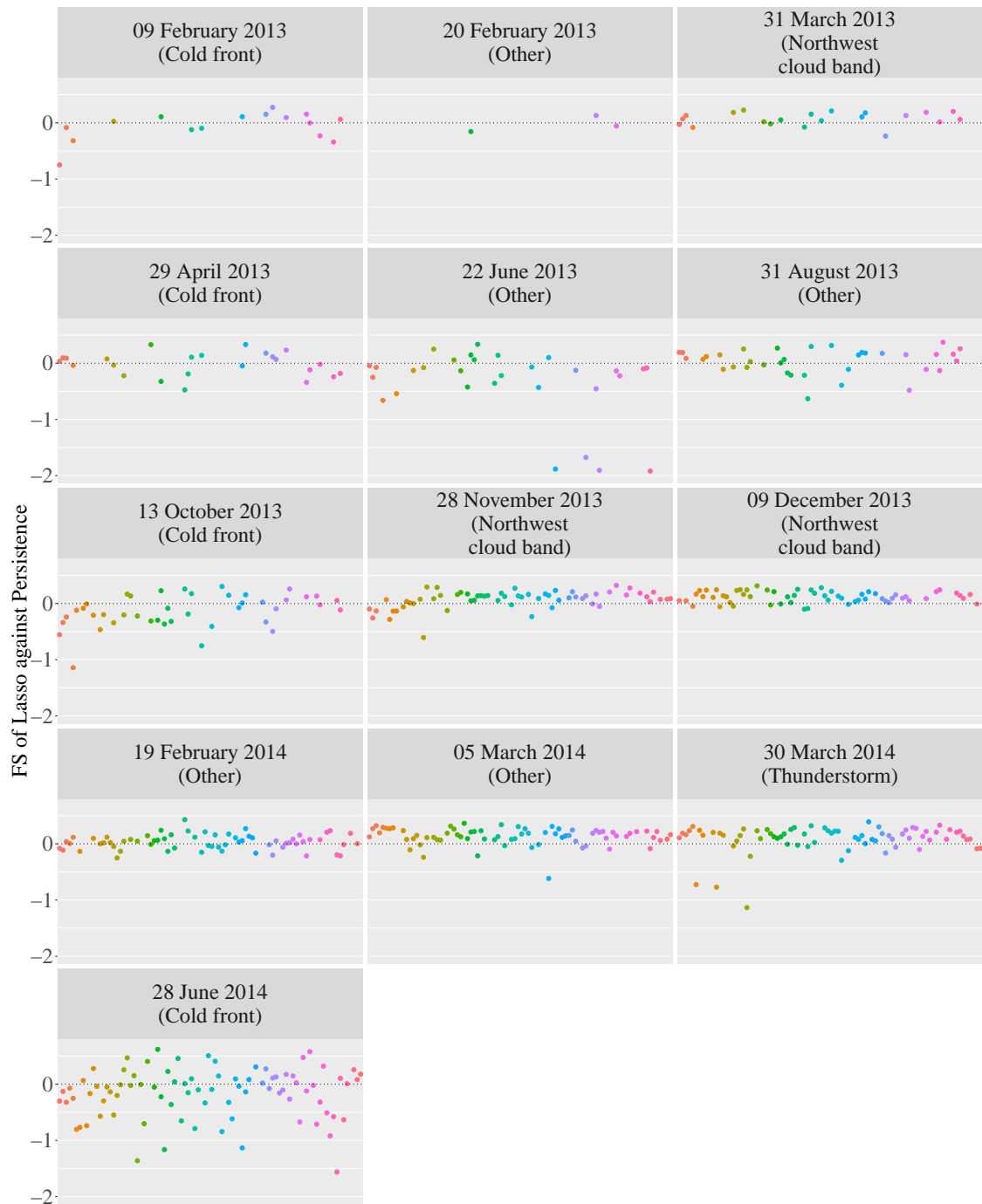
Fig. 3.1: FS plots of lasso against Pers method of individual systems for all the days with critical collective ramp events. The station colour coding is following that of Fig. 1.7
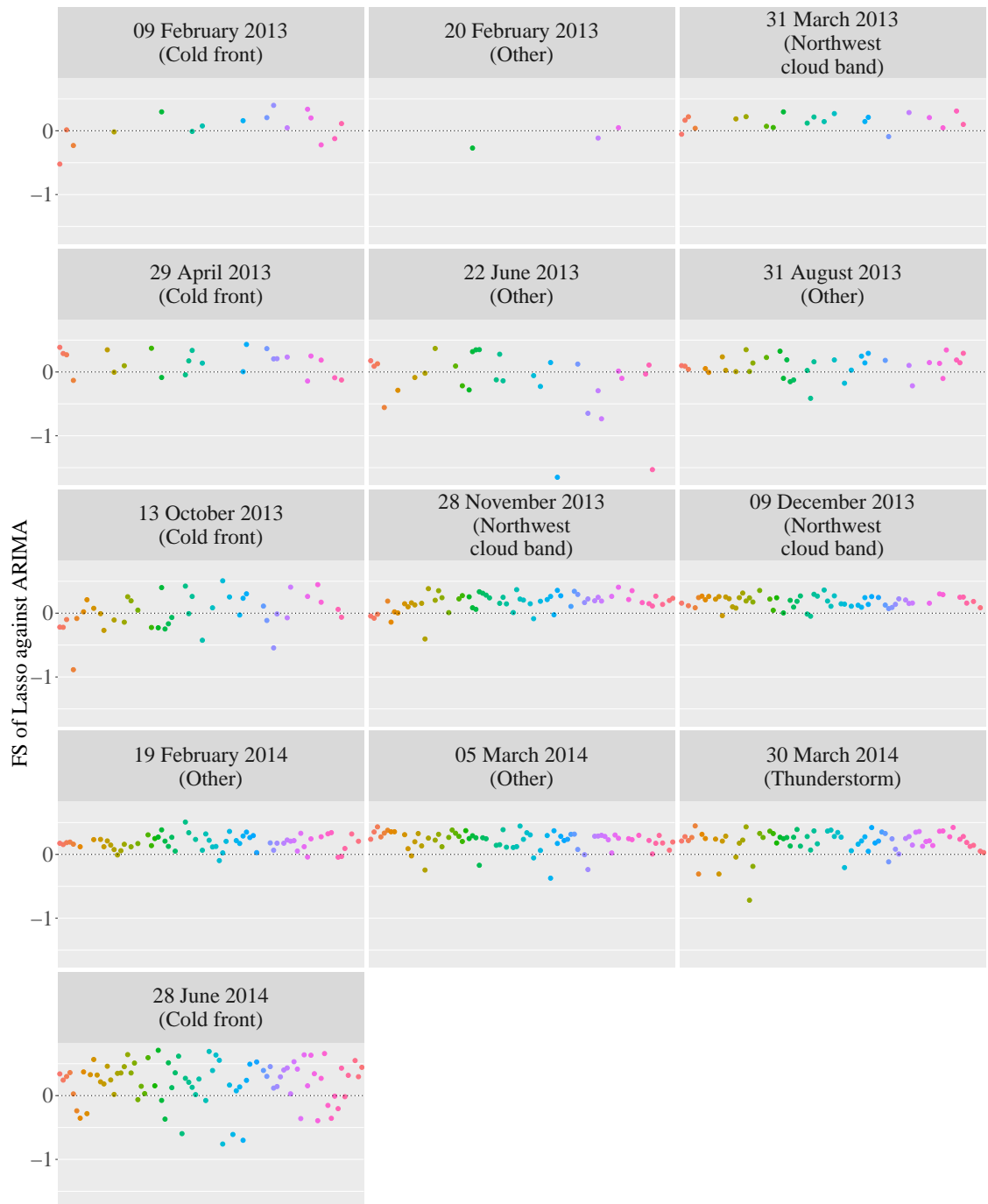
Fig. 3.2: As in Fig. 3.1, but for lasso against ARIMA.

Table 3.1: Forecast accuracies for collective outputs

Date	Event type	No of systems	Total capacity (kW)	Pers rRMSE (%)	FS				Best method	Best rRMSE (%)
					lasso.IND	ARIMA.IND	lasso.BU	ARIMA.BU		
09 February 2013	Cold front	16	95.200	9.9	0.19	-0.07	0.22	-0.02	lasso.BU	7.7
20 February 2013	Other	3	41.700	54.8	0.15	0.21	0.14	0.24	ARIMA.BU	41.7
31 March 2013	Northwest cloud band	21	121.200	13.1	0.08	-0.08	0.12	-0.27	lasso.BU	11.5
29 April 2013	Cold front	24	139.200	9.9	-0.23	-0.32	0.20	-0.24	lasso.BU	7.9
22 June 2013	Other	30	168.200	9.2	-1.25	-0.35	-0.61	-0.43	Pers	9.2
31 August 2013	Other	36	186.500	7.2	-0.04	0.07	0.14	-0.03	lasso.BU	6.2
13 October 2013	Cold front	42	225.080	6.6	-0.32	-0.14	-0.24	-0.26	Pers	6.6
28 November 2013	Northwest cloud band	61	301.682	10.7	-0.14	-0.39	0.25	-0.04	lasso.BU	8.0
09 December 2013	Northwest cloud band	61	306.182	14.5	0.10	0.05	0.15	-0.14	lasso.BU	12.3
19 February 2014	Other	65	331.082	11.0	0.03	-0.18	0.00	-0.43	lasso.IND	10.7
05 March 2014	Other	73	359.842	11.0	-0.25	0.04	0.16	-0.27	lasso.BU	9.2
30 March 2014	Thunderstorm	74	361.102	12.4	0.19	-0.05	0.14	-0.42	lasso.IND	10.1
28 June 2014	Cold front	84	400.982	25.6	-0.46	0.22	-0.09	-1.08	ARIMA.IND	19.9

The ramp event types selected for forecasting (following those reported in Wellby and Engerer (2016)).

The accuracies of all forecasts are then measured using forecast skill (FS)

$$FS = 1 - \frac{RMSE_f}{RMSE_{benchmark}} \qquad (3.1)$$

Table 3.1 shows the forecast accuracies for collective outputs with smart persistence method (Pers) becoming the benchmark for FS calculation. It can be seen that the lasso.BU method is the overall best method, and it is noted that lasso.BU method works particularly better than the others during northwest cloud band phenomena. Extremely inaccurate forecast results are obtained for 20 February 2013 due to the limited available data. Plots of the observed and forecast time series of the collective power outputs are available in appendix A.

Fig. 3.1 shows the FS of lasso and Pers of individual systems for all the days listed in Table 3.1. For completeness, Fig. 3.2 visualizes the comparison of lasso and ARIMA methods. In line with Table 3.1, Fig. 3.1 indicates that lasso method is able to do forecasts of individual systems power outputs more accurately when northwest cloud band phenomena occur. In contrast, during cold front phenomena, the forecast accuracies degrade. It is not reflected in Fig. 3.2 since the ARIMA method seems to be worst than the Pers method.

3.3 Conclusion

Five-minute-ahead forecasts of collective power outputs of a network of PV systems in the ACT, Australia, have been performed on days with critical collective ramp events. The results show that generally the bottom-up lasso method, where the power outputs of the individual systems are first forecast using lasso parameter selection, and then the individual forecast values are summed up as the forecast collective output, performs the best. This method performs particularly well on days with northwest cloud band phenomena. We further observe that the applied forecasting methods improve upon persistence and in all cases but one give forecast skill results comparable to other statistical methods of short time forecasting.

The particular significance of this result is that it has been demonstrated that acceptable forecasting can be performed using the output from PV power generation systems rather than relying upon an expensive network of irradiance monitoring stations. This advantage is of particular significance in situations where a relatively sudden increase in the number of small scale domestic PV systems occurs, potentially impacting grid operation. A suitably modified version of the forecasting technique demonstrated in this chapter offers the propsect of a low-cost, operational forecasting system, which may be deployed by as grid operator to facilitate short term grid stability.

Chapter 4

Very short-term forecasting using spatio-temporal kriging

4.1 Introduction

This chapter presents work on irradiance forecasting with forecast horizons ≤ 5 minutes, which we refer to as very short term forecasting. One potential application of such short-term forecasts is to anticipate extreme ramp events in large (MW-scale) utility PV systems (Engerer and Mills, 2014; Gould, 2013). Ancillary generators may be deployed if near–future negative ramp events are able to be forecast. Another potential application, as mentioned by Achleitner et al. (2014), is

---

This chapter is based on:

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

- A. W. Aryaputera, D. Yang, L. Zhao, and W. M. Walsh, "Very short-term irradiance forecasting at unobserved locations using spatio-temporal kriging with polynomial anisotropy fitting," *25th International Photovoltaic Science and Engineering Conference*, Busan, Korea, 2015 (oral presentation)

smoothing using ultra–capacitors.

There have been many previous approaches to very short term irradiance forecasting reported in the literature. Yang et al. (2014c) utilized a sky imager in San Diego, California for a forecasting system with forecast horizon up to 15 minutes. Although sky cameras provide detailed information about cloud position and movement, significant difficulties arise in creating a accurate models to forecast from them. Forecasting errors may be caused by inaccurate cloud base height approximation and assumptions of uniform cloud motion which do not always hold. In addition, occurrences of haze, which is often detected as thin clouds, leads to additional forecasting errors. For nowcasting, the relative root mean square error (rRMSE) is found to be 17.7%. The error increases quasi–linearly as forecast horizon gets longer.

Dambreville et al. (2014) used ground measurements and satellite–based HelioClim–3 data to forecast irradiance 15–minute to 1–hour ahead. Satellite images enable the use of spatio–temporal correlations. Their proposed method has forecast skill (FS) of 20.2%, 26.3%, 27.7%, and 29.3% for forecast horizons of 15, 30, 45, and 60 minutes respectively. Here, FS is defined as in Eq. (1.3) with persistence as the benchmarking method. Chu et al. (2015a) applied a multi–layer perceptron (MLP) approach and a genetic algorithm (GA) method to forecast solar irradiance ramps (with 1–minute resolution) 10 minutes in advance in San Diego and Folsom, California. The FS achieved are 10.1% and 6.2% for 10–minute horizons in San Diego and Folsom respectively. Chu et al. (2015b) tried to improve three existing forecast methods (cloud tracking (Chow et al., 2011), autoregressive integrated moving average (ARIMA) (Box et al., 2011), and k–nearest neighbour (Pedro and Coimbra, 2012) models) using GA. After GA had been implemented, for cloud tracking models (which improve results the most), the FS values increase from -71.3%, -30.7%, and -18.9% to 15.1%, 21.8%, and 26.2% for forecast horizons of 5, 10, and 15 minutes respectively.

Recently, investigating spatio–temporal correlations between measurements of solar irradiance sensors has become a research trend in the solar irradiance resource field. As an example, Yang et al. (2015b) performed very short–term forecasting by exploiting the spatio–temporal correlations in a dense network of irradiance sensors of Oahu, Hawaii for days dominated by broken clouds which had previously been investigated by Hinkelman (2013). In spatio–temporal–based forecasting, the irradiance value for the next time step is predicted based on the past and current irradiance values of the corresponding and neighbouring stations. The particular method used is the least absolute shrinkage and selection operator (lasso) (Efron et al., 2004; Tibshirani, 2011) which was shown to reach FS (averaged over all stations) of 19% in the case of 10–second ahead forecasting. This method is more accurate when compared with the standard ordinary least square (OLS) whose FS value is, for the same forecast horizon, 8% (in the case when wind direction is known). As the name suggests, the main difference between the lasso and OLS lies on the ability of lasso to eliminate insignificant predictors in order to diminish forecasting noise.

Similarly, Achleitner et al. (2014) designed solar power prediction system using a network of sensors built with the knowledge of dominant wind direction. The so-called Peak Matching Algorithm (PMA) implemented in this system is able to reach 97.24% accuracy for 6 testing days. Higher accuracy may be achieved compared to that of Yang et al. (2015b) owing to the fact that power output collected from a large solar field is temporally smoother than irradiance recording from a point sensor due to geographic smoothing effects (Lave et al., 2013).

Although lasso and PMA seem to be the state–of–the–art for very short–term irradiance forecasting using spatio–temporal correlations, they have a disadvantage since there is no way to predict irradiance at unobserved locations (where no historical data are available for these particular sites). In order to resolve this issue, this chapter presents a spatio–temporal forecasting using the kriging method which

includes a fitting of spatio–temporal correlation matrix (Cressie and Wikle, 2011; Gneiting et al., 2007b). Fitting is the key to enable prediction of solar irradiance at unobserved locations.

In this chapter, several spatio–temporal–based methods for very short–term surface irradiance forecasting (forecast horizons ≤ 5 minutes) are investigated. The forecast horizons are chosen in accordance with the size of the sensor network available. Once a network with larger area and higher density is available, there is a possibility to extend the forecast horizons (Engerer and Mills, 2014; Gould, 2013) and thus widen the usage of the forecasting results presented here.

This chapter is organized as follows. Section 4.2 explains the irradiance data from a station network on Oahu Island, Hawaii, provided by the National Renewable Energy Laboratory (NREL). Section 4.3 describes the kriging method. Section 4.4 displays the exploratory analyses and forecasting results. Section 4.5 presents the discussion based on the forecasting results. Finally, section 4.6 presents the conclusion of this chapter.

4.2   Irradiance data

As mentioned in section 4.1, the irradiance data used in this chapter is taken from the NREL station network in Oahu, Hawaii (Sengupta and Andreas, 2010). The layout of the network is shown in Fig. 1.6. For most of the stations, the global horizontal irradiance (GHI) data (with 1–second resolution) is available for 18 March 2010 to 31 November 2011. Dominant wind during this period, especially during 13 days mentioned by Hinkelman (2013), blows towards 60° West from South. In this chapter, since we are interested in integration of wind information into the forecasting system, only data from these 13 days, during which irradiance is also heavily influenced by broken clouds, is used. These days have also become the research subjects of Hinkelman (2013), Arias-Castro et al. (2014), and Yang

et al. (2015b).

As in the work of Yang et al. (2015b), the GHI datasets are averaged over intervals ($\bar{t}$), namely 10, 20, 30, 40, 50, 60, 120, 180, 240, and 300 seconds, to facilitate the forecasting with different forecast horizons. The smallest time aggregation (10–second) is chosen since the smallest along–wind distance for wind–parallel station pairs is 89 m (for AP1 and DH3) while the average wind speed during these 13 days is $\approx$ 10 m/s (Hinkelman, 2013). Thus, the average time needed for clouds to move from AP1 to DH3 is around 8.9 seconds.

In order to eliminate seasonal patterns in the GHI time series, the data are transformed into clear sky index (k_{T}) by dividing the data by calculated/modeled clear sky irradiance (I_{clr}). Then, to obtain zero mean values, the indices are subtracted by the mean of indices of each corresponding station. In the following sections, only such data with solar zenith angle smaller than 80° are utilized.

4.3   Kriging and spatio–temporal function fitting

Kriging is a method to predict the value of a certain variable in an unobserved location, given some point observation data distributed in a geographical space. This method was developed in 1950s by D. G. Krige, a mining engineer from South Africa (Krige, 1951; Matheron, 1963; Cressie, 1993). In the next step, this method was extended for spatio–temporal applications (Gneiting et al., 2007b; Cressie and Wikle, 2011), for example time–forward kriging (Yang et al., 2013), and thus making it as one of forecasting methods. The method used especially for cases which involve spatial processes with time series.

4.3.1 The kriging formulation

Consider a spatio–temporal process $\{z(\boldsymbol{s};t) : \boldsymbol{s} \in \mathcal{D}_s \subset \mathbb{R}^d, t \in \mathcal{D}_t \subset \mathbb{R}^1\}$ evolves through the spatio–temporal index set $\mathcal{D}_s \times \mathcal{D}_t$. Given observations:

$$\boldsymbol{Z} \equiv \left(z(\boldsymbol{s}_1;t_1), \cdots, z(\boldsymbol{s}_p;t_p)\right)^\top \tag{4.1}$$

at p space–time coordinates where $^\top$ stands for matrix transpose operation, predictions at unobserved coordinates $(\boldsymbol{s}_0;t_0)$ can be obtained using the simple kriging predictor:

$$z^*(\boldsymbol{s}_0;t_0) = \mu(\boldsymbol{s}_0;t_0) + \boldsymbol{c}(\boldsymbol{s}_0;t_0)^\top \boldsymbol{\mathfrak{S}}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}) \tag{4.2}$$

where $\boldsymbol{\mathfrak{S}} \equiv \mathtt{cor}(\boldsymbol{Z})$, $\boldsymbol{c}(\boldsymbol{s}_0;t_0) \equiv \mathtt{cor}(z(\boldsymbol{s}_0;t_0), \boldsymbol{Z})$, $\mathtt{cor}(\cdot)$ is the Pearson's (empirical) correlation function, and $\boldsymbol{\mu} \equiv \mathbb{E}(\boldsymbol{Z})$ (Cressie and Huang, 1999). When the number of temporal indices at each station is identical, we can decompose $\boldsymbol{Z}$ such that:

$$\boldsymbol{Z} = \left(\boldsymbol{z}^{(1)\top}, \boldsymbol{z}^{(2)\top}, \cdots, \boldsymbol{z}^{(m_{\text{lag}})\top}\right)^\top \tag{4.3}$$

and

$$\boldsymbol{z}^{(k)} = \left(z(\boldsymbol{s}_1;t_k), z(\boldsymbol{s}_2;t_k), \cdots, z(\boldsymbol{s}_n;t_k)\right)^\top, \quad k = 1, \cdots, m_{\text{lag}} \tag{4.4}$$

where n and m_{lag} are numbers of spatial and temporal indices, $nm_{\text{lag}} = p$. In other words, m_{lag} also represents the maximum time lag which is considered in the forecasting algorithm. For example, if m_{lag} is equal to three, observation values of the three newest time steps are considered in order to forecast observation value of the next time step.

The $\boldsymbol{\mathfrak{S}}$ in Eq. (4.2) is:

$$\boldsymbol{\mathfrak{S}} = \begin{pmatrix} \boldsymbol{\mathfrak{S}}_0 & \boldsymbol{\mathfrak{S}}_1^\top & \cdots & \boldsymbol{\mathfrak{S}}_{m_{\text{lag}}-1}^\top \\ \boldsymbol{\mathfrak{S}}_1 & \boldsymbol{\mathfrak{S}}_0 & \cdots & \boldsymbol{\mathfrak{S}}_{m_{\text{lag}}-2}^\top \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mathfrak{S}}_{m_{\text{lag}}-1} & \boldsymbol{\mathfrak{S}}_{m_{\text{lag}}-2} & \cdots & \boldsymbol{\mathfrak{S}}_0 \end{pmatrix}, \quad \in \mathbb{R}^{nm_{\text{lag}} \times nm_{\text{lag}}} \tag{4.5}$$

and

$$\mathfrak{S}_u = \begin{pmatrix} \mathfrak{S}_{11,u} & \mathfrak{S}_{12,u} & \cdots & \mathfrak{S}_{1n,u} \\ \mathfrak{S}_{21,u} & \mathfrak{S}_{22,u} & \cdots & \mathfrak{S}_{2n,u} \\ \vdots & \vdots & \ddots & \vdots \\ \mathfrak{S}_{N1,u} & \mathfrak{S}_{N2,u} & \cdots & \mathfrak{S}_{nn,u} \end{pmatrix}, \quad \in \mathbb{R}^{n \times n} \tag{4.6}$$

where $u = 0, \cdots, m_{\text{lag}} - 1$ and

$$\begin{aligned} \mathfrak{S}_{ij,u} &= \mathfrak{S}_{ij,k-l} \\ &= \text{cor}(z(\boldsymbol{s}_i; t_k), z(\boldsymbol{s}_j; t_l)) \\ &= \text{cor}(z(\boldsymbol{s}_i; t_k), z(\boldsymbol{s}_j; t_k - u)), \quad \forall k > l \end{aligned} \tag{4.7}$$

For one time step ahead prediction, the $\boldsymbol{c}$ in Eq. (4.2) can be expanded as:

$$\boldsymbol{c} = \begin{pmatrix} \boldsymbol{c}_{m_{\text{lag}}}^\top & \cdots & \boldsymbol{c}_2^\top & \boldsymbol{c}_1^\top \end{pmatrix}^\top, \quad \in \mathbb{R}^{nm_{\text{lag}} \times 1} \tag{4.8}$$

and

$$\boldsymbol{c}_u = \begin{pmatrix} c_{01,u} & c_{02,u} & \cdots & c_{0n,u} \end{pmatrix}^\top, \quad \in \mathbb{R}^{n \times 1} \tag{4.9}$$

where $u = m_{\text{lag}}, \cdots, 1$ and

$$c_{0j,u} = \text{cor}(z(\boldsymbol{s}_0; t_k), z(\boldsymbol{s}_j; t_k - u)) \tag{4.10}$$

Suppose there is a time series (which can be considered as a collection of sequential samples) at each spatio–temporal index, the empirical estimates of the correlations in Eqs. (4.7) and (4.10) can be calculated. However, when the goal is to estimate the irradiance at an unobserved location, correlations in Eq. (4.10) cannot be estimated directly due to the absence of data at the forecast location. In this case, fitted correlation functions are needed to estimate $\boldsymbol{c}$ in Eq. (4.8). A fitted correlation function is defined as:

$$\mathcal{C}(\boldsymbol{h}; u) \equiv \rho\left(z(\boldsymbol{s}; t), z(\boldsymbol{r}; q)\right) \tag{4.11}$$

where spatial difference $\boldsymbol{h} = \boldsymbol{s} - \boldsymbol{r}$ and temporal difference $u = t - q$, and $\rho\left(\cdot\right)$ symbolizes a fitted correlation function between two time series. It should be noted that in this chapter, the value of u is always related to $\bar{t}$. For example, $u = 0, 1, \cdots, 5$ refer to time lags of $0, 50, \cdots, 250$ seconds respectively for $\bar{t} = 50$ seconds.

4.3.2 Spatio–temporal function fitting

Besides the empirical one (EMP), four classes of fitted correlation functions are considered in this chapter, namely, separable (SEP), fully symmetric (F.SYM), general stationary (STAT), and polynomial (POLY). We say a correlation function is separable if:

$$\mathcal{C}_{\mathsf{SEP}}\left(\boldsymbol{h}; u\right) = \mathcal{C}_{\mathrm{S}}\left(\boldsymbol{h}\right)\mathcal{C}_{\mathrm{T}}\left(u\right) \tag{4.12}$$

where $\mathcal{C}_{\mathrm{S}}\left(\boldsymbol{h}\right)$ is a purely spatial correlation function, and $\mathcal{C}_{\mathrm{T}}\left(u\right)$ is a purely temporal correlation function. $\mathcal{C}_{\mathrm{S}}\left(\boldsymbol{h}\right)$ and $\mathcal{C}_{\mathrm{T}}\left(u\right)$ are shown in Eqs. (4.13) and (4.14) respectively

$$\mathcal{C}_{\mathrm{S}}\left(\boldsymbol{h}\right) = \left(1 - \eta\right)\exp\left(-c||\boldsymbol{h}||\right) + \eta\delta_{\boldsymbol{h}=0} \tag{4.13}$$

$$\mathcal{C}_{\mathrm{T}}\left(u\right) = \left(1 + a|u|^{2\vartheta}\right)^{-1} \tag{4.14}$$

where η, c, a, and ϑ are fitted parameters. Specifically, variable η is called the nugget effect. Eqs. (4.13) and (4.14) are monotone functions. Eq. (4.13) is a convex combination between an exponential function and a nugget effect. The nugget effect occurs due to variability in small spatial scale and/or measurement error. It causes discontinuities at zero distance. Eq. (4.14) is a Cauchy type correlation function with limited smoothness at small time lag which suits typical correlation–time lag relationship. The absence of the nugget effect is assumed since temporal aggregation, mentioned in section 4.2, tends to smooth out discontinuities.

Fully symmetric fitting is a fitting technique where spatial and temporal correlation functions are fitted simultaneously. Thus, the resulting function is generally

not separable. This function is shown in Eq. (4.15):

$$
\begin{aligned}
\mathcal{C}_{\text{F.SYM}}\left(\boldsymbol{h}; u\right) = {} & \frac{1 - \eta}{1 + a|u|^{2\vartheta}} \\
& \times \left[\exp\left(-\frac{c\|\boldsymbol{h}\|}{(1 + a|u|^{2\vartheta})^{\frac{\varrho}{2}}} \right) + \frac{\eta}{1 - \eta}\delta_{\boldsymbol{h}=0} \right]
\end{aligned}
\tag{4.15}
$$

where ϱ is the fitted space–time interaction. As mentioned by Gneiting et al. (2007b), there is a restriction of $\varrho \in [0, 1]$. The function will turn to be separable when $\varrho = 0$. Eq. (4.15) is a fully symmetric correlation function since it satisfies Eq. (4.16).

$$
\mathcal{C}\left(\boldsymbol{h}; u\right) = \mathcal{C}\left(\boldsymbol{h}; -u\right) = \mathcal{C}\left(-\boldsymbol{h}; u\right) = \mathcal{C}\left(-\boldsymbol{h}; -u\right)
\tag{4.16}
$$

General stationary fitting is similar with the symmetric fitting, except that it considers anisotropy. In the field of solar resource, anisotropy is typically caused by wind. General stationary correlation function can thus be formulated as:

$$
\mathcal{C}_{\text{STAT}}\left(\boldsymbol{h}; u\right) = (1 - \lambda)\,\mathcal{C}_{\text{F.SYM}}\left(\boldsymbol{h}; u\right) + \lambda \mathcal{C}_{\text{LGR}}\left(\boldsymbol{h}; u\right)
\tag{4.17}
$$

where λ is a fitted parameter and $\mathcal{C}_{\text{LGR}}\left(\boldsymbol{h}; u\right)$ is a function describing the anisotropy in the along wind direction. An example of $\mathcal{C}_{\text{LGR}}\left(\boldsymbol{h}; u\right)$, as proposed by Gneiting et al. (2007b) is:

$$
\mathcal{C}_{\text{LGR}}\left(\boldsymbol{h}; u\right) = \left(1 - \frac{1}{2v}|h_1 - vu| \right)_{+}
\tag{4.18}
$$

where v is a fitted parameter and h_1 is the along wind component of $\boldsymbol{h}$.

Finally, anisotropic correlations can also be fitted using polynomials. In this manner, the correlation function can be formulated as:

$$
\mathcal{C}_{\text{POLY}}\left(\boldsymbol{h}; u\right) = \mathcal{C}_{\text{F.SYM}}\left(\boldsymbol{h}; u\right) + \lambda \mathcal{C}_{\text{WIND.POLY}}\left(\boldsymbol{h}; u\right)
\tag{4.19}
$$

where λ is a fitted parameter and $\mathcal{C}_{\text{WIND.POLY}}\left(\boldsymbol{h}; u\right)$ has similar role with that of $\mathcal{C}_{\text{LGR}}\left(\boldsymbol{h}; u\right)$. The difference between $\mathcal{C}_{\text{WIND.POLY}}\left(\boldsymbol{h}; u\right)$ and $\mathcal{C}_{\text{LGR}}\left(\boldsymbol{h}; u\right)$ will be apparent in section 4.4.1.1.

4.4 Exploratory analyses and results

4.4.1 Exploratory analyses

Before extensive forecasting is performed, some exploratory analyses are conducted to examine the empirical spatio–temporal correlation matrix. Data from a single day, namely 5 August 2010, is used in section 4.4.1.1 to demonstrate the forecasting at observed locations. For section 4.4.1.2 which validates forecasting at an unobserved location, the whole 13 days are tested. As an illustration, m_{lag} is taken to be 1 (m_{lag} is the maximum time lag considered in the forecasting calculation, as mentioned in section 4.3.1), $\bar{t}$ is taken to be 50 seconds (the 1–second data are averaged into 50–second data), and only 50% of the data of each day are predicted (since in the forecasting process we need to split the data into training and testing parts). These 50% data correspond to those of the later half of each day. In order to get a more detailed understanding of the performance of the spatio–temporal models, sections 4.4.2 and 4.4.3 present the FS values for various forecast horizons/temporal aggregations, maximum time lag m_{lag} values, and learning data lengths for the whole 13 days.

As mentioned in section 4.2 and following Eq. (4.2), in the very first step, the irradiance values are first divided by the corresponding clear sky irradiance and then subtracted by the mean value of the corresponding station k_{T}. The adopted clear sky irradiance model was proposed by Perez et al. (2002). The Linke turbidity used in the model was formulated by Ineichen (2008).

4.4.1.1 Observed locations

In this section, forecasting at observed locations is demonstrated, meaning historical data from all stations are included in the learning process. Fig. 4.1 shows the purely spatial correlation plot of the observation values. After the points in Fig. 4.1

are fitted to $\mathcal{C}_S(\boldsymbol{h})$ in Eq. (4.13), it is estimated that $\hat{\eta} = 0$ and $\hat{c} = 0.822$. All the fittings in this chapter, except for those of STAT correlation functions, are done using limited–memory Broyden–Fletcher–Goldfarb–Shanno algorithm with boundary conditions (L-BFGS-B) (Byrd et al., 1995) integrated in the function `optim` of the R statistical software (R Core Team, 2015). The fitting of STAT correlation function will be explained later.
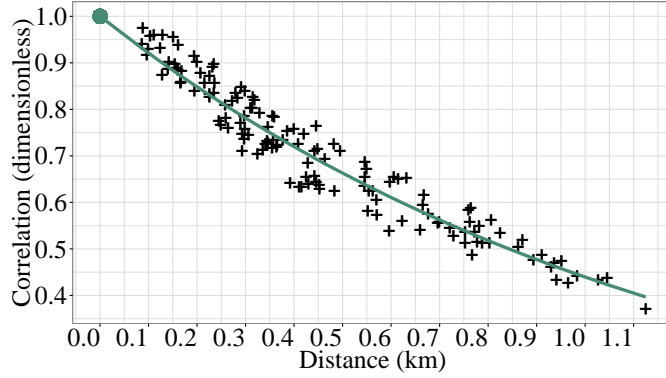
Fig. 4.1: Empirical and fitted purely spatial correlation of 50% learning data on 5 August 2010 for $\bar{t} = 50$ seconds (50–second average).

Similarly, fitting can be done on $\mathcal{C}_T(u)$ as shown in Fig. 4.2. It is estimated that $\hat{a} = 0.325$ and $\hat{\vartheta} = 0.754$ after fitting is done according to Eq. (4.14). Simple multiplication between $\mathcal{C}_S(\boldsymbol{h})$ and $\mathcal{C}_T(u)$ results in SEP correlation function (Eq. (4.12)). For F.SYM correlation function, fitting can be done based on Eq. (4.15). In this way, it is estimated that $\hat{\varrho} = 1$ given that the values of $\hat{\eta}$, $\hat{c}$, $\hat{a}$, and $\hat{\vartheta}$ from the SEP correlation function fitting are retained, following Gneiting et al. (2007b).

The next step is to investigate the asymmetry with respect with along and cross wind displacement vectors. As mentioned, in section 4.2, the wind direction is taken to be 60° West from South owing to the prior knowledge of the dominant wind direction in this area. In the future, wind direction forecast may be included, for example, by making use of intra–day weather forecasting model, such
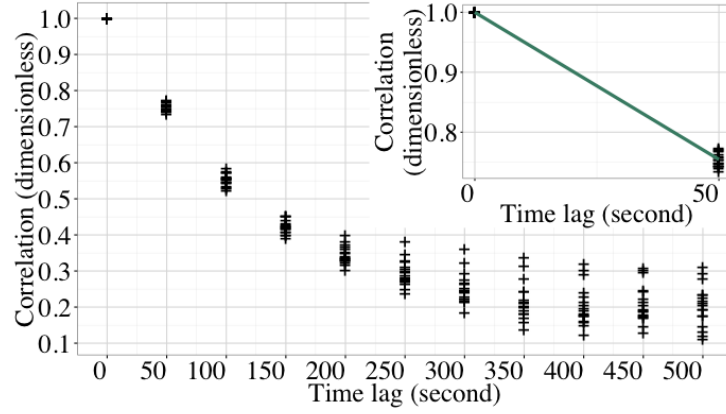
Fig. 4.2: Empirical purely temporal correlation of 50% learning data on 5 August 2010 for $\bar{t} = 50$ seconds and maximum time lag $m_{\mathrm{lag}} = 10$. Inset: Fitted temporal correlations with maximum time lag $m_{\mathrm{lag}} = 1$.

as Global Forecasting System (GFS) (NOAA, 2015a), North American Mesoscale Model (NAM) (NOAA, 2015b), or European Centre for Medium–Range Weather Forecasts (ECMWF) (ECMWF, 2015).

First, the modeling of POLY correlation function is described here. As an illustration, asymmetry between stations DH8 and AP1 for lag u is computed as the difference between $\mathsf{cor}(z(\boldsymbol{s}_{\mathrm{DH8}}; t_k), z(\boldsymbol{s}_{\mathrm{AP1}}; t_k - u))$ and $\mathsf{cor}(z(\boldsymbol{s}_{\mathrm{AP1}}; t_k), z(\boldsymbol{s}_{\mathrm{DH8}}; t_k - u))$. Two asymmetry models are visualized in Fig. 4.3. They are polynomial line and polynomial surface fitted models. Surface fitting is made possible by including the information of cross wind distance. From this points onwards, POLY correlation function which utilizes polynomial line fitting will be called POLY.A while the other will be called POLY.B. Referring to Eq. (4.19), we are able to put the 'asymmetry' information into $\mathcal{C}_{\mathrm{WIND.POLY}}(\boldsymbol{h}, u)$ as stated in Eqs. (4.20) and (4.21) for POLY.A and POLY.B respectively

$$
\begin{aligned}
\mathcal{C}_{\mathrm{WIND.POLY,A}}(\boldsymbol{h}, u) = H\left[u\right] H\left[h_1\right] \{ & K_{A,u}\left(h_1\right)^4 + L_{A,u}\left(h_1\right)^3 \\
& + M_{A,u}\left(h_1\right)^2 + N_{A,u}\left(h_1\right) \}
\end{aligned}
\tag{4.20}
$$
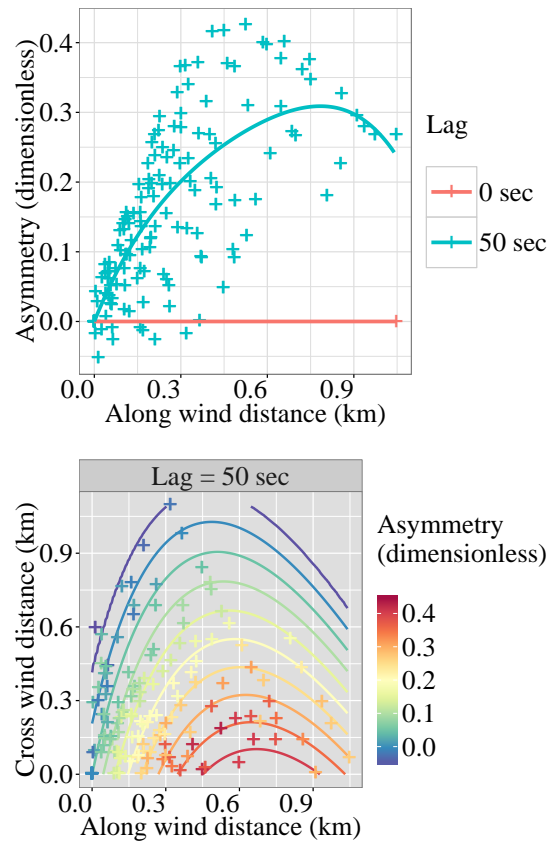
Fig. 4.3: (top) Empirical and fitted asymmetry in correlation with respect to along wind distance of 50% learning data on 5 August 2010 for $\bar{t} = 50$ seconds and maximum time lag $m_{\text{lag}} = 1$. (bottom) Empirical and fitted asymmetry in correlation with respect to along and cross wind distances.

$$\mathcal{C}_{\text{WIND.POLY,B}}\left(\boldsymbol{h}, u\right) = H\left[u\right] H\left[h_1\right] \left\{ K_{B,u}\left(h_1\right)^2 + L_{B,u}\left(h_2\right)^2 \right.$$
$$+ M_{B,u}\left(h_1\right)\left(h_2\right) + N_{B,u}\left(h_1\right)$$
$$\left. + O_{B,u}\left(h_2\right) + P_{B,u}\right\} \tag{4.21}$$

where h_2 is the cross wind component of $\boldsymbol{h}$, K, L, M, N, O, and P are coefficients which are fitted separately for different u, and $H\left[\cdot\right]$ is the Heaviside step function as defined in Eq. (4.22)

$$H\left[x\right] = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{4.22}$$

As mentioned in section 4.2, the dominant wind direction is taken to be $60°$ West from South following Hinkelman (2013). The fitting coefficients are found to be as follows: $\hat{K}_{A,1} = -0.850$, $\hat{L}_{A,1} = 1.727$, $\hat{M}_{A,1} = -1.643$, $\hat{N}_{A,1} = 1.030$, $\hat{K}_{B,1} = -0.837$, $\hat{L}_{B,1} = -0.003$, $\hat{M}_{B,1} = -0.330$, $\hat{N}_{B,1} = 1.155$, $\hat{O}_{B,1} = -0.238$, and $\hat{P}_{B,1} = 0.049$. Regarding the along wind component h_1 mentioned previously, as an illustration, referring to the network layout in Fig. 1.6, $\text{cor}(z(\boldsymbol{s}_{\text{DH8}}; t_k), z(\boldsymbol{s}_{\text{AP1}}; t_k - u))$ has positive h_1 while $\text{cor}(z(\boldsymbol{s}_{\text{AP1}}; t_k), z(\boldsymbol{s}_{\text{DH8}}; t_k - u))$ has negative h_1.

By utilizing Eqs. (4.20) and (4.21), the suitable values of $\hat{\lambda}$ for two of them are found to be 0.955 (for POLY.A) and 0.797 (for POLY.B). Those $\hat{\lambda}$ values are chosen in such a way that the forecast of the learning data gives the smallest RMSE. In addition, here, the values of $\hat{\eta}$, $\hat{c}$, $\hat{a}$, $\hat{\vartheta}$, and $\hat{\varrho}$ from the fittings of SEP and F.SYM correlation functions are retained.

For STAT correlation function (refer to Eqs. 4.17 and 4.18), the fittings of parameters η, c, a, ϑ, ϱ, v, and λ are done simultaneously using the function `GenSA` (Xiang et al., 2013). This correlation function is treated differently since utilization of function `optim` does not lead to good results.

After all correlation matrices have been obtained, forecasting can be performed. Fig. 4.4 shows the rRMSE for GHI forecasting (of 50% testing data) of

each station and forecasting method on 5 August 2010 with $\bar{t} = 50$ seconds and maximum time lag $m_{\text{lag}} = 1$. The results in Fig. 4.4 are obtained by fitting parameters η, c, a, ϑ, ϱ, K, L, M, N, O, P, and λ separately for each time step (learning data become like a moving window with fixed length) since this way yields slightly better results compared to those when the coefficients are fitted only once per day. Exception is applied for STAT correlation function since the function GenSA typically takes longer time (~ 1.5 minute) for fitting. Thus, fitting is only done once for each day for this particular correlation function. The stations in Fig. 4.4 are sorted according to the dominant wind direction. Besides kriging with the five spatio–temporal correlation functions mentioned previously (EMP, SEP, F.SYM, STAT, and POLY), three univariate models are used for benchmarking purpose, namely persistence (Pers), ARIMA model (ARIMA) (Yang et al., 2012), and exponential smoothing (ETS) (Yang et al., 2015a). The Pers method (Inman et al., 2013) is performed following Eq. (4.23)

$$\hat{I}_{t_k+1} = k_{\text{T},t_k} \times I_{\text{clr},t_k+1} = \frac{I_{t_k}}{I_{\text{clr},t_k}} \times I_{\text{clr},t_k+1} \tag{4.23}$$

where $\hat{I}_{t_k+1}$ is the estimated irradiance at one time step ahead, I_{clr,t_k} and I_{clr,t_k+1} are computed/modeled clear sky irradiance at current time and one time step ahead respectively, I_{t_k} is the measured current irradiance, and k_{T,t_k} is the current clear sky index. The rRMSE is computed according to Eq. (1.2).

Fig. 4.4 reveals that, in this case, EMP kriging is able to produce the best forecasting result in terms of median rRMSE. In this criterion, the second and third best results are achieved by POLY.B and POLY.A kriging respectively. However, STAT kriging has smaller 75% quantile than that of POLY.A kriging. Another feature noted in the figure is that stations which are preceded by other stations in the upwind direction (for example station DH10 is preceded by stations DH3 and AP1, while station DH1 is not preceded by any station in the upwind direction) tend to have lower rRMSE, especially for EMP and POLY kriging.
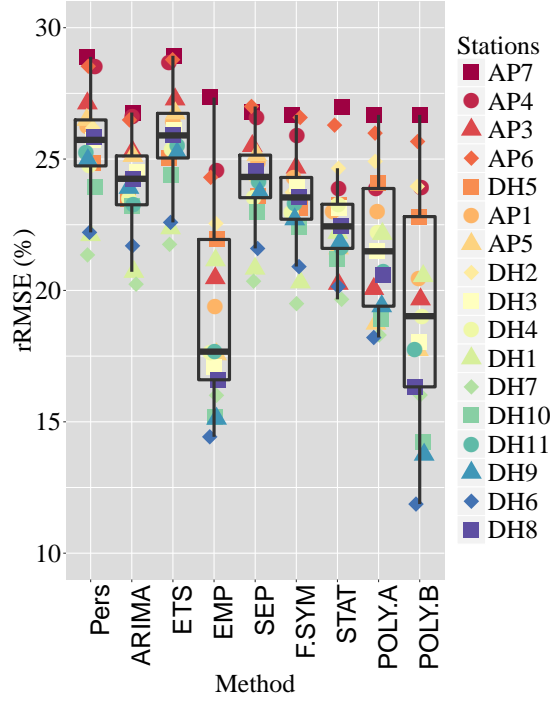
Fig. 4.4: Box plot of irradiance forecasting rRMSE at observed locations (5 August 2010, $\bar{t} = 50$ seconds, 50% learning data, maximum time lag $m_{\text{lag}} = 1$). The ends of the whiskers refer to the highest and lowest data within $1.5 \times$ inter-quantile-range.

4.4.1.2 Unobserved locations

In order to validate irradiance forecast at an unobserved location, one station is not included when correlation matrices are constructed. In this case, station DH10 is chosen as the testing station. Obviously, EMP kriging cannot be done for prediction at an unobserved location since it is assumed that we do not have any historical data at those points. Thus, only SEP, F.SYM, STAT, and POLY kriging are applicable.

By referring to section 4.2 and Eq. (4.2), after calculating $\boldsymbol{c}(\boldsymbol{s}_{\text{DH10}}; t_k + 1)^\top \mathfrak{S}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu})$, we need to add back the mean of k_{T} at station DH10 before multiplying the result with the appropriate I_{clr, t_k+1} to get the predicted irradiance on location $\boldsymbol{s}_{\text{DH10}}$ at time $t_k + 1$. Since it is assumed that we do not have any data on the location of station DH10, the mean of k_{T} of all learning stations is added to $\boldsymbol{c}(\boldsymbol{s}_{\text{DH10}}; t_k + 1)^\top \mathfrak{S}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu})$ instead.
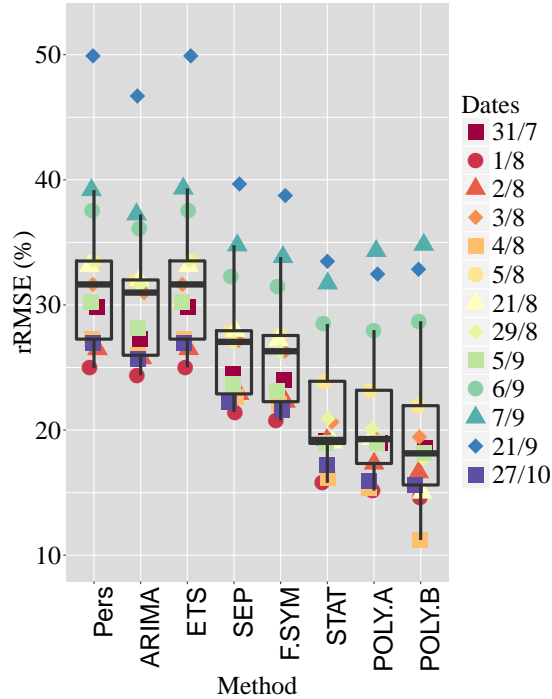
Fig. 4.5: Box plot of irradiance forecasting rRMSE at unobserved location AP3 (13 days dominated by broken clouds in 2010, $\bar{t} = 50$ seconds, 50% learning data, maximum time lag $m_{\mathrm{lag}} = 1$).

For benchmarking purposes, the mean values of univariate forecasting results (derived by Pers, ARIMA, and ETS methods) of learning stations are adopted. Fig. 4.5 shows the rRMSE of one time step ahead irradiance prediction on station AP3 on all the 13 days dominated by broken clouds (mentioned in section 4.2) with $\bar{t} = 50$ seconds, 50% learning data, and maximum time lag $m_{\mathrm{lag}} = 1$. The figure shows that on most days, POLY.B kriging is able to predict irradiance at AP3 with highest accuracy in terms of median rRMSE. According to this criterion, the performances of STAT and POLY.A kriging are comparable.

4.4.2 Forecasting with various forecast horizons

In this section, the accuracy of the proposed models are tested for various forecast horizons. In addition, for forecasting at observed locations, various maximum time lag m_{lag} values are tested. The m_{lag} value with the smallest RMSE is chosen for
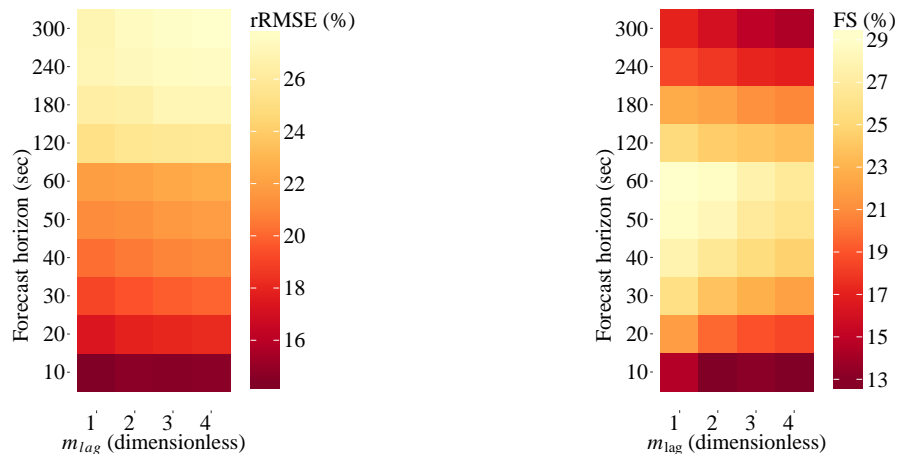
Fig. 4.6: (left) rRMSE of irradiance prediction at observed locations (averaged over all stations and the 13 days dominated by broken clouds in 2010) of POLY.B kriging for various forecast horizons and maximum time lag m_{lag} values (50% learning data). (right) As the left one but for FS.

more detailed evaluation including forecasting at unobserved locations. It should be noted that the meaning of m_{lag} here depends on the forecast horizon since the irradiance data are aggregated according to it (as mentioned in section 4.2). For example, when the forecast horizon is 120 seconds and $m_{\text{lag}} = 3$, it means aggregated data at 120, 240, and 360 seconds (corresponding to time step t/now, $t - 1$, and $t - 2$ respectively) before time $t + 1$ are included in the forecasting calculation.

4.4.2.1 Observed locations

Fig. 4.6 shows the effect of maximum time lag m_{lag} values and forecast horizon on rRMSE and FS (averaged over all stations and the 13 days dominated by broken clouds in 2010) of POLY.B kriging. The length of the learning data is 50%. The reason of focusing on POLY.B kriging here is because, as mentioned in section 4.1, ultimately we would like to find the most suitable fitted spatio–temporal correlation function as a first step to do forecasting at unobserved locations. Fig. 4.6 reveals that as the forecast horizon gets longer, as expected, the rRMSE becomes worse. In terms of m_{lag}, it seems that as m_{lag} gets larger, the learning data introduce more

error. For FS, it is observed that FS generally decreases as m_{lag} increases. In terms of forecast horizon, FS has its peak at forecast horizon near 60 seconds. Similar trends occur for POLY.A kriging (not shown here). Thus, in the next part, including forecasting at unobserved locations, $m_{\text{lag}} = 1$ is chosen to inspect the accuracy of each method.

Fig. 4.7 displays the FS of several models with various forecast horizons, averaged over 13 days dominated by broken clouds in 2010 (maximum time lag $m_{\text{lag}} = 1$, 50% learning data). It seems that EMP kriging, in general, performs the best among different methods for sub–100–second forecast horizons. However, as the forecast horizon increases, the accuracy of EMP kriging degrades. For forecast horizons longer than 100 seconds, POLY.A and POLY.B kriging perform the best in overall. The second model outperforms the first one for forecast horizons smaller than 100 seconds. STAT kriging also shows better accuracies than most of the methods, but it is generally less accurate than POLY.A and POLY.B kriging.

The FS of different stations follow the trend spotted in Fig. 4.4, namely stations which are preceded by other stations in the upwind direction tend to have higher FS. The other interesting feature observed in Fig. 4.7 is different "peaks" of EMP kriging FS lines are located at different forecast horizons for different stations. For example, for station AP3, the peak resides at forecast horizon of 50 second, but that of station DH3 is located at forecast horizon of 20-30 second. Intuitively, it happens due to the distance of each station to the preceding stations in the upwind direction. In this case, the distance of station AP3 to AP7 is longer than those of station DH3 to AP1. Larger distance implies more suitability for prediction at longer forecast horizon. This observation is consistent, especially with Hinkelman (2013) and Lonij et al. (2013).

Similar plots with different maximum time lag m_{lag} are also investigated. However, here, only such plot of stations AP7, AP1, and DH8 for $m_{\text{lag}} = 4$ is displayed (in Fig. 4.8). No other plots are shown since similar trend is observed

in every station and other m_{lag} values. Degradation in FS when m_{lag} increases is very obvious, especially for EMP kriging. The performances of POLY.A and POLY.B kriging seem to diminish as well, but overall they still outperform the other methods (except for sub–60–second forecast horizons in some stations where EMP kriging has the highest accuracy).

4.4.2.2   Unobserved locations

Fig. 4.9 shows the FS for irradiance forecasting at unobserved locations, averaged over the 13 days dominated by broken clouds in 2010 (maximum time lag $m_{\mathrm{lag}} = 1$, 50% learning data). For this case, forecasting at each location is done separately. Each time, 16 stations are used as learning stations, leaving one out as a testing station.

Fig. 4.9 reveals that the accuracies of POLY.A and POLY.B kriging generally outperform those of the other algorithms. POLY.B kriging is sligthly more accurate compared to POLY.A kriging for forecast horizons shorter than 100 seconds. The accuracies of all methods become comparable again as the forecast horizon approaches 300 seconds. There is some degradation of FS at $\bar{t} = 50$ or 60 seconds at some "downwind" stations, such as stations DH6 and DH8. It can be explained using Fig. 4.10. Fig. 4.10 shows the accuracy of two type of Pers forecasting: observed (section 4.4.1.1) and unobserved (section 4.4.1.2). The second type of Pers forecasting may reach low rRMSE for certain forecast horizon at stations located at a downwind zone (for example at station DH8 for forecast horizon of 50 second). It is justifiable since indeed past clear sky indeces of stations in the upwind zone have high correlation with current clear sky indeces of downwind zone.

4.4.3 Forecasting with various training data lengths

Fig. 4.11 shows the comparison of FS for irradiance prediction at observed locations between EMP, POLY.A, and POLY.B kriging for various amounts of learning data and forecast horizons. The figure reveals that the accuracy of EMP kriging is very sensitive to the amounts of learning data. As fewer learning data are available, the FS of EMP kriging drops significantly, especially for longer forecast horizons. In contrast, this is not the case for POLY kriging.

Fig. 4.12 illustrates the FS for irradiance prediction at unobserved locations of POLY.A and POLY.B kriging for various amounts of learning data and forecast horizons. The figure shows that learning data percentage, in general, does not affect much the forecasting accuracies of these two kriging methods: a condition which is also observed in Fig. 4.11.

4.5 Discussion

Forecasting on observed and unobserved locations have been performed with the kriging methods with asymmetric correlation functions to show that good performances can be achieved compared with simpler, non-spatial forecasting methods. For example, for station AP3, the FS for irradiance prediction reaches 0.37 (as shown in Fig. 4.9) even when historical data of the corresponding station is not included in the forecasting algorithm. The fact that the proposed anisotropic methods do not have equal performance for all stations reemphasizes the importance of sensor network design plannning, especially if the network is going to be utilized for forecasting purposes (Lonij et al., 2013). Forecasting can indeed be performed on locations without historical data, but the accuracy is constrained by the locations of neighbouring stations.

Another important point is regarding the presence of asymmetry in the

spatio–temporal correlations between stations clear sky indexes as visualized in Fig. 4.3. Asymmetry is a key element in order for POLY.A and POLY.B kriging to outperform SEP and F.SYM kriging methods for forecasting at unobserved locations. As pointed out by Yang et al. (2015b) and shown in Fig. 4.5, POLY.A and POLY.B kriging do not outperform simpler kriging methods for the 7 September 2010 case due to the lack of anisotropy in the spatio–temporal correlations on that day. Thus, this method is suitable for areas where persistent wind in one direction takes place most of the time such as Oahu Island. Changes of wind direction for certain periods may not be an issue given that accurate wind forecasting system exists. Thus, integration with numerical weather model may be an option for development of these methods.

Lastly, as expressed in section 4.1, this chapter only performs irradiance forecasting for sub–5–minute forecast horizons. Principally, it is due to the size of the network. Larger network (with high density) is necessary in order to accomplish accurate forecasting at longer forecast horizons.

4.6   Conclusion

This chapter has exhibited very short–term (less than 5–minute–ahead) irradiance forecasting using spatio–temporal kriging for a dense solar network in days dominated by broken clouds. Forecasting is also done at unobserved locations, namely where irradiance sensors are not installed. The result shows that a forecasting technique which makes use of spatio–temporal correlation, given that spatio–temporal correlations between sensors in a network are high, is able to outperform conventional univariate model (which mainly relies on temporal autocorrelations), such as persistence, ARIMA, and ETS, both for observed and unobserved locations.

Forecasting at unobserved locations is made possible by fittings on empirical correlation matrix. Fittings are done using various functions, namely spatio–

temporally separable, fully symmetric, and anisotropic fitting. Anisotropic fitting with polynomial, which makes use of the knowledge of dominant wind direction, has been proven to produce the best forecast for observed (beyond 100–second forecast horizon) and unobserved (sub–5–minute forecast horizons) locations. The highest FS achieved by the polynomial fit kriging are 0.43 and 0.37 (both for forecast horizon of 50 seconds) for prediction at observed and unobserved locations respectively on days dominated by broken clouds.
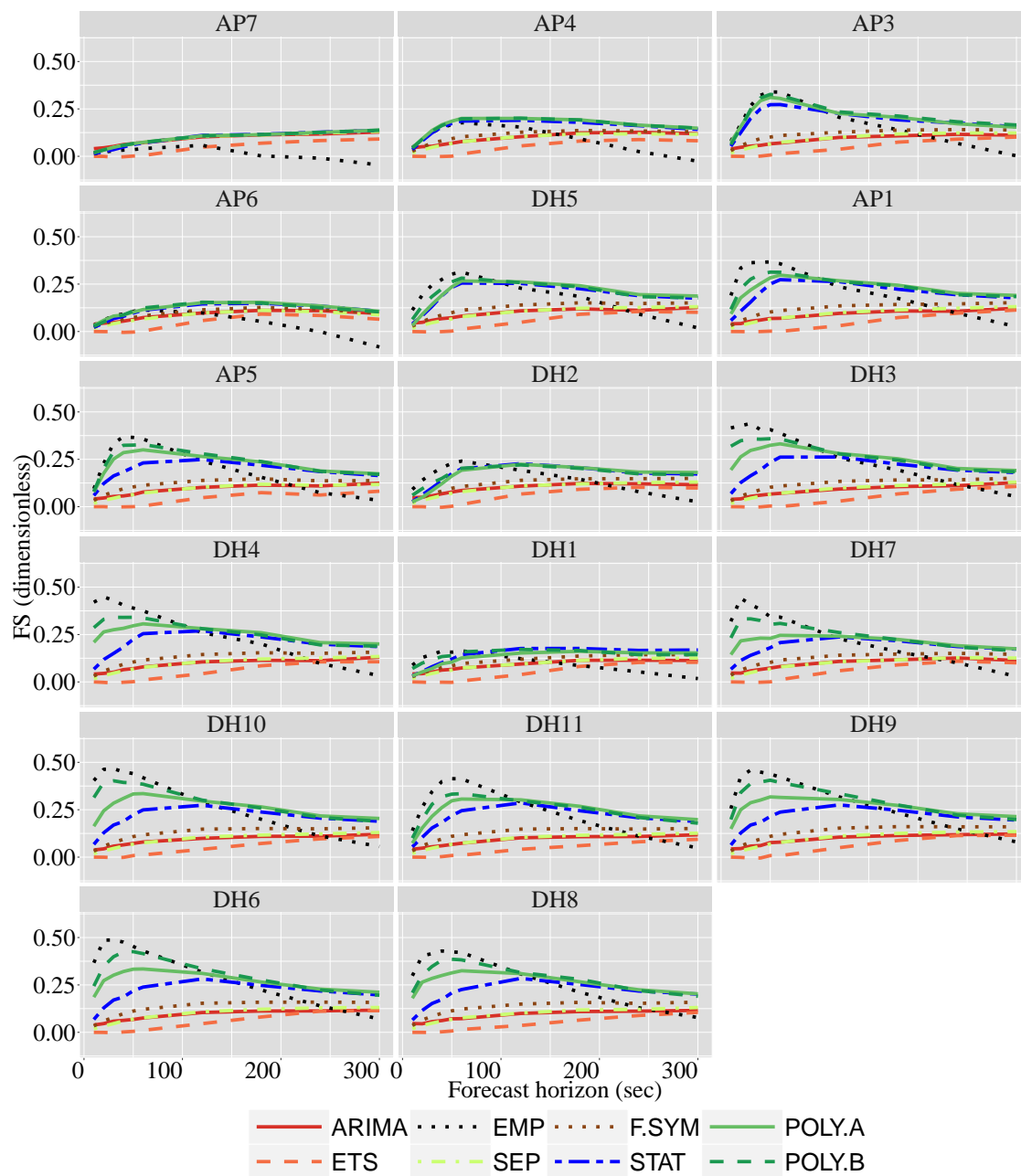
Fig. 4.7: FS for irradiance prediction at observed locations for various forecast horizons (averaged over the 13 days dominated by broken clouds in 2010, maximum time lag $m_{lag} = 1$, 50% learning data)
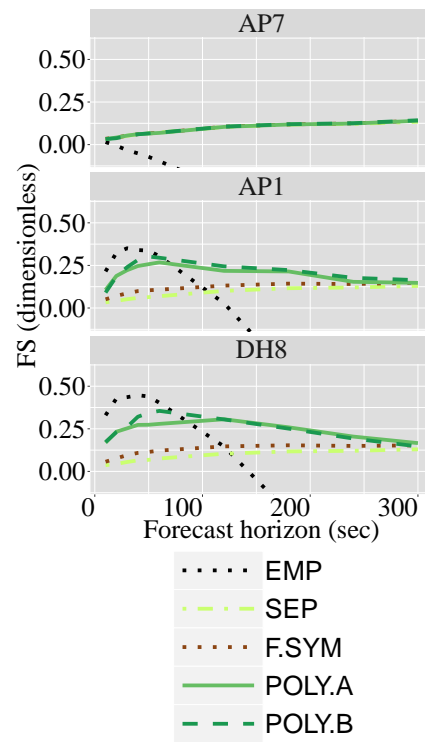
Fig. 4.8: FS for irradiance prediction at observed locations for various forecast horizons (averaged over the 13 days dominated by broken clouds in 2010, maximum time lag $m_{\mathrm{lag}} = 4$, 50% learning data)
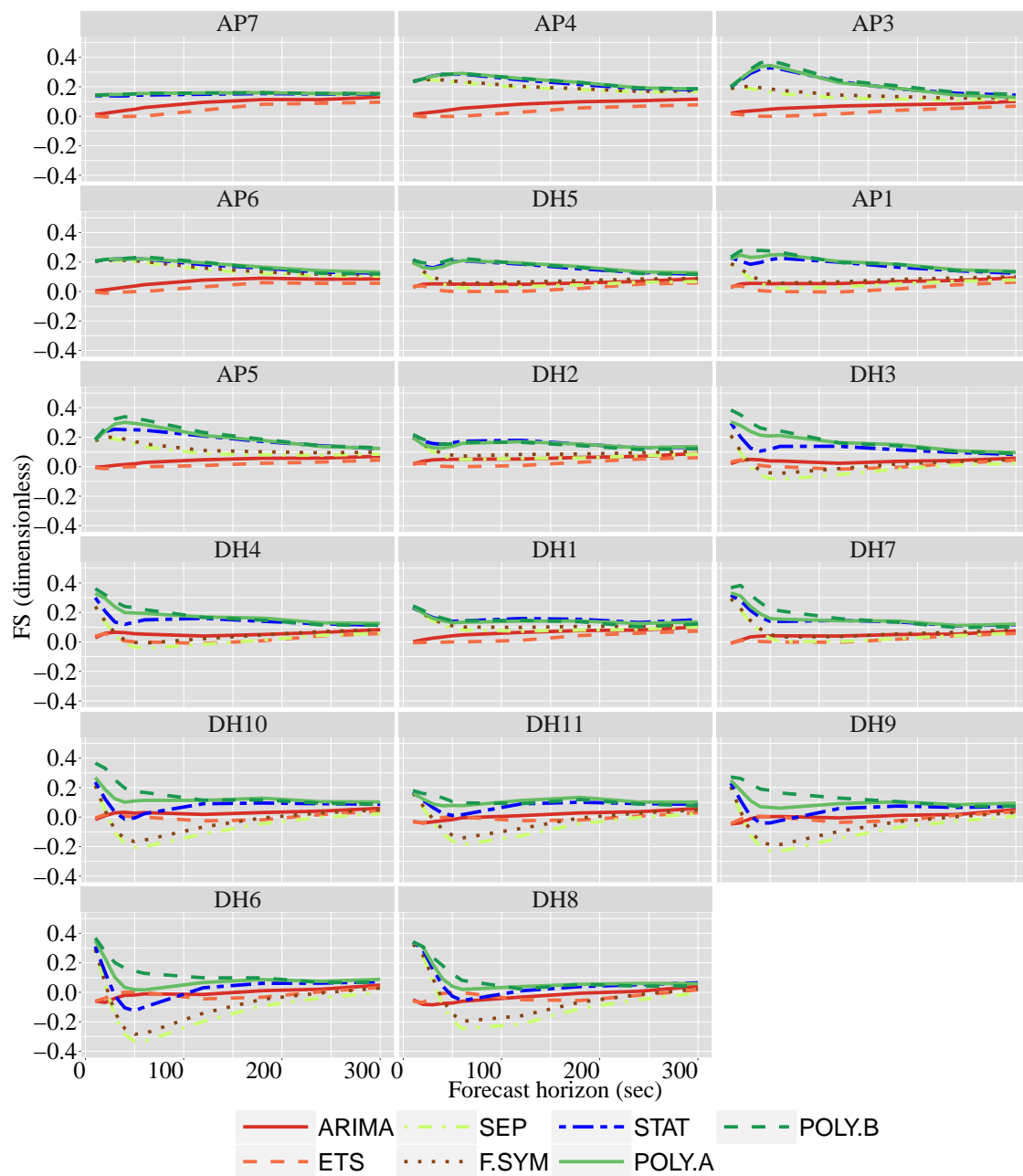
Fig. 4.9: FS for irradiance prediction at unobserved locations for various forecast horizons (averaged over the 13 days dominated by broken clouds in 2010, maximum time lag $m_{\text{lag}} = 1$, 50% learning data)
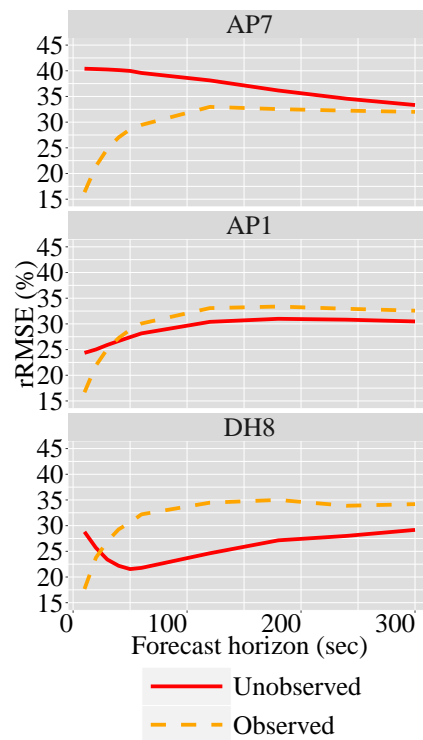
Fig. 4.10: rRMSE of persistence of observed and unobserved locations (averaged over the 13 days dominated by broken clouds in 2010, 50% learning data)
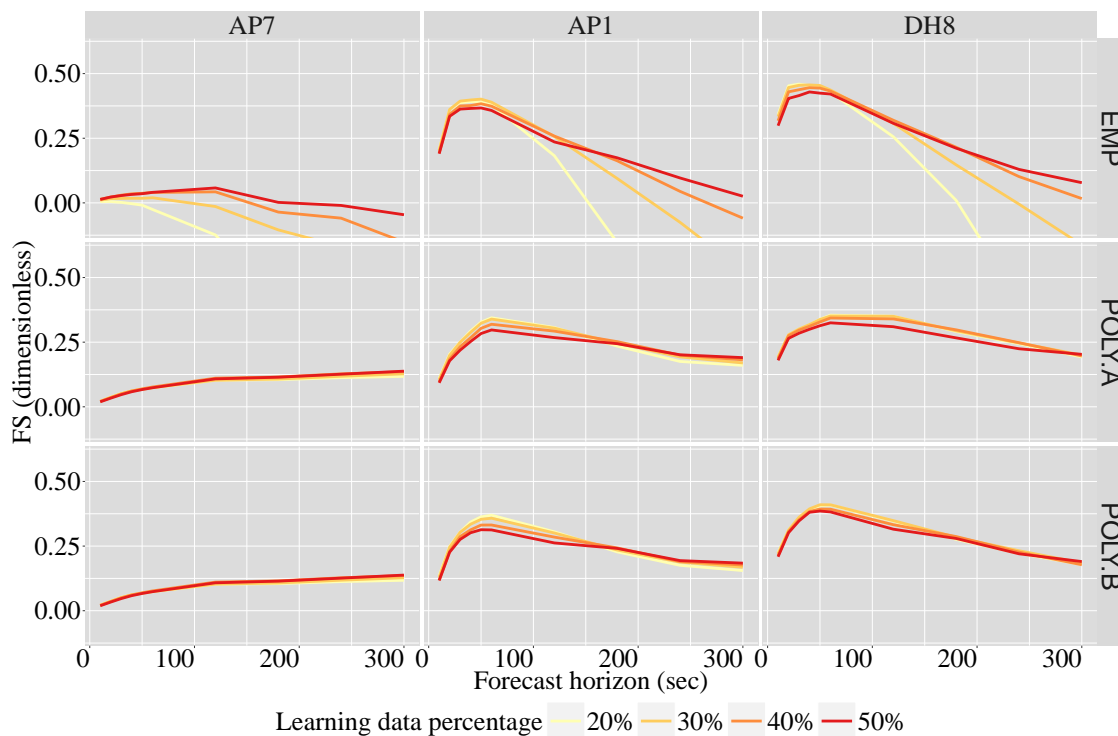
Fig. 4.11: FS for irradiance prediction at observed locations for various forecast horizons and learning data percentages (averaged over the 13 days dominated by broken clouds in 2010, maximum time lag $m_{\text{lag}} = 1$)
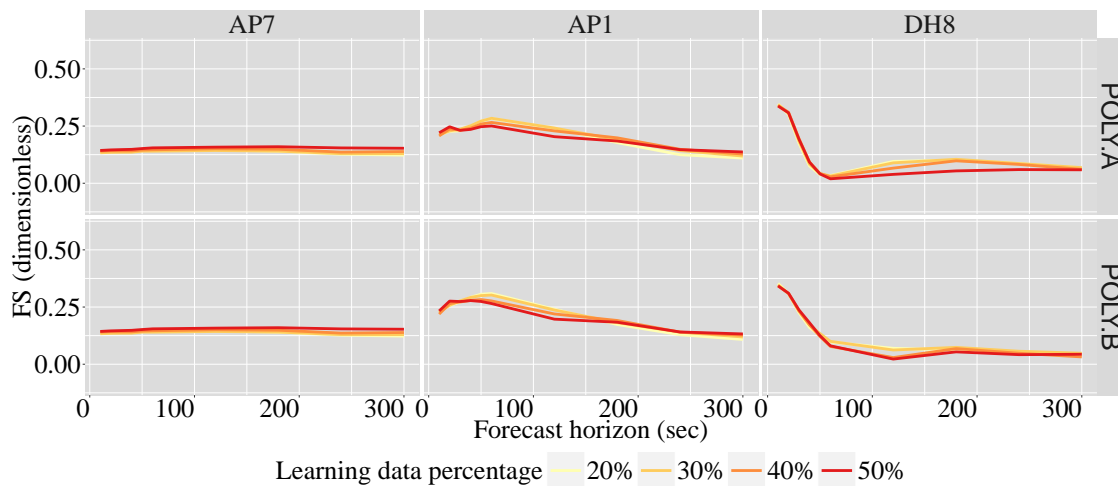
Fig. 4.12: FS for irradiance prediction at unobserved locations for various forecast horizons and learning data percentages (averaged over the 13 days dominated by broken clouds in 2010, maximum time lag $m_{\text{lag}} = 1$)

Chapter 5

Modeling variability and geographic smoothing

5.1 Introduction

The undesirable variability of solar irradiance, especially in the case of electricity grids with a high penetration of PV generation, has been addressed in chapters 2 and 4. In addition, chapter 4 offers a possible solution using very-short-term irradiance forecasting. Another consideration which must be taken into account, especially in the framework of resource assessment, is the geographic smoothing effect, which is assessed empirically in chapter 2. Knowledge of this effect helps, for example, system planners to decide the requirements of reserve capacity to lessen the worst case of extreme ramp events (Hoff and Perez, 2012).

Lave et al. (2013) designed a wavelet-based variability model (WVM) to simulate the variability of a large solar power plant using inputs of irradiance data

---

This chapter is based on: L. Zhao, A. W. Aryaputera, D. Yang, and W. M. Walsh, "Modeling of solar variability and geographic smoothing effect," *PV Asia Scientific Conference*, Singapore, 2015 (oral presentation).

from several sensors. This method requires a function which describes correlations between ramp events in terms of pair-station distances and time resolutions. Lave et al. (2013) constructed this function empirically using data from 6 irradiance sensors. Arias-Castro et al. (2014) made an attempt to design the same type of correlation function, called the Anisotropic Correlation Model (ACM), using merely along- and cross-wind distances of pair stations, estimated cloud diameter, cloud speed, and time resolutions. Thus, the method enables the simulation of solar power plant ramp events using irradiance data from only one sensor. Earlier, Perez et al. (2011) had proposed similar correlation model, which depends on pair-station distances, cloud speed, and time resolutions.

Monger et al. (2016) proposed a completely different approach which involves the modeling of irradiance correlations (rather than those of ramp events) using data from a sensor network. Irradiance values at unobserved locations are then predicted using the kriging method. The aggregated simulated (and measured) irradiance values can then be utilized to estimate the distribution of ramp events over an area, on which a solar power plant is going to be built.

This chapter presents simulations of geographic smoothing using the geostatistical space-time models of Gneiting et al. (2007b). These methods have been used for irradiance forecasting at unobserved locations in chapter 4 while here they are utilized to simulate the irradiance ramp event. These methods are more similar to that of Monger et al. (2016) than WVM (Lave et al., 2013) since they attempt to model the spatio-temporal process of irradiance values rather than that of ramp events. In this chapter, almost-exactly-the-same methods as those in chapter 4 are employed to simulate the ramp events. The results are then compared with those of several existing methods (Lave et al., 2013; Arias-Castro et al., 2014; Perez et al., 2011; Monger et al., 2016). The difference between the space-time models of Gneiting et al. (2007b) with that of Monger et al. (2016) is the former ones are able to use the historical observation data for prediction while the latter is not.

This chapter is structured as follows. Section 5.1 presents the short introduction and literature reviews. Section 5.2 mentions briefly the data used in this chapter. Sections 5.3 and 5.4 describe the methods used and results achieved in this chapter. Finally, the conclusion is written in section 5.5.

5.2 Irradiance data

The irradiance data used in this chapter are the same as those employed in chapter 4, namely the irradiance data of the National Renewable Energy Laboratory (NREL) network in Oahu, Hawaii (Sengupta and Andreas, 2010). These data have been described in sections 1.5.1.4 and 4.2, and the map of the sensor network is visualized in Fig. 1.6. The 13 days referred by Hinkelman (2013) are used due to the existence of wind with persistent direction towards 60° West from South and broken (cumulus) clouds.

5.3 Methods

The kriging method has been described in chapter 4. However, in this chapter, it is used only for spatial prediction (without time-forward processes) since no forecasting is performed here.

5.3.1 The kriging formulation

The kriging formulation used in this chapter is similar to that of section 4.3.1, specifically Eq. (4.2), with some exceptions, since data with time lag/$u = 0$ can be used for spatial-prediction here. Thus, instead of following Eq. (4.3), $\boldsymbol{Z}$ is expressed as:

$$\boldsymbol{Z} = \left(\boldsymbol{z}^{(0)\top}, \boldsymbol{z}^{(1)\top}, \cdots, \boldsymbol{z}^{(m)\top} \right)^{\top} \tag{5.1}$$

and consequently, instead of complying with Eq. (4.5), the $\mathfrak{S}$ is formularized as:

$$\mathfrak{S} = \begin{pmatrix} \mathfrak{S}_0 & \mathfrak{S}_1^\top & \cdots & \mathfrak{S}_m^\top \\ \mathfrak{S}_1 & \mathfrak{S}_0 & \cdots & \mathfrak{S}_{m-1}^\top \\ \vdots & \vdots & \ddots & \vdots \\ \mathfrak{S}_m & \mathfrak{S}_{m-1} & \cdots & \mathfrak{S}_0 \end{pmatrix}, \quad \in \mathbb{R}^{n(m+1) \times n(m+1)} \tag{5.2}$$

and similarly the c in Eq. (4.2) can be expanded as:

$$c = \begin{pmatrix} c_m^\top & \cdots & c_1^\top & c_0^\top \end{pmatrix}^\top, \quad \in \mathbb{R}^{n(m+1) \times 1} \tag{5.3}$$

with c_u and $c_{0j,u}$ remain as in Eqs. (4.9) and (4.10) respectively. Fitting of correlation function, which is notated in Eq. (4.11), is indispensable here due to the nature of the problem, which is spatial prediction.

As mentioned in section 5.1, the main difference of the kriging methods employed in this chapter with that of Monger et al. (2016) is the formers are able to use of historical values for spatial prediction, as shown in Eq. (4.2), while the latter is not. Historical values may increase the spatial prediction accuracy due to the persistent wind direction and cumulus clouds, mentioned in section 5.2. The other differences will be addressed in section 5.3.2 since they are related to the modeling of spatio-temporal correlation functions.

5.3.2 Classes of fitted spatio-temporal function and ramp events simulation process

There are three classes of fitted correlation functions used in this chapter, namely separable (SEP), fully symmetric (F.SYM), and general stationary (STAT). Their functions have been listed in Eqs. (4.12), (4.15), and (4.17). The fitting of the parameters are done simultaneously for each class using the function `GenSA` (Xiang et al., 2013).

Eqs. (4.12), (4.15), and (4.17) show that the three correlation functions contain the information regarding the relation between present observation values with those in the past. This is not the case for the variogram model of Monger et al. (2016). On top of that, STAT correlation function covers the asymmetry in the spatio-temporal correlation, which arises, for example, due to wind with prominent direction.

In order to simulate the ramp events, the observations of five stations are chosen as the learning data on which the spatio-temporal correlation functions are built. These stations are AP3, AP1, DH10, AP5, and DH2. Then, based on the correlation functions, similar with the method of Monger et al. (2016), irradiance values at the locations of the remaining stations are predicted using the kriging method. Subsequently, the spatially-averaged simulated (and measured) irradiance values are used to approximate the ramp events distribution of aggregated observed irradiance values.

Intuitively, the kriging method should underestimate the irradiance ramp rate distribution on any single spot since this method exercises prediction based on weighted average of observed values, as shown mathematically in Eq. (4.2). However, kriging turns out to give relatively good results for the case of aggregated irradiance ramp events, as will be shown in section 5.4.

5.3.3 Benchmarking methods

As mentioned in section 5.1, to benchmark the performances of SEP, F.SYM, and STAT kriging methods, several models from literatures are adopted. They are Lave.WVM (Lave et al., 2013), ACM.WVM (Arias-Castro et al., 2014), Perez.WVM (Perez et al., 2011), and Monger (Monger et al., 2016). It should be noted that among these four methods, only Lave.WVM and Monger require irradiance data from multiple learning stations, which are used to compute the correlations of

ramp events of some station pairs. The same five learning stations mentioned in section 5.3.2 are used for both methods. In contrast, ACM.WVM and Perez.WVM only require irradiance data from one station, provided that the distances between unknown stations, timescale, and cloud speed are available. Here, the same parameters for ACM.WVM and Perez.WVM are taken from Arias-Castro et al. (2014) since the same set of irradiance data is used.

A more straightforward benchmarking method is also employed. It is derived by taking the spatial average of irradiance data of the five learning stations (AGG.L).

5.4   Results and discussion

There are two types of simulations performed in this chapter. The first one (category I) estimates the spatially-aggregated irradiance ramp events distributions of all stations. In this case, the observed ramp events of learning stations are included in the spatial aggregation. However, since ACM.WVM and Perez.WVM take irradiance data from only one station, to make the comparison as fair as possible, the second type of simulation (category II) is adopted. It approximates the same thing as in category I, but only for the testing stations.

The differences between cumulative density functions (CDFs) of the simulated ramp events and those of the simulated ones are visualized in Figs. 5.1 - 5.4, following Arias-Castro et al. (2014). Perfect simulations will result in horizontal lines (zero difference between the simulated and measured CDFs). For the easiness of visualization, Figs. 5.1 and 5.3 show the performances of kriging method with the three classes of spatio-temporal functions (mentioned in section 5.3.2) and AGG.L, while Figs. 5.2 and 5.4 exhibit the performances of STAT kriging of Gneiting et al. (2007b) and other methods available in the literatures. Furthermore, Figs. 5.1 and 5.2 display the results for category I simulation while the other two figures display the results for category II simulation.
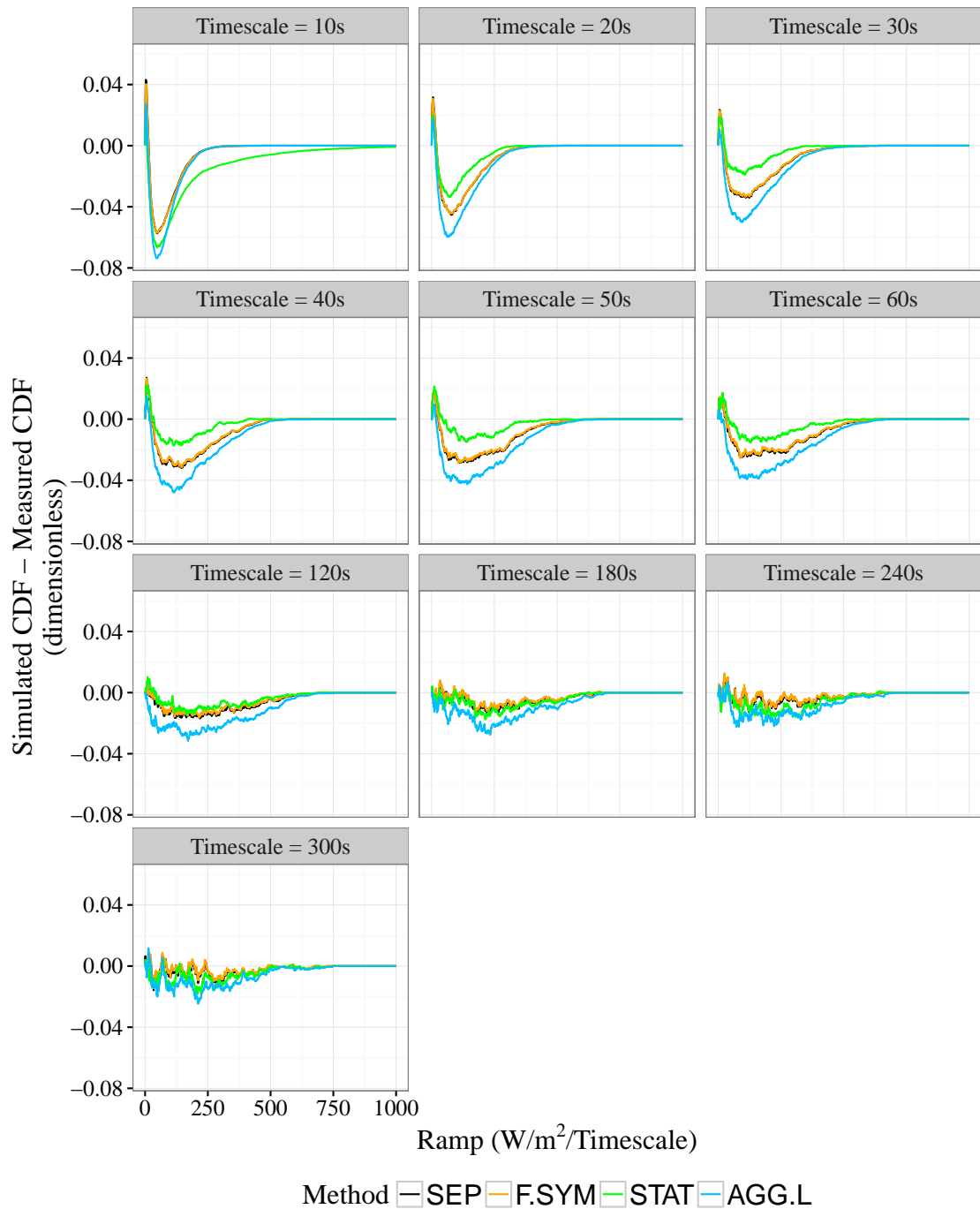
Fig. 5.1: Difference between the simulated and measured ramps of the category I simulation at various timescales. The simulation methods are SEP, F.SYM, and STAT (Gneiting et al., 2007b). For benchmarking purpose, the result of a simple spatial aggregation of irradiance values of learning stations (AGG.L) is also included.
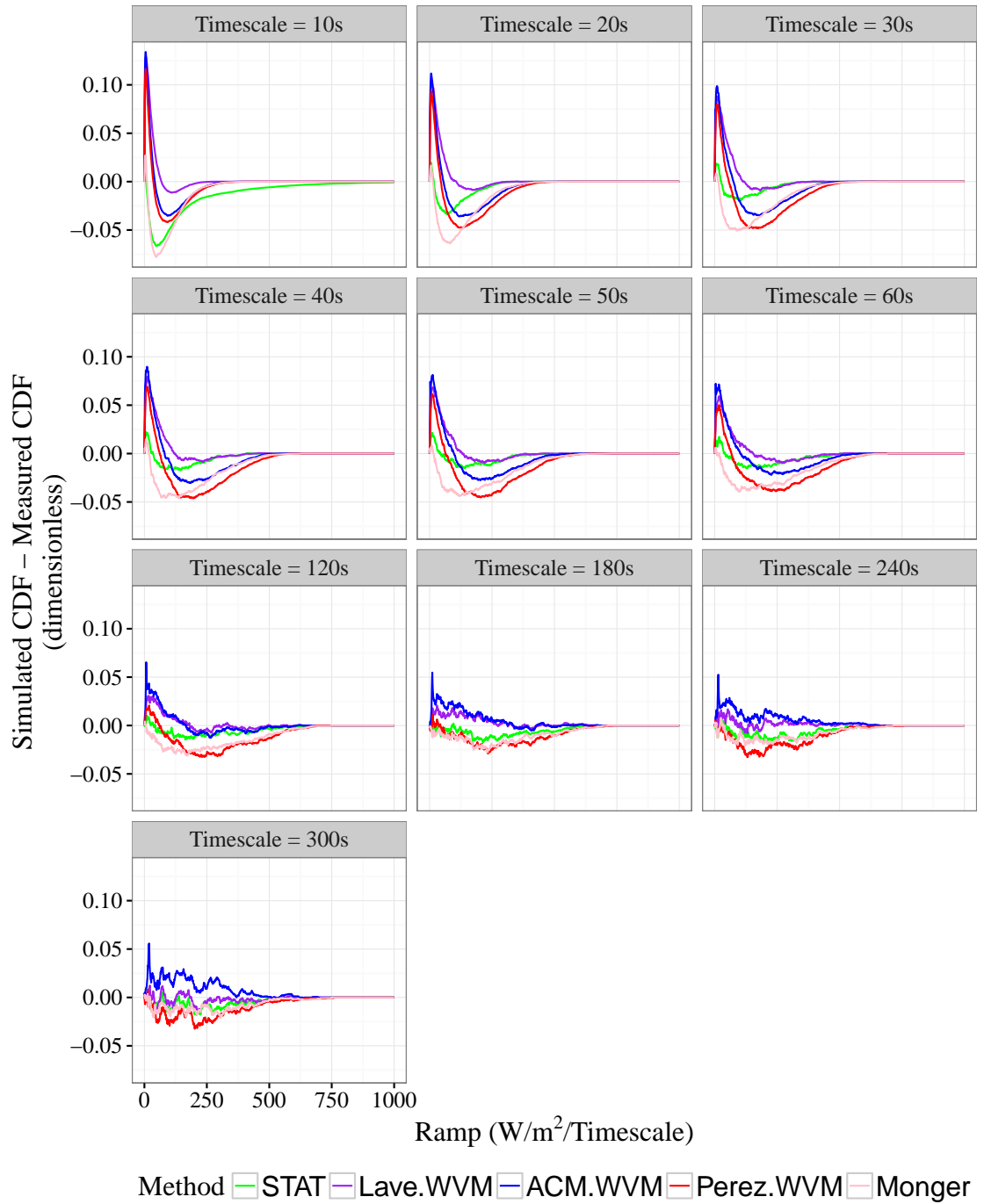
Fig. 5.2: As in Fig. 5.1, but the simulation methods are STAT (Gneiting et al., 2007b), Lave.WVM (Lave et al., 2013), ACM.WVM (Arias-Castro et al., 2014), Perez.WVM (Perez et al., 2011), and Monger (Monger et al., 2016).
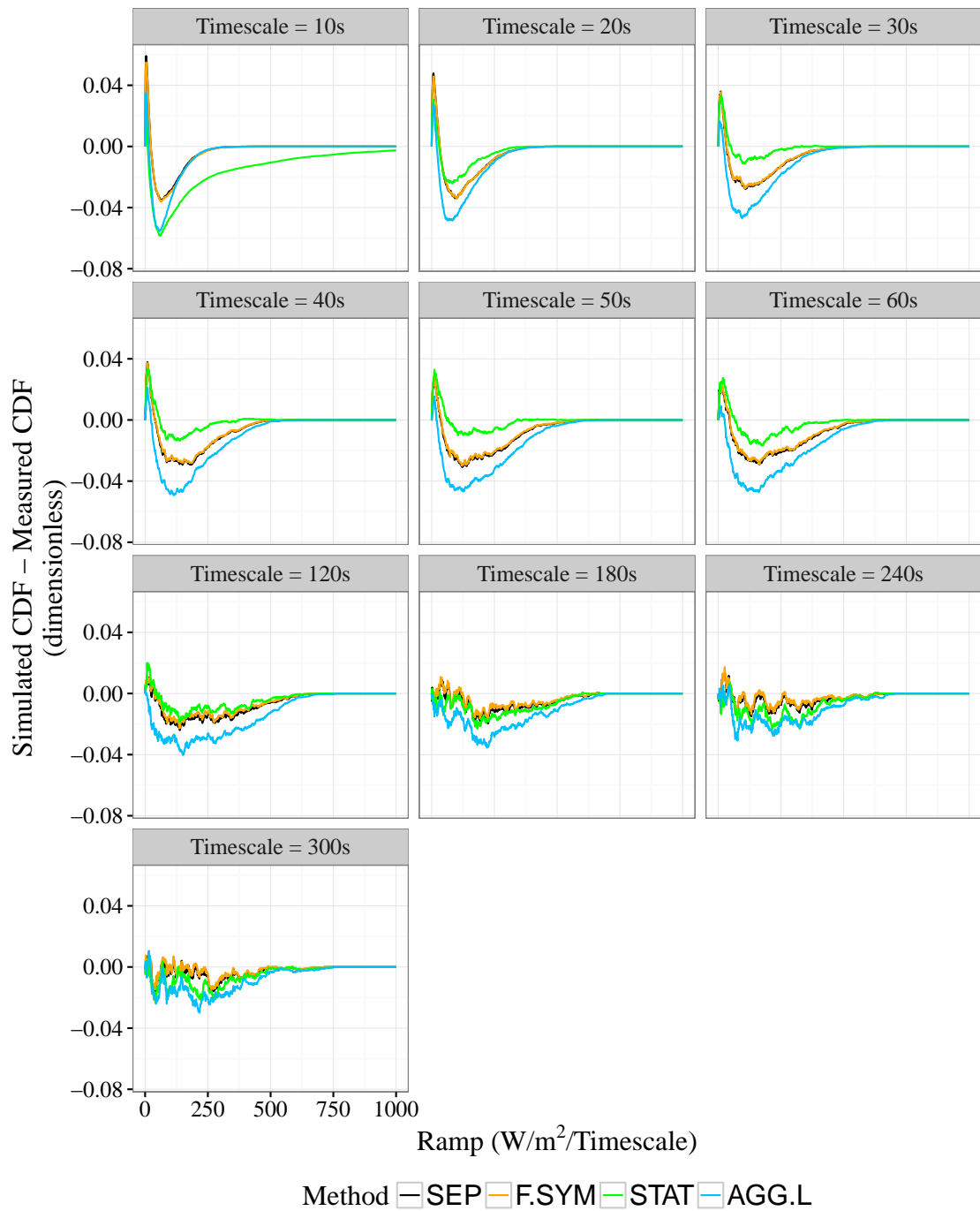
Fig. 5.3: As in Fig. 5.1, but of the category II simulation.
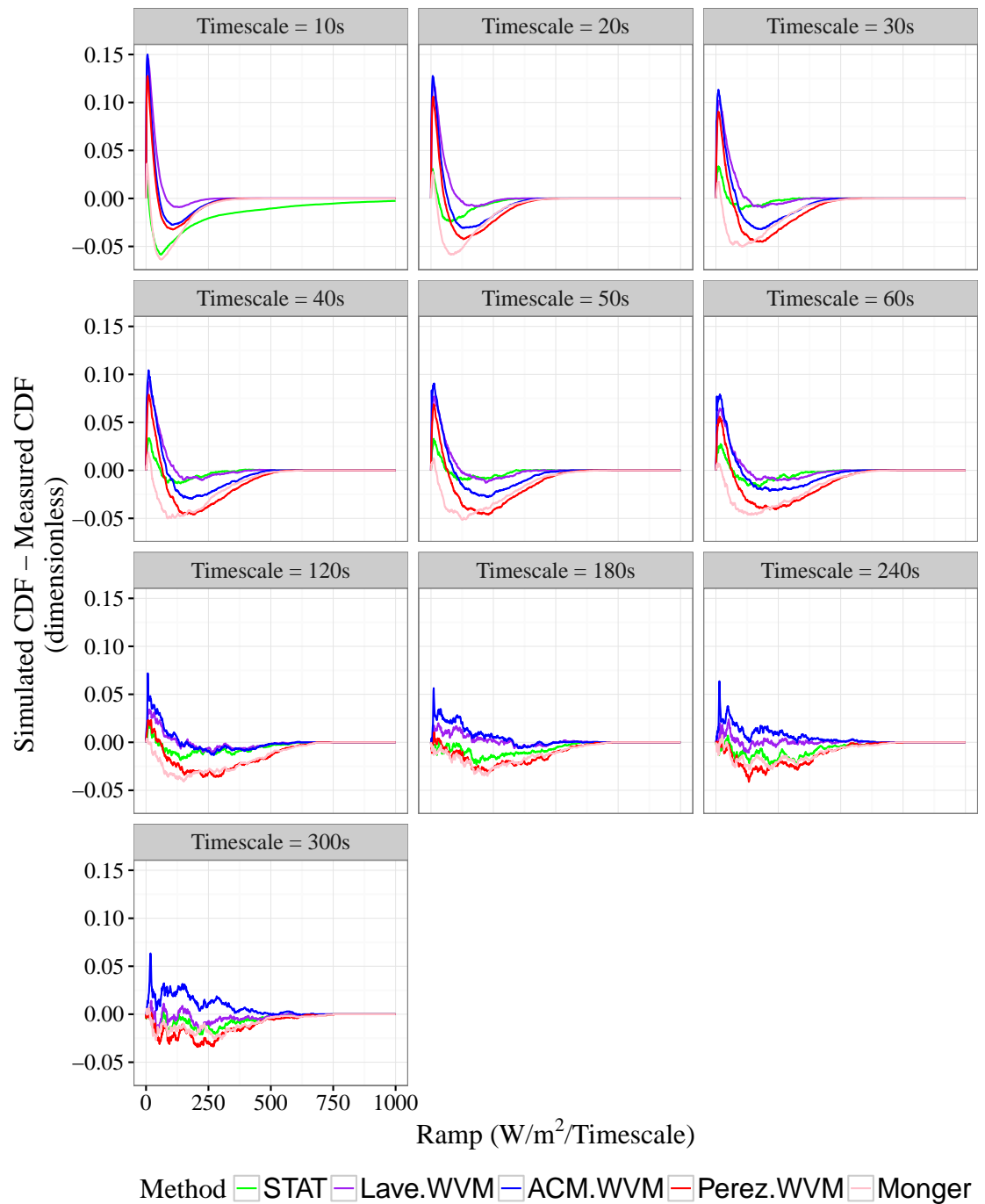
Fig. 5.4: As in Fig. 5.2, but of the category II simulation.
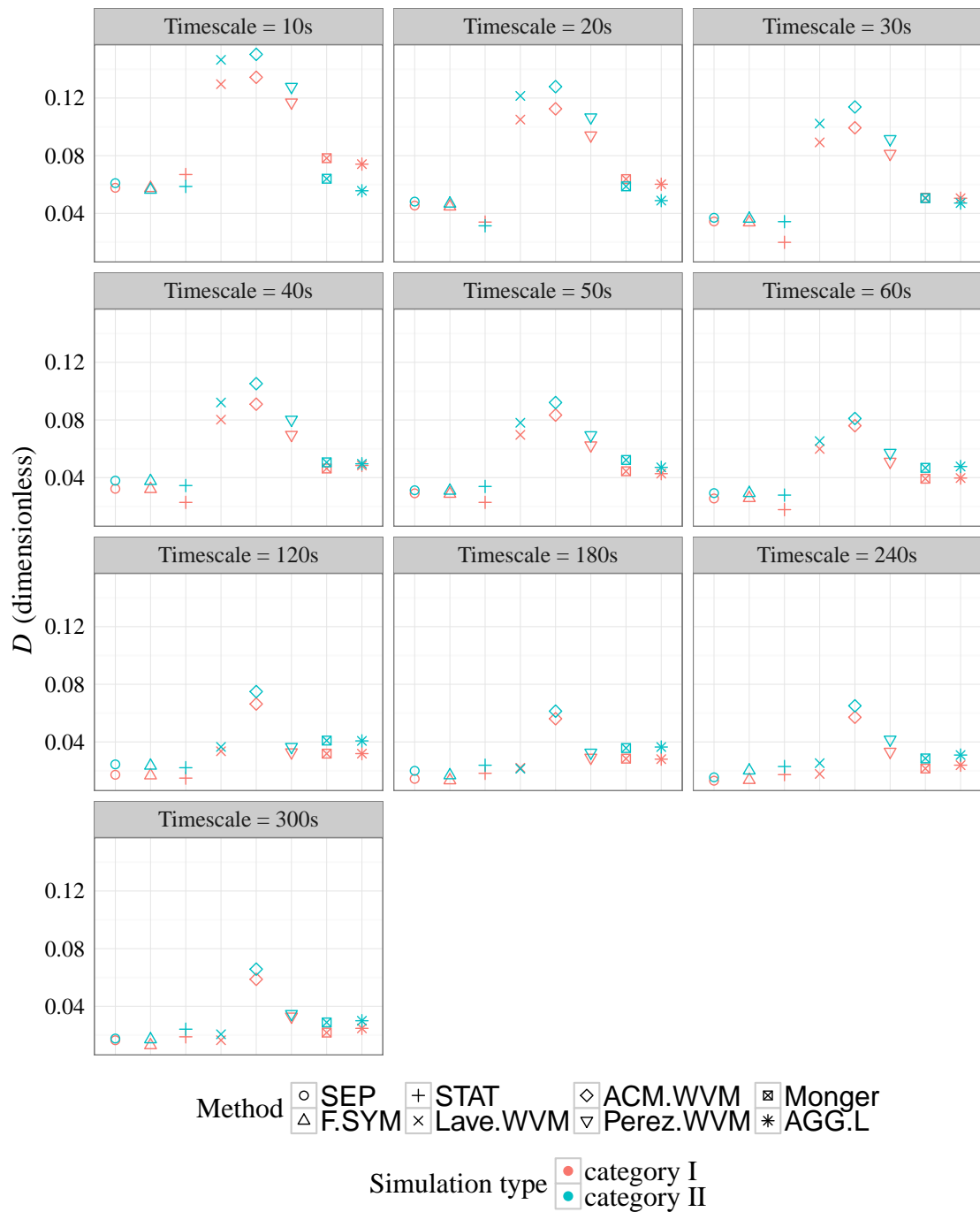
Fig. 5.5: Plots of D values of KS tests for various methods and timescales.

On the other hand, Fig. 5.5 visualizes the comparison of D values of the two-sample Kolmogorov-Smirnov (KS) tests, whose formula is written in Eq. (1.4), of all the methods in various temporal aggregations. Smaller D value signifies that the simulated CDF is more similar to the observed one.

As can be seen from Figs. 5.1 - 5.5, SEP, F.SYM, and STAT kriging methods typically produce ramp events distributions which are the closest to those of the reality. Fig. 5.5 shows that STAT kriging achieves the best results among the three for both aggregated all and testing stations ramp events simulations in some timescales: 20s, 30s, 40s, 60s, and 120s. As the timescale gets longer, the D values of all the methods (except those of ACM.WVM) tend to be similar to each other.

The benchmark method AGG.L appears to perform relatively well, indicating that the geographic smoothing effect caused by aggregating the measurements of all stations are similar with that caused by aggregating only the five learning stations.

5.5 Conclusion

The geographic smoothing effect in a site in Oahu, Hawaii, has been modeled using various methods available in the literature. Days dominated by cumulus clouds are chosen since they possess the most extreme ramp events. Different correlation functions, which are fed into the kriging method, have been shown to give similar results. However, the simulated irradiance time series which are the closest to reality are produced by the model which includes information of dominant wind direction.

Chapter 6

Optimizing solar PV system orientations

6.1 Introduction and literature review

The annual average solar irradiance received on a photovoltaic (PV) panel is affected by its orientation, namely tilt and azimuth angle. In the absence of clouds, the optimum orientation can be derived from astronomical considerations. Local atmospheric tendencies complicate matters, so an orientation derived from astronomy alone will rarely be optimal, although panel tilt and azimuth angles at any site remain strong functions of the site latitude (Lave and Kleissl, 2011). In the continental United States (US), a panel receives the most irradiance when it is tilted southward (Lubitz, 2011) while in Australia, a north-facing panel receives the most (Yan et al., 2013). The optimum tilt increases as the latitude increases. However, this rule does not strictly hold, especially if the solar irradiance in an area is strongly affected by local earth terrain features and in special cases like near the equator. It may happen, for instance, that a high mountain lies at the west of a site which blocks the sun in the afternoon (Lubitz, 2011) or where cloud cover is typically more prominent at a certain time of day due to local meteorological

conditions (Khoo et al., 2014).

Predictions of optimum orientation are usually done using a solar transposition model which predicts the received solar irradiance on tilted angle (I_t) from global horizontal irradiance (GHI) and diffuse horizontal irradiance (DHI) observational data. The optimum panel tilt and azimuth angles are chosen based on the configuration which yields maximum solar radiation over one year. For the diffuse irradiance component, numerous models have been proposed. The Perez model (Perez et al., 1987, 1988, 1990a), adopted in this work, is one such model which has been widely used. Noorian et al. (2008) shows that, in general, this model predicts I_t more accurately than other models.

Lave and Kleissl (2011) produced a PV panel optimization map for the US using the satellite-based spatially-continuous solar radiation data from the National Renewable Energy Laboratory (NREL). These data had been obtained by processing the US Geostationary Operational Environmental Satellites (GOES) imagery using the State University of New York (SUNY) - Albany model. Improvements of the map are possible as uncertainties emerge from modelling process which range from 8 to 25% (Lave and Kleissl, 2011; Wilcox, 2012). More precise ground measurement data, such as NREL's typical meteorological year (TMY) data, can be utilized to generate a more accurate solar radiation and PV panel optimization map. The main challenge is how to integrate spatially-discrete data into spatially-continuous maps.

Previous works attempt to solve similar problems. Zhang et al. (2013) implemented a similar method in a completely different application: to create a more refined surface potential image with submicrometer resolution using a high resolution topography image. Earlier, Qian and Wu (2008) showed integration of high- and low-accuracy experimental data: integration was done in order to improve the crude–but–fast computer simulation results using data derived from more accurate

but computationally intensive models.

By adopting the approach of Qian and Wu (2008) and Zhang et al. (2013) to solve our problem, we can consider satellite images as high-resolution, low-accuracy data and ground measurement results as low-resolution, high-accuracy data. Integration of both data will result in new high resolution data with better accuracy. The actual steps are described below. Optimum PV panel orientation can be calculated for each point in SUNY and TMY data using transposition model. After optimum tilt and azimuth angles of those points have been derived, a universal kriging interpolation can be implemented on TMY data results by incorporating those of SUNY data using ordinary least squares fits. With these steps, we can expect a spatially-continuous map with higher accuracy compared to those derived from solely satellite image.

This chapter is organized as follows. Section 6.2 describes briefly the data used in this chapter. Section 6.3 explains the method used to produce the refined solar and optimum PV orientation maps. Section 6.4 shows the calculation of optimum PV orientation, map generation, and map validation. Sections 6.5 and 6.6 present the transposition model verification and simulation results respectively. The chapter presents a conclusion in section 6.7. Solar transposition and Perez diffuse irradiance model are explained in appendix B.

6.2   Data

As mentioned in section 6.1, the SUNY and TMY data are used in this chapter. The data are obtained from the National Solar Radiation Database (NSRDB) managed by NREL. The details of these data are mentioned in sections 1.5.1.6 and 1.5.1.7. Since SUNY data only cover continental US, in this work, only 925 stations are used. These stations are located within 124ºW to 67ºW and 24ºN to 48ºN. This area covers four time zones, namely Pacific Time Zone, Mountain Time Zone, Central

Fig. 6.1: Locations of TMY stations used in this work

Time Zone, and Eastern Time Zone. The locations of these 925 stations are shown in Fig. 6.1.

6.3 Methods

In this section, we explain the methods used to generate refined solar radation and optimum PV orientation maps after individual optimum orientation in each SUNY and TMY location has been computed. Calculation of individual optimum orientation is presented in section 6.4.1.

The maps are generated based on kriging which is a method to interpolate certain spatially-discrete variable based on random spatial processes (Cressie, 1990; Christou, 2014). This method was proposed empirically by a South African mining engineer, D. G. Krige (Krige, 1951), and developed further by some authors, especially Matheron (1963) and Cressie (1993). There are some variations of kriging,

for instance simple kriging, ordinary kriging, and universal kriging.

In this chapter, we use ordinary kriging and universal kriging since these two methods consider spatial trend in data. It should be noted that the two kriging methods are different from the one in chapters 4 and 5 since, here, we are not dealing with time series data.

6.3.1 Ordinary kriging

Suppose we have $\mathbf{Z} = (Z(\mathbf{s}_1), Z(\mathbf{s}_2), ..., Z(\mathbf{s}_n))^\top$, which is a set of observed data in spatial locations $\mathbf{s}_1$, $\mathbf{s}_2$, ..., $\mathbf{s}_n$, kriging helps us to estimate the value of $Z(\mathbf{s}_0)$ at $\mathbf{s}_0$. The data may be represented in Eq. (6.1).

$$Z(\mathbf{s}_i) = \mu + \mathfrak{d}(\mathbf{s}_i) \tag{6.1}$$

where μ is the mean of $\mathbf{Z}$ and $\mathfrak{d}(\mathbf{s}_i)$ is a zero-mean error process with a determined covariance function.

Kriging interpolation is done using weighted average of observed data. The predictor assumption is written in Eq. (6.2) with Eq. (6.3) as a requirement:

$$\hat{Z}(\mathbf{s}_0) = \mu + \left\{ \sum_{i=1}^{n} \lambda_i \Big(Z(\mathbf{s}_i) - \mu \Big) \right\} \tag{6.2}$$

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{6.3}$$

where $\hat{Z}(\mathbf{s}_0)$ is the predicted value of an unobserved value $Z(\mathbf{s}_0)$ at spatial location $\mathbf{s}_0$ and λ_i is the weight for $Z(\mathbf{s}_i)$ in $Z(\mathbf{s}_0)$ prediction case.

The optimum set of weights $\boldsymbol{\lambda}^*$ is chosen in such a way that it minimizes the mean square prediction error written in Eq. (6.4). In order to satisfy Eq. (6.3), another term is added, and the optimization problem becomes that in Eq. (6.5) with m as a Lagrange multiplier and n is the number of observed points.

$$\boldsymbol{\lambda}^* = \operatorname*{argmin}_{\boldsymbol{\lambda}} F(\boldsymbol{\lambda}) = \operatorname*{argmin}_{\boldsymbol{\lambda}} E\left[Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0) \right]^2 \tag{6.4}$$

$$(\boldsymbol{\lambda}^*, m^*) = \operatorname*{argmin}_{\boldsymbol{\lambda},m} F(\boldsymbol{\lambda}, m) = \operatorname*{argmin}_{\boldsymbol{\lambda},m} \left[E \left[Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0) \right]^2 - 2m \left(\sum_{i=1}^{n} \lambda_i - 1 \right) \right]$$
(6.5)

The objective function of Eq. (6.5) could be further expressed as Eq. (6.6) assuming that $\gamma(\cdot)$ or variogram is expressed as in Eq. (6.7). Here it is assumed that the variogram only depends on the distance between two observation locations (i.e. data are isotropic).

$$2 \sum_{i=1}^{n} \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) - \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) - 2m \left(\sum_{i=1}^{n} \lambda_i - 1 \right)$$
(6.6)

$$2\gamma\left[(\mathbf{s}_i + h) - \mathbf{s}_i\right] = 2\gamma(h) = \operatorname{var}\left[Z(\mathbf{s}_i + h) - Z(\mathbf{s}_i)\right]$$
(6.7)

In order to obtain the values of $\gamma(\cdot)$, empirical variogram $\hat{\gamma}(\cdot)$ needs to be calculated. As indicated in Pebesma (2001), it can be calculated based on Eq. (6.8)

$$\hat{\gamma}\left[\bar{h}_j\right] = \frac{1}{2n(N_j)} \sum_{(j_1, j_2) \in N_j} \left\{ \mathfrak{d}(\mathbf{s}_{j_1}) - \mathfrak{d}(\mathbf{s}_{j_2}) \right\}^2$$
(6.8)

where $\bar{h}_j$ is the average of all distances (h) of pairs in set N_j and $n(\cdot)$ is the number of ordered pairs inside the set N_j. Set N_j contains ordered pairs (j_1, j_2) whose h values fall within the same region of distance tolerance. This averaging process is usually called variogram binning. Binning is required in order to discover a clear trend in a variogram.

The values of $\gamma(\cdot)$ are derived by fitting $\hat{\gamma}(\cdot)$ using an isotropic variogram model. Some examples of these models are linear, exponential, rational quadratic, wave, and power semivariograms (Cressie, 1993). After the function of $\gamma(\cdot)$ has been obtained, the optimum set of weights $(\boldsymbol{\lambda}^*)$ can be calculated. In the next step, Eq. (6.6) is differentiated with respect to $\lambda_1, \lambda_2, \dots, \lambda_n$ and then set to zero (see Eq. (6.9)). Finally, the optimum set of weights could be calculated based on

Eq. (6.10)

$$\gamma(\mathbf{s}_0 - \mathbf{s}_i) - \sum_{j=1}^{n} \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) - m = 0 \tag{6.9}$$

$$\mathbf{\Lambda}_O = \mathbf{\Gamma}_O^{-1} \gamma_O \tag{6.10}$$

where

$$\mathbf{\Lambda}_O = (\lambda_1^*, \ldots, \lambda_n^*, m^*)^\top \tag{6.11}$$

$$\boldsymbol{\gamma}_O = (\gamma(\mathbf{s}_0 - \mathbf{s}_1), \ldots, \gamma(\mathbf{s}_0 - \mathbf{s}_n), 1)^\top \tag{6.12}$$

$$\mathbf{\Gamma}_O = \begin{pmatrix} \gamma(\mathbf{s}_1 - \mathbf{s}_1) & \gamma(\mathbf{s}_1 - \mathbf{s}_2) & \ldots & \gamma(\mathbf{s}_1 - \mathbf{s}_n) & 1 \\ \gamma(\mathbf{s}_2 - \mathbf{s}_1) & \gamma(\mathbf{s}_2 - \mathbf{s}_2) & \ldots & \gamma(\mathbf{s}_2 - \mathbf{s}_n) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(\mathbf{s}_n - \mathbf{s}_1) & \gamma(\mathbf{s}_n - \mathbf{s}_2) & \ldots & \gamma(\mathbf{s}_n - \mathbf{s}_n) & 1 \\ 1 & 1 & \ldots & 1 & 0 \end{pmatrix} \tag{6.13}$$

6.3.2 Universal kriging

Most of the time, μ in Eq. (6.1) is not constant but built of a function which depends on $\mathbf{s}$. In this case, the spatial data are written as in Eq. (6.14):

$$Z(\mathbf{s}_i) = \sum_{j=1}^{p+1} f_{j-1}(\mathbf{s}_i) \upsilon_{j-1} + \mathfrak{d}(\mathbf{s}_i) \tag{6.14}$$

where $f(\mathbf{s}_i)$ is a set of determined functions $\{f_0(\mathbf{s}_i), \ldots, f_p(\mathbf{s}_i)\}$ and $f_0(\mathbf{s}_i)$ usually has a value of 1 to accomodate bias term. This explains the upper value of j in the summation in Eq. (6.14) which is set to $p+1$ instead of p. The values of the series $\boldsymbol{\upsilon}$ are obtained using ordinary least squares fit.

In order to create an unbiased estimation (i.e. expectation of the estimate is the true parameter) (Christou, 2014), apart from Eq. (6.3), Eq. (6.15) must be

satisfied

$$\sum_{i=1}^{n} \lambda_i f_j(\mathbf{s}_i) = \mathbb{E}\left\{f_j(\mathbf{s}_0)\right\} \tag{6.15}$$

The minimization problem and optimum set of weights calculation are presented in Eqs. (6.16) and (6.17) respectively:

$$(\boldsymbol{\lambda}^*, \boldsymbol{m}^*) = \operatorname*{argmin}_{\boldsymbol{\lambda}, \boldsymbol{m}} F(\boldsymbol{\lambda}, \boldsymbol{m}) = \operatorname*{argmin}_{\boldsymbol{\lambda}, \boldsymbol{m}} \left[2\sum_{i}^{n} \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \right.$$
$$\left. -2\sum_{j=1}^{p+1} m_{j-1} \left[\sum_{i=1}^{n} \lambda_i f_{j-1}(\mathbf{s}_i) - f_{j-1}(\mathbf{s}_0) \right] \right] \tag{6.16}$$

$$\boldsymbol{\Lambda}_U = \mathfrak{F}_U^{-1} \gamma_U \tag{6.17}$$

where

$$\boldsymbol{\Lambda}_U = (\lambda_1^*, \dots, \lambda_n^*, m_0^*, \dots, m_p^*)^\top \tag{6.18}$$

$$\boldsymbol{\gamma}_U = (\gamma(\mathbf{s}_0 - \mathbf{s}_1), \dots, \gamma(\mathbf{s}_0 - \mathbf{s}_n), f_0(\mathbf{s}_0), \dots, f_p(\mathbf{s}_0))^\top \tag{6.19}$$

and $\mathfrak{F}_U$ is a symmetric matrix with dimension $(n + p + 1) \times (n + p + 1)$ as represented in Eq. (6.20):

$$\mathfrak{F}_U = \left(\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{B}^\top & \mathbf{0} \end{array} \right) \tag{6.20}$$

where

$$\mathbf{A} = \begin{pmatrix} \gamma(\mathbf{s}_1 - \mathbf{s}_1) & \gamma(\mathbf{s}_1 - \mathbf{s}_2) & \gamma(\mathbf{s}_1 - \mathbf{s}_3) & \dots & \gamma(\mathbf{s}_1 - \mathbf{s}_n) \\ \gamma(\mathbf{s}_2 - \mathbf{s}_1) & \gamma(\mathbf{s}_2 - \mathbf{s}_2) & \gamma(\mathbf{s}_2 - \mathbf{s}_3) & \dots & \gamma(\mathbf{s}_2 - \mathbf{s}_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(\mathbf{s}_n - \mathbf{s}_1) & \gamma(\mathbf{s}_n - \mathbf{s}_2) & \gamma(\mathbf{s}_n - \mathbf{s}_3) & \dots & \gamma(\mathbf{s}_n - \mathbf{s}_n) \end{pmatrix} \tag{6.21}$$

$$\mathbf{B} = \begin{pmatrix} f_0(\mathbf{s}_1) & f_1(\mathbf{s}_1) & \dots & f_p(\mathbf{s}_1) \\ f_0(\mathbf{s}_2) & f_1(\mathbf{s}_2) & \dots & f_p(\mathbf{s}_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(\mathbf{s}_n) & f_1(\mathbf{s}_n) & \dots & f_p(\mathbf{s}_n) \end{pmatrix} \quad (6.22)$$

The empirical variogram for universal kriging is the same as the one in Eq. (6.8), except that $\mathfrak{d}(\mathbf{s}_i)$ follows Eq. (6.14).

6.4 Optimum PV orientation, map generation, and map validation

6.4.1 Computation of optimum panel orientation

Optimum panel orientation, consisting of optimum panel tilt (ω_{opt}) and azimuth angles (β_{opt}), was calculated for each point in SUNY and TMY data by utilizing `optim` command in the R statistical software (The R Core Team, 2014). The objective of the optimization is to maximize the value of annual solar radiation ($E_{\mathrm{t.opt}}$), which is the summation of I_{t} for a period of one year. The calculation of I_{t} is presented in appendix B. The adopted optimization method is limited-memory Broyden–Fletcher–Goldfarb–Shanno (BFGS) with box constraints (Byrd et al., 1995; Nocedal and Wright, 1999).

6.4.2 Map generation and validation

In this work, three kinds of map are generated, namely ω_{opt}, β_{opt}, and $E_{\mathrm{t.opt}}$ maps. For each kind, there are four maps produced. The first maps were generated using solely SUNY data results. The second and third maps were derived from ordinary kriging (OK) and universal kriging (UK) of TMY data results. The fourth maps are similar as the third ones, except for the SUNY data results which were incorporated

as part of the features ($f(\mathbf{s}_i)$ in Eq. (6.14)) of the universal kriging (UK.SUNY). From this point onwards, based on the derivation method, the first, second, third, and fourth maps are referred as SUNY, OK, UK, and UK.SUNY maps respectively.

The R package automap (Hiemstra et al., 2008) is adopted for kriging operations. This package makes use of gstat package (Pebesma, 2001, 2004). The features used in UK maps are chosen from geographical information (longitude and latitude) based on Akaike information criterion (AIC). This feature selection method is explained in detail in Akaike (1974) and Venables and Ripley (2002). In R, model selection using AIC can be performed using step command (The R Core Team, 2014). As mentioned, for UK.SUNY maps, besides geographical information, SUNY data results (ω_{opt}, β_{opt}, and $E_{\text{t.opt}}$) are also taken into account.

In these simulations, as stated previously, there are only 925 out of 1,020 TMY stations that can be used since the rests are located outside the SUNY map area. The accuracy of each map is measured using mean bias error (MBE), relative mean bias error (rMBE), root mean square error (RMSE), and relative root mean square error (rRMSE). For OK, UK, and UK.SUNY maps, the errors are computed using cross validation, namely kriging are done 925 times where each time 924 stations are used as learning data, leaving one station for validation.

6.5 Validation of transposition model

6.5.1 Validation using SRRL observation data

The accuracy of irradiance transposition model (which includes the Perez diffuse irradiance model) is verified using measurement data provided by NREL which were taken at the Solar Radiation Research Laboratory (SRRL) in Golden, Colorado (39.74°N, 105.18°W) (NREL, 2014a). It consists of hourly-averaged GHI, DHI, I_t on a 40°tilted panel facing southward, and I_t on a 2-axis tracking panel. Here, the
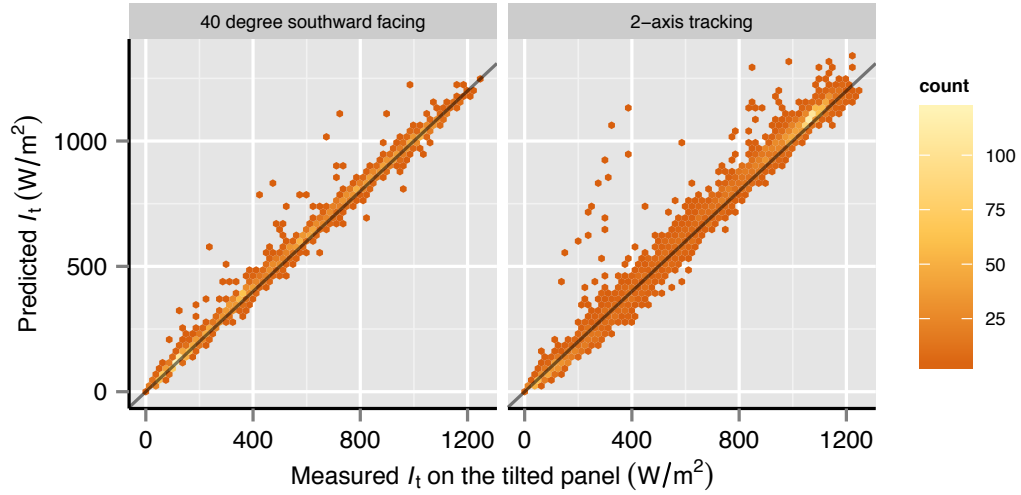
Fig. 6.2: Scatter plots of estimation and measurement values of I_t

data used are the ones of 13 October 2013 until 12 October 2014. The measurement devices are as follows. Kipp & Zonen CM 22 pyranometers measured the GHI and DHI values. Eppley PSP pyranometer measured the tilted global irradiance. Kipp & Zonen CM21 pyranometer which was attached on Licor LI-2020 solar tracker measured the global irradiance of tracking panel (Lave and Kleissl, 2011; NREL, 2014a).

Verification was done only on data with solar zenith angle (θ_z) smaller than 80°. Following Gueymard (2009), the albedo is fixed at 0.2. This assumption is applied since individual instantenous albedo is not provided in each SUNY and TMY data location. The plots of the measured and calculated I_t on the fixed and tracking panel are shown in Fig. 6.2. The rRMSE values of the fixed tilted and 2-axis tracking configurations are found as 5.68% and 8.17% respectively. These errors are found to be comparable with the ones in Gueymard (2009) and Lave and Kleissl (2011) and thus we note it is justifiable to proceed with this tranposition method.

6.5.2   Validation using PVWatts calculator

The PVWatts Calculator (NREL, 2014b) is a web application developed by NREL to produce quick computation about solar energy potential in some areas worldwide. It adopts the NREL TMY and SolarAnywhere data for continental US. In other parts of the world, the following databases are used: Solar and Wind Energy Resource Assessment Programme (SWERA), The ASHRAE International Weather for Energy Calculations Version 1.1 (IWEC), and Canadian Weather for Energy Calculations (CWEC).

Here, PVWatts Calculator is used to benchmark the transposition model in five points in the US, namely: Barstow-Daggett Airport (California/CA), Quillayute State Airport (Washington/WA), Denver International Airport (Colorado/CO), Madison Dane County Regional Airport (Wisconsin/WI), and Jacksonville International Airport (Florida/FL). These sites are chosen based on the variations in terms of climates and geographical conditions (Lubitz, 2011). Moreover, the TMY3 data of these sites are available in the web application. The annual solar irradiance on panel with optimum fixed orientation and 2-axis tracking system according to PVWatts Calculator and our calculation are shown in Table 6.1. The optimum fixed orientation are chosen based on the algorithm explained in Section 6.4.1. The 2-axis tracking system is simulated by assuming the panel is always perpendicular to the sun. In Table 6.1 and the rests of the chapter, panel azimuth angles (β) are stated in degrees where 0° and 90° refer to the north and east direction respectively.

The obtained rMBE values are within 1.6% and shows that the implemented model is consistent with the established one.

Table 6.1: Validation of Perez transposition model using PVWatts calculator

Location name	Longitude (°W)	Latitude (°N)	Altitude (m)	ω_{opt} (°)	β_{opt} (°)	Annual global radiation on panel with optimum fixed orientation (MWh/m²)			Annual global radiation on 2-axis tracking panel (MWh/m²)		
						Perez	PVWatts	rMBE (%)	Perez	PVWatts	rMBE (%)
Barstow-Daggett Airport, CA	116.800	34.850	586	32.69	178.37	2.40	2.40	-0.03	3.37	3.41	-1.39
Quillayute State Airport, WA	124.567	47.933	55	35.74	190.84	1.27	1.28	-0.07	1.60	1.61	-1.06
Denver International Airport, CO	104.650	39.833	1650	37.19	174.49	2.00	2.01	-0.48	2.67	2.71	-1.59
Madison Dane County Regional Airport, WI	89.330	43.130	262	34.78	179.12	1.64	1.65	-0.33	2.10	2.12	-0.85
Jacksonville International Airport, FL	81.700	30.500	8	27.73	177.36	1.82	1.82	0.02	2.35	2.36	-0.57

6.6 Simulation results

Figs. 6.3 – 6.5 show the maps of ω_{opt}, β_{opt}, and $E_{\text{t.opt}}$ respectively which are obtained using SUNY and TMY data. The OK, UK, and UK.SUNY maps in those figures are generated using the 925 TMY stations shown in Fig. 6.1. The SUNY results seem identical with those obtained by Lave and Kleissl (2011). As mentioned there, SUNY map of Fig. 6.4 shows discontinuity in four states, namely Montana, Wyoming, Colorado, and New Mexico. Possibly, it occurs due to time shift errors for evening irradiance values of SUNY satellite (Nottrott and Kleissl, 2010).
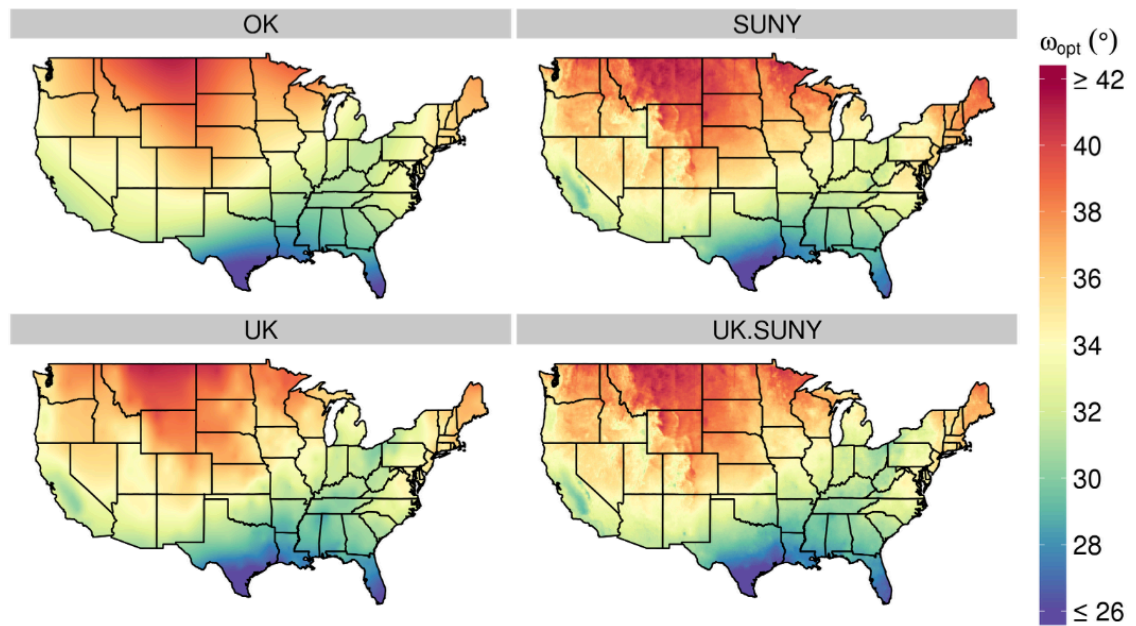
Fig. 6.3: ω_{opt} maps: OK map is derived from TMY data using ordinary kriging; UK map is derived from TMY data using universal kriging, SUNY map is derived from SUNY data, and UK.SUNY map is derived from TMY and SUNY data using universal kriging.

The error values of each map, obtained from cross validation mentioned in section 6.4.2, are shown in Tables 6.2 – 6.4. The tables show that maps which utilize ground measurement data possess lower bias error than those of purely

Table 6.2: Validation of ω_{opt} maps

Map	MBE (°)	rMBE (%)	RMSE (°)	rRMSE (%)
SUNY	0.569	1.714	1.240	3.737
OK	-0.001	-0.002	1.221	3.679
UK	0.003	0.008	1.157	3.485
UK.SUNY	0.003	0.009	1.060	3.193
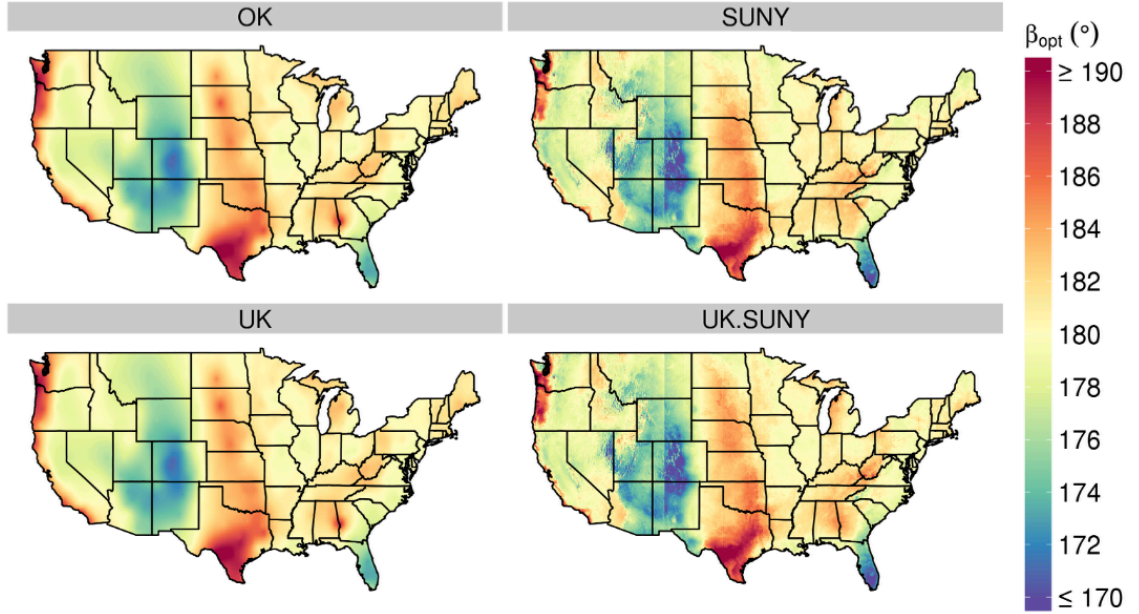
Fig. 6.4: Same as Fig. 6.3, but for β_{opt} maps

SUNY results. In terms of RMSE, UK.SUNY maps consistently reach the lowest errors among the four methods although there are only small error differences.

The difference of absolute error values of SUNY and UK.SUNY maps are compared in Figs. 6.6 – 6.8. Figs. 6.6 and 6.8 show that corrections for ω_{opt} and $E_{\text{t.opt}}$ mainly occur in the northeastern US while there is not so much changing at the west part of the country in terms of these two parameters. Fig. 6.8 shows that the proposed method is able to correct some $E_{\text{t.opt}}$ values in the Florida state. In Fig. 6.7, obvious improvements of β_{opt} occur in the coastline of the Oregon state and the west part of the Pennsylvania state.

Table 6.3: Validation of β_{opt} maps

Map	MBE (°)	rMBE (%)	RMSE (°)	rRMSE (%)
SUNY	-0.160	-0.089	2.096	1.159
OK	0.006	0.003	2.383	1.318
UK	0.006	0.003	2.383	1.318
UK.SUNY	0.007	0.004	2.063	1.141
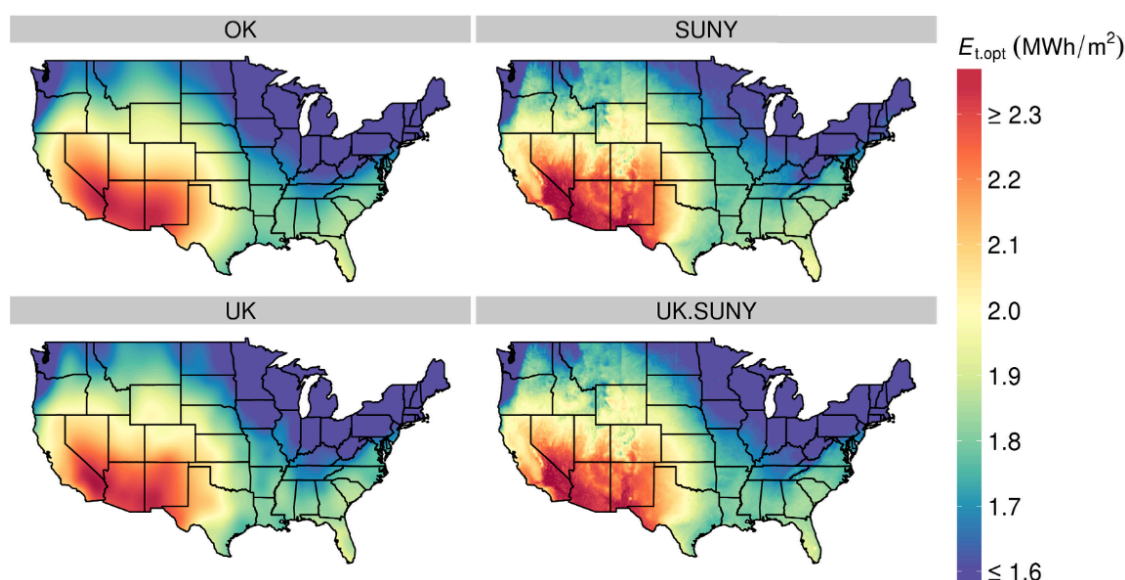
Fig. 6.5: Same as Fig. 6.3, but for $E_{\text{t.opt}}$ maps

6.7 Conclusion

A method to integrate satellite-derived SUNY data and ground-based TMY data to produce more accurate US solar radiation and optimum photovoltaic panel orientation maps have been proposed. The method is able to eliminate bias error of the same kinds of map derived using solely SUNY data. Furthermore, maps produced by the proposed method consistently have less RMSE compared to purely satellite-based and ground-measurement-based maps. The error differences, however, are not huge.

Table 6.4: Validation of $E_{\text{t.opt}}$ maps

Map	MBE (Wh/m^2)	rMBE (%)	RMSE (Wh/m^2)	rRMSE (%)
SUNY	32,369.284	1.824	82,186.966	4.632
OK	-98.625	-0.006	81,124.407	4.572
UK	-226.607	-0.013	79,709.121	4.493
UK.SUNY	11.465	0.001	72,002.220	4.058
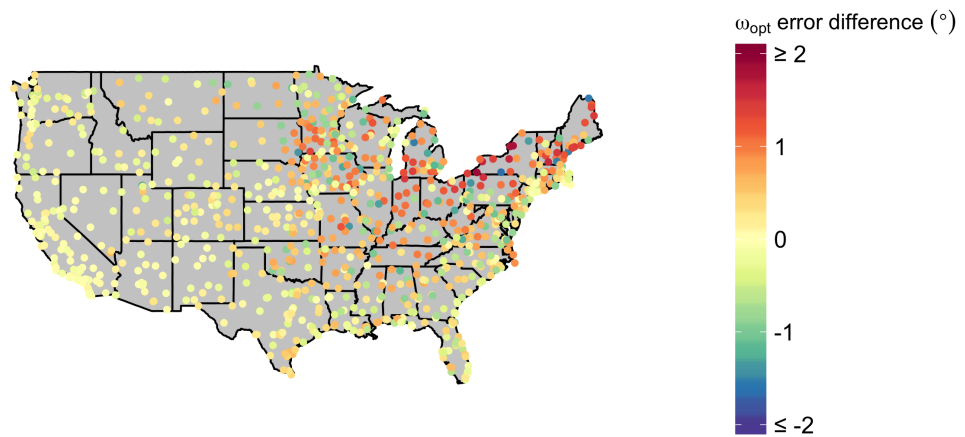
Fig. 6.6: Absolute error difference between SUNY and UK.SUNY ω_{opt} maps
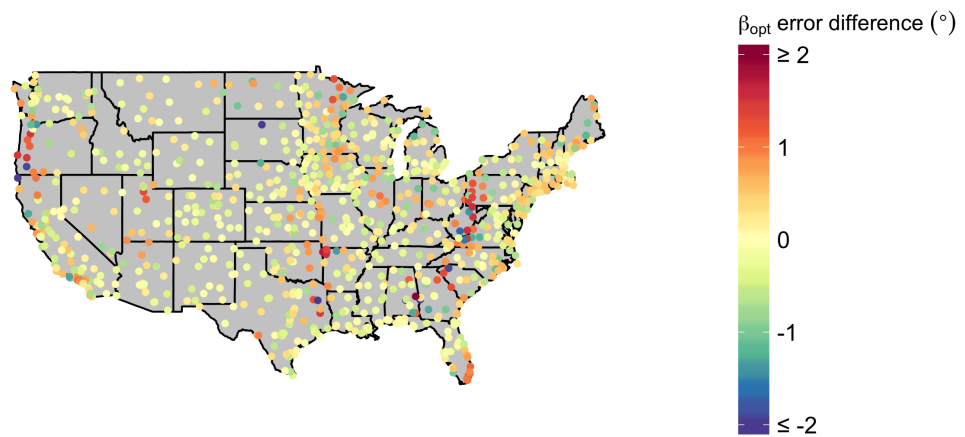
Fig. 6.7: Same as Fig. 6.6, but for β_{opt} maps
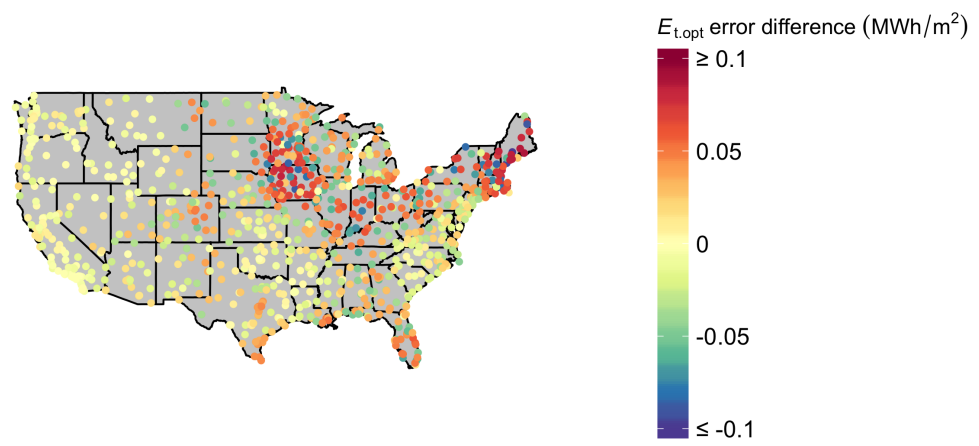
Fig. 6.8: Same as Fig. 6.6, but for $E_{\text{t.opt}}$ maps

Chapter 7

Numerical weather prediction applied to Singapore irradiance forecasts

7.1 Introduction

As previously stated in this thesis, electric power grid operators need quantitative meteorological forecasts to mitigate the impact of the variability of energy generated by variable renewable sources, solar in particular, on a variety of timescales (Pelland et al., 2013b). Previous chapters of this thesis have considered short and very-short-term forecasts. Grid management forecasting on such timescales is presently only possible using statistical methods of the type applied in this thesis. For longer forecasting horizons, however, it is possible to formulate physical models of the atmosphere, and to evolve them according to the relevant physical

This chapter is based on: A. W. Aryaputera, D. Yang, and W. M. Walsh, "Day-ahead solar irradiance forecasting in a tropical environment," *Journal of Solar Energy Engineering*, 137(5):051009, 2015. doi: http://doi.org/10.1115/1.4030231.

models (Inman et al., 2013). Such numerical weather prediction (NWP) methods are generally adopted for that beyond ∼1 hour (Diagne et al., 2013) and indeed typically are updated every 6, 12 or 24 hours. In this chapter, day–ahead solar irradiance forecasting in a tropical environment, namely, Singapore, is investigated in the context of currently available NWP models.

While short-term statistical forecasts are required in the process of instantaneous balancing of supply and demand of the electricity grid, longer-term forecasts are required for generation and transmission planning, and by electricity futures markets. Singapore does not share in the need for long distance transmission planning, but as the amount of local PV increases, generation balance between the competing companies will require management. Similarly the forthcoming futures market will look towards forecasting for both demand and supply forecasts. Such considerations often take place on a 24 hour cycle.

For day–ahead solar irradiance forecasting, NWP models have already been shown as the most promising approaches to be viable for operational deployment (Lara-Fanego et al., 2012). NWP models build and evolve physical models of the atmosphere on selected sub-regions (Bjerknes, 1904). The success of an NWP model relies on the knowledge about the initial state of the atmosphere and the physical laws which govern the evolution of the atmosphere. Input data may be derived from a combination of global models and observational data.

There is a rich literature on NWP–based irradiance forecasting with many studies conducted using different NWP models and over a variety of locations. Zamora et al. (2005) compared the outputs of the Fifth-Generation Penn State/National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5) with observational data from the United States. Similar studies have been conducted in Spain using both the MM5 and the Weather Research and Forecasting (WRF) model (Ruiz-Arias et al., 2008; Lara-Fanego et al., 2012). Lorenz et al. (2009) considered

the European centre for medium–range weather forecasts (ECMWF) model while Mathiesen and Kleissl (2011) compared the North American model (NAM), the global forecast system (GFS), and the ECMWF model.

The state of the art of NWP-based solar irradiance forecasting appears to be based on statistical correction (Lorenz et al., 2009; Diagne et al., 2014; Pelland et al., 2013a) or machine learning (Cornaro et al., 2015) on multiple instances of a single model or on an ensemble of different models. For WRF NWP models, concerns remain about the best physical assumptions and atmospheric modeling made within the models, particularly in regard to solar irradiance forecasting, for which the models have generally not been specifically designed or optimized. Local geographical conditions can significantly impact the forecasting skill of NWP models (Inman et al., 2013). Most works in this field have considered mid–latitude sites, particularly Germany, Spain, and the United States (US) due to the concentration of renewable energy facilities there. Although it is still limited, there is also an example of such work for a sub-tropical location (Diagne et al., 2014).

In a tropical environment such as Southeast Asia, there are some reports on the performance of NWP (He et al., 2013; Tursilowati et al., 2011; Vaid, 2013; Li et al., 2013; Laux et al., 2013; Chotamonsak et al., 2011), but none of these specifically addressed the forecasting of global horizontal irradiance (GHI), which is required for PV power generation purposes. To address the issue of how well NWP performs in the tropics, this chapter tests the reliability of the Advanced Research WRF (ARW) model (Skamarock et al., 2008) to forecast day–ahead GHI in Singapore.

In solar engineering, the physical laws of motion and thermodynamics that characterize the atmosphere are rarely scrutinized in detail. As NWP models output hundreds of parameters in each run, including irradiance, researchers simply run NWP models and study their outputs (Lorenz et al., 2009; Mathiesen and Kleissl,

2011; Perez et al., 2013). However, the forecast errors of the NWP models are often compared only to the persistence model (persistence forecasts assume tomorrow's irradiance is identical to today's (Lara-Fanego et al., 2012)) due to the lack of other benchmarking models. We therefore consider seasonal stochastic models in this chapter. More specifically, exponential smoothing (ETS) models and seasonal autoregressive integrated moving average (seasonal ARIMA) models are used to perform day–ahead solar irradiance forecasting.

Stochastic methods are frequently used for intra–hour irradiance forecasting (Yang et al., 2012; Dong et al., 2013; Yang et al., 2014a; Dong et al., 2014). However, seasonal models have rarely been considered. If we consider an hourly irradiance time series $\{y_t\}$, the diurnal cycle due to the Earth's rotation can be modeled accurately using seasonal stochastic models. As statistical models possess significant advantage in computational speed as compared to the NWP models, the performance of such models is therefore worth investigating.
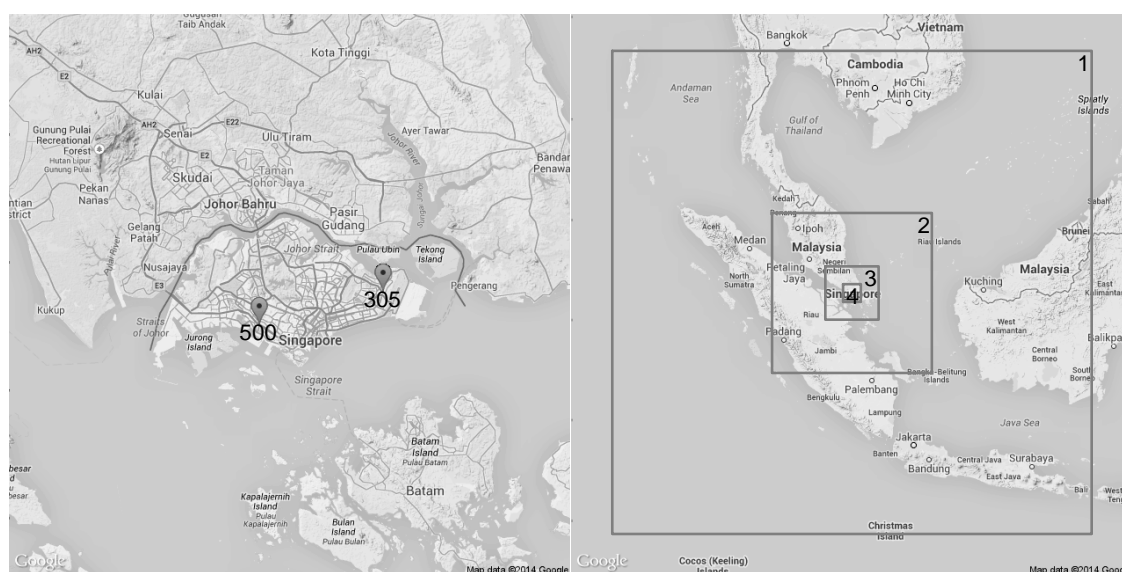
Fig. 7.1: (left) Map of Singapore. Stations 305 and 500 are the same as stations 5 and 2 respectively in Fig. 1.3. (right) WRF simulation domain. Map data source: Google Maps.

Before we discuss the WRF forecasts and the stochastic models in sections 7.2

and 7.3 respectively, the ground measurement data used in this chapter is described.

7.1.1 Observation data

GHI data from two stations, namely, S305 (1.353°N, 103.965°E) and S500 (1.301°N, 103.771°E), maintained by the Solar Energy Research Institute of Singapore (SERIS) are used as the "true" GHI for performance evaluation of various forecasting models. The locations of the stations are shown in Fig. 7.1 (left). An industrial standard pyranometer CMP11 from Kipp & Zonen is installed horizontally at each site.

The experiment is performed on data from four different periods in the year 2013, namely, 24 January to 7 February (period I), 24 April to 8 May (period II), 24 July to 7 August (period III), and 24 October to 7 November (period IV). The choices of day fit the seasonal phases in Singapore, which are northeast monsoon season (December-early March), inter-monsoon period (late March-May), southwest monsoon season (June-September), and inter-monsoon period (October-November) (National Environment Agency, 2009). Apart from that, only hourly–averaged data from 8AM to 6PM Singapore time (SGT) are considered in the data analysis, so as to exclude periods of greatest solar zenith angle.

7.2 Numerical weather prediction method

7.2.1 WRF model

The WRF mesoscale NWP models have been developed since the late 1990s by the NCAR, National Oceanic and Atmospheric Administration (NOAA), and other institutes in the US. WRF has been used widely for real–time forecasting and analysis of historical weather phenomena. Currently, there are two types of WRF dynamical cores, which are developed separately, namely the ARW model by the NCAR, and Nonhydrostatic Mesoscale Model (NMM) by NOAA and National Centers for

Environmental Prediction (NCEP).
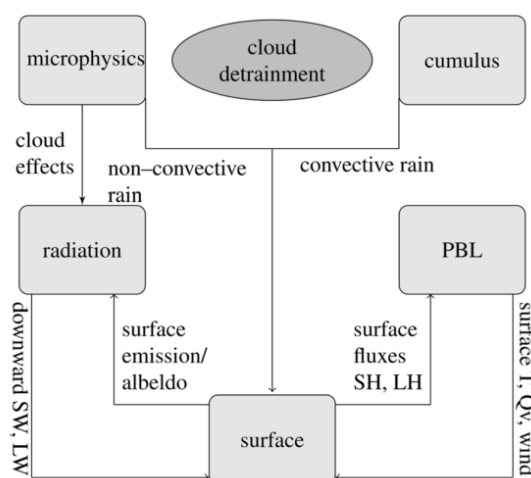
Fig. 7.2: Interactions of WRF parameterizations (adopted from Dudhia (2014))

7.2.2 Simulation setup

The simulations are conducted using the ARW model version 3.5.1 (Skamarock et al., 2008). The Global Forecast System (GFS) model from NCEP is utilized as the simulation boundary condition. The temporal and spatial resolutions are 3 hours and $0.5° \times 0.5°$ respectively. The GFS data used are the ones of 12 PM UTC (or 8 PM SGT) on the previous day. The simulation spin up time is 11.5 hours. Spin up is necessary for the calculation of the microphysics (He et al., 2013).

The domain selection is shown in Fig. 7.1 (right). The geographical data resolutions for the domains are 10 minutes, 5 minutes, 2 minutes, and 30 seconds from the largest to the smallest domain. In the same sequence, the grid resolutions are 27, 9, 3, and 1 km. Two way nesting is adopted in order to enable exchange of information between domains during simulation. The physics configuration setup is shown in Table 7.1.

The WRF schemes are related with each other as depicted in Dudhia (2014) and Fig. 7.2. The cumulus physics determines the cloud detrainment which affects

Table 7.1: Physics configuration setup

Parameter	Scheme Selection
Microphysics	WRF Single–Moment 6–class (Hong and Lim, 2006)
Longwave Radiation	Rapid Radiative Transfer Model (RRTM) (Mlawer et al., 1997)
Shortwave Radiation	Dudhia (Dudhia, 1989)
Surface Layer	MM5 Similarity (Paulson, 1970)
Land Surfaces	Noah Land Surface Model (Chen, 2007)
Planetary Boundary Layer	Yonsei University (Hong et al., 2006)
Cumulus Parameterization	Grell–Freitas scheme (Grell and Freitas, 2013) (except for domain 4 which does not use any scheme)

the microphysics scheme. Cumulus and microphysics are responsible for convective and non convective rain respectively. Besides, the microsphysics scheme is responsible in the formation of the clouds and thus influences the longwave (LW) and shortwave (SW) radiation schemes. The solar radiation affects the surface layer physics. The surface layer scheme controls the surface radiation emission phenomena, which in turn affects the radiation scheme, and the heat exchange between the earth surface and the atmosphere, which influences the planetary boundary layer (PBL) physics. Heat exchange consists of sensible and latent heat (SH and LH). PBL itself determines the temperature (T) and moisture (Qv) in the lower troposphere, which is located in the lowermost part of the earth atmosphere. In the tropics, the troposhere may reach a thickness of 16 km. In this layer, most weather phenomena occur (Gregory et al., 2009).

The physics configurations, in particular the radiation, land surfaces, and PBL, are chosen according to He et al. (2013), since it utilized the WRF model for Singapore rain forecast. Dudhia's model is chosen as the shortwave scheme due to its efficiency in the computation of radiation absorption and scattering in cloudy and clear sky conditions (Hummon, 2014). Rapid Radiative Transfer Model (RRTM) is implemented as the longwave scheme since it generates detailed absorption spectrum by considering water vapour, carbon dioxide, methane, nitrous oxide, and ozone (Ruiz-Arias et al., 2008). However, since this work focuses on solar en-

ergy, the shortwave scheme is more important (Ruiz-Arias et al., 2008). The choice of PBL scheme is critical since, as mentioned, it determines the properties of the lower troposhere (He et al., 2013). For simulation which requires high resolution as in the case of Singapore, the Yonsei University scheme (YSU) is considered to be efficient (He et al., 2013). YSU is a revised algorithm of vertical diffusion which was developed from Medium-Range Forecast (MRF) PBL by Hong and Pan in 1996 (Hong and Pan, 1996; Hong et al., 2006). The problem of the old model is it often overestimates the mixing when the wind is strong (Hong et al., 2006).

The Grell–Freitas (GF) scheme is adopted as the cumulus/convection scheme of this work since it has been tested on the tropics (Grell and Freitas, 2013; Grell et al., 2013). This scheme was new in WRF version 3.5. It was developed from the Grell and Devenyi's model (Grell and Freitas, 2013) which implemented stochastic approach. Aerosols effect is included by incorporating cloud condensation nuclei dependent cloud water to rain conversion and cloud drop evaporation (Grell and Freitas, 2013).

The WRF Single–Moment 6–class is implemented as the microphysics scheme instead of the Thompson scheme used by He et al. (2013) since the latter one sometimes overestimates cloud cover. The former one has been made in operation by the NCAR (Wang et al., 2008).

7.3 Stochastic processes

As alternatives to the WRF model, two stochastic methods are introduced in this section to perform day–ahead forecasting, namely, the ETS model and the seasonal ARIMA model. Unlike the WRF model, stochastic methods consider the irradiance from a statistical point of view. More specifically, these methods aim at identifying the properties and the stochastic nature of an irradiance time series.

7.3.1 Exponential smoothing

ETS considers time series as a combination of various components, namely, the trend ($\mathcal{T}$), seasonal ($\mathcal{S}$) and error ($\mathcal{E}$) components. Furthermore, the trend component can be further separated into the level (ℓ) and growth (b) components. The state space formulation is constructed using different combinations of these components.

The seasonal component can be either additive ($\mathcal{T} + \mathcal{S}$) or multiplicative ($\mathcal{T} \times \mathcal{S}$) to the trend component. As the irradiance transient fluctuates about a well–known clear sky model, the additive formulation is more appropriate than the multiplicative one. $\mathcal{T}$ is modeled using the level and growth components, following Hyndman et al. (2002):

$$
\begin{aligned}
\text{None}: \quad & \mathcal{T}_h = \ell \\
\text{Additive}: \quad & \mathcal{T}_h = \ell + bh \\
\text{Additive damped}: \quad & \mathcal{T}_h = \ell + (\phi + \phi^2 + \cdots + \phi^h)b \\
\text{Multiplicative}: \quad & \mathcal{T}_h = \ell b^h \\
\text{Multiplicative damped}: \quad & \mathcal{T}_h = \ell b^{(\phi + \phi^2 + \cdots + \phi^h)}
\end{aligned}
$$

where $\mathcal{T}_h$ is forecast trend over the next h time steps, $0 < \phi < 1$ is a damping parameter. Based on these five models, the formulae for ETS is given in Table 7.2. In each case, $\widehat{y}_{t+h|t}$ denotes an h–steps–ahead forecast using information up to time t; ℓ_t denotes the level; b_t denotes the slope; s_t denotes the seasonal components and m denotes the number of seasons in the data. α^*, β^*, γ and ϕ are coefficients; $\phi_h = \phi + \phi^2 + \cdots \phi^h$ and $h_m^+ = [(h - 1) \bmod m] + 1$.

For each case shown in Table 7.2, the error component can be either additive, $(\mathcal{T} + \mathcal{S}) + \mathcal{E}$, or multiplicative, $(\mathcal{T} + \mathcal{S}) \times \mathcal{E}$. This leads to 10 different ETS models. Their state space formulation can be found in Hyndman et al. (2008).

Table 7.2: Formulae for recursive calculations and point forecasts.

$$
\begin{array}{ll}
\text{N} &
\begin{aligned}
\ell_t &= \alpha^*(y_t - s_{t-m}) + (1-\alpha^*)\ell_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1}) + (1-\gamma)s_{t-m} \\
\widehat{y}_{t+h|t} &= \ell_t + s_{t-m+h_m^+}
\end{aligned} \\[2em]
\text{A} &
\begin{aligned}
\ell_t &= \alpha^*(y_t - s_{t-m}) + (1-\alpha^*)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \\
\widehat{y}_{t+h|t} &= \ell_t + hb_t + s_{t-m+h_m^+}
\end{aligned} \\[2em]
\text{A}_d &
\begin{aligned}
\ell_t &= \alpha^*(y_t - s_{t-m}) + (1-\alpha^*)(\ell_{t-1} + \phi b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1-\gamma)s_{t-m} \\
\widehat{y}_{t+h|t} &= \ell_t + \phi_h b_t + s_{t-m+h_m^+}
\end{aligned} \\[2em]
\text{M} &
\begin{aligned}
\ell_t &= \alpha^*(y_t - s_{t-m}) + (1-\alpha^*)\ell_{t-1}b_{t-1} \\
b_t &= \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma(y_t - \ell_{t-1}b_{t-1}) + (1-\gamma)s_{t-m} \\
\widehat{y}_{t+h|t} &= \ell_t b_t^h + s_{t-m+h_m^+}
\end{aligned} \\[2em]
\text{M}_d &
\begin{aligned}
\ell_t &= \alpha^*(y_t - s_{t-m}) + (1-\alpha^*)\ell_{t-1}b_{t-1}^\phi \\
b_t &= \beta^*(\ell_t/\ell_{t-1}) + (1-\beta^*)b_{t-1}^\phi \\
s_t &= \gamma(y_t - \ell_{t-1}b_{t-1}^\phi) + (1-\gamma)s_{t-m} \\
\widehat{y}_{t+h|t} &= \ell_t b_t^{\phi_h} + s_{t-m+h_m^+}
\end{aligned}
\end{array}
$$

7.3.2 Seasonal ARIMA

We introduce the Box and Jenkins method (Box et al., 2011) by considering the ARMA(p, q) model:

$$
y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{7.1}
$$

where $\phi_1 \cdots \phi_p$ are the autoregressive parameters to be estimated and $\theta_1 \cdots \theta_q$ are the moving average parameters to be estimated. The random process $\{a_t\}$ describers the random shocks from each time step t, it has zero mean and constant variance. The ARMA model has process order (p, q), which indicates that the forecast is a linear combination of observations from p previous steps and the errors from q previous steps.

If we define a backshift operator $\mathcal{B}$ so that $\mathcal{B}y_t = y_{t-1}$ and more generally $\mathcal{B}^j y_t = y_{t-j}$, Eq. (7.1) can be written as:

$$
(1 - \phi_1\mathcal{B} - \cdots - \phi_p\mathcal{B}^p)y_t = (1 - \theta_1\mathcal{B} - \cdots - \theta_q\mathcal{B}^q)a_t \tag{7.2}
$$

This may be abbreviated even further by writing:

$$
\phi(\mathcal{B})y_t = \theta(\mathcal{B})a_t \tag{7.3}
$$

where

$$\phi(\mathcal{B}) = 1 - \phi_1\mathcal{B} - \phi_2\mathcal{B}^2 - \cdots - \phi_p\mathcal{B}^p \tag{7.4}$$

$$\theta(\mathcal{B}) = 1 - \theta_1\mathcal{B} - \theta_2\mathcal{B}^2 - \cdots - \theta_q\mathcal{B}^q \tag{7.5}$$

To extend the stationary ARMA model to the ARIMA model which can handle non–stationary time series, we define the difference operator $\nabla = 1 - \mathcal{B}$ or, more generally, $\nabla^d = (1 - \mathcal{B})^d$. We also find the first differences of the series, i.e., $z_t = y_t - y_{t-1}$. This can be re–written as:

$$z_t = (1 - \mathcal{B})y_t = \nabla y_t \tag{7.6}$$

If an ARMA model is constructed using the differenced time series $\{z_t\}$, i.e, $\phi(\mathcal{B})z_t = \theta(\mathcal{B})a_t$, it is equivalent to the ARIMA$(p, 1, q)$ model constructed using the original time series y_t, i.e., $\phi(\mathcal{B})\nabla y_t = \theta(\mathcal{B})a_t$. Generalization of the equation gives the ARIMA(p, d, q) model:

$$\phi(\mathcal{B})\nabla^d y_t = \theta(\mathcal{B})a_t \tag{7.7}$$

where p, d and q are process orders.

The non–seasonal model in Eq. (7.7) can be extended to a seasonal model. Suppose the time series $\{y_t\}$ has s time periods in a cycle, we can define $\mathcal{B}^s y_t = y_{t-s}$. Thus, a seasonal ARIMA model with process order $(p, d, q) \times (P, D, Q)_s$ is:

$$\phi(\mathcal{B})\Phi(\mathcal{B}^s)\nabla^d\nabla_s^D y_t = \theta(\mathcal{B})\Theta(\mathcal{B}^s)a_t \tag{7.8}$$

For example, seasonal ARIMA$(0, 1, 0) \times (1, 0, 1)_s$ model is:

$$\Phi(\mathcal{B}^s)\nabla^d y_t = \Theta(\mathcal{B}^s)a_t$$

$$(1 - \Phi_1\mathcal{B}^s)(y_t - y_{t-1}) = (1 - \Theta_1\mathcal{B}^s)a_t$$

$$y_t - y_{t-1} - \Phi_1 y_{t-s} + \Phi_1 y_{t-s-1} = a_t - \Theta_1 a_{t-s}$$

In other words, the forecast at time t, $\widehat{y}_t$, is found via:

$$y_t = y_{t-1} + \Phi_1 y_{t-s} - \Phi_1 y_{t-s-1} + a_t - \Theta_1 a_{t-s} \tag{7.9}$$

7.3.3   Model selection and parameter estimation

To forecast the irradiance on a particular day, we need to select one model from
Table 7.2 which best describes the historical data. Similarly model selection is
required for seasonal ARIMA implementation. There are many ways to perform a
model selection using the given data. As our forecasting application is operational,
automatic model identification and parameter estimation is essential. We use the
Akaike information criterion (AIC) (Ricci, 2005) for model selection:

$$\text{AIC} = -2 \ln \mathcal{L} + 2k \tag{7.10}$$

where $\mathcal{L}$ is the likelihood function for the model and k is the number of parameters
in the model. For a detailed description on the likelihood calculation, we refer the
readers to Hyndman et al. (2008). The idea of AIC is to maximize the likelihood
and penalize the model complexity. Therefore, the model with the smallest AIC
should be chosen. After a particular model is selected based on the AIC, maximum
likelihood estimation is used to obtain the model parameters.

7.4   Results and discussions

In order to validate the performance of WRF, the observational data are categorized
according to daily clear sky index. The indices are calculated by following Perez
et al. (1990b) as shown in Eq. (7.11):

$$k_{\text{T.daily}} = \frac{\int_{8\text{h}}^{18\text{h}} I \, dt}{\int_{8\text{h}}^{18\text{h}} I_{\text{o}} \, dt} \approx \frac{\sum\limits_{8\text{h}}^{18\text{h}} I}{\sum\limits_{8\text{h}}^{18\text{h}} I_{\text{o}}} \tag{7.11}$$

where $k_{\text{T.daily}}$ is the daily clear sky index, I is the hourly–averaged GHI, and I_{o}
is the extraterrestrial irradiance. Furthermore, as proposed in Lara-Fanego et al.

(2012), when $k_{\text{T.daily}}$ is less than 0.4, a particular day is marked as overcast; when $k_{\text{T.daily}}$ is in between 0.4 and 0.65, a day is considered as cloudy; finally, clear sky condition is indicated by index greater than 0.65. The extraterrestrial irradiance is given by:

$$I_{\text{o}} = I_{\text{sc}} \left(1 + 0.033 \cos \frac{360 d_{\text{n}}}{365} \right) \cos \theta_{\text{z}} \qquad (7.12)$$

where I_{sc} is the solar constant (1367 W/m^2), d_{n} is the day sequence in the year, and θ_{z} is the solar zenith angle.

Forecast performance is evaluated by calculating the root mean square error (RMSE) and relative root mean square error (rRMSE) according to Eqs. (1.1) and (1.2) respectively.

For persistence forecasting, rather than adopting the previous day hourly irradiance values, we follow the method proposed in Perez et al. (2013); Beyer et al. (2009) in which daily clear sky index of the previous day multiplied by the forecast day I_{o} is taken to predict the forecast day irradiance. The latter method has been proven to have higher accuracy compared to the earlier one.

In order to compare different forecasting results from various locations (which indeed have different persistence forecasting accuracy), forecast skill (FS) parameter (Eq. (1.3)) is considered.

7.4.1 Relation between WRF simulation grid resolution and accuracy

As illustrated in Fig. 7.1, the simulation has four domains with different resolutions. Since the Singapore area is covered by all these domains in the simulation, the GHI forecasting result from each domain could be compared. The plots of the RMSE from different domains are visualized in Fig. 7.3. The results show that as the resolution becomes more refined, the accuracy of the WRF gets higher. Thus, from this point onwards, the result of domain 4, which has the highest resolution, will
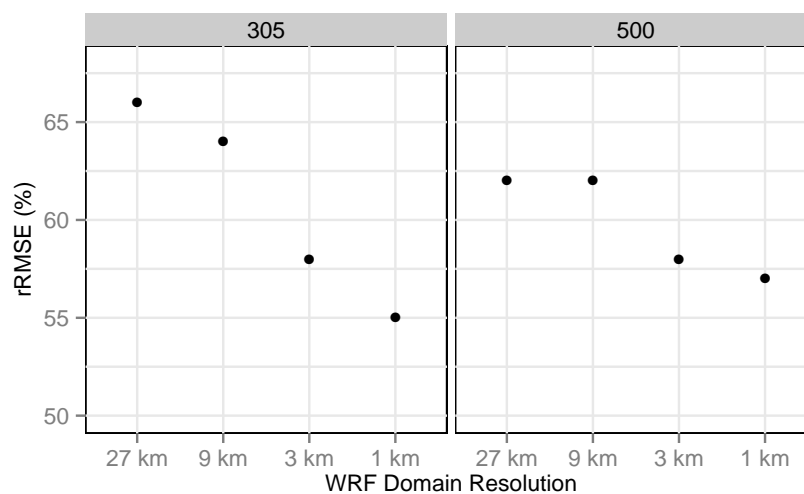
Fig. 7.3: Comparison between rRMSE and simulation grid resolutions

be used to represent the WRF performance.

7.4.2 Evaluation of WRF simulation results

Tables 7.3 and 7.4 show the persistence method, seasonal ARIMA, ETS, WRF, and average of ETS and WRF output errors for all the stations as a function of sky condition and time of the year. Overall, it is shown that seasonal ARIMA produces the worst forecast. Meanwhile, the performances of persistence, ETS, and WRF forecasts are comparable. Averaging the forecasting result of ETS and WRF yields the lowest error. The overall RMSE of this method in the two stations is 49%.

There is a correlation between the accuracy of persistence method with those of the other methods. In period I and IV, the overall RMSE values of persistence method are lower than those in the other two periods. This indicates larger variance between one day to the next one in period I and IV. This pattern also occurs in the RMSE of the other methods.

In order to compare different forecasting result in different places with different persistence accuracy, as mentioned in the beginning of section 7.4, FS can be

Table 7.3: GHI forecast errors for station 305

Sky Condition (No of Days)	Pers. RMSE	Seasonal ARIMA RMSE	ETS RMSE	WRF RMSE	mean (ETS, WRF) RMSE
Period I					
CS (5)	128 (21)	210 (35)	**124 (21)**	228 (38)	163 (27)
CL (6)	211 (40)	220 (42)	**175 (34)**	260 (50)	192 (37)
OC (4)	277 (111)	353 (142)	330 (133)	**212 (85)**	249 (100)
OA (15)	209 (44)	260 (55)	215 (45)	237 (50)	**200 (42)**
Period II					
CS (1)	**151 (24)**	468 (76)	419 (68)	209 (34)	313 (51)
CL (4)	322 (70)	330 (72)	277 (60)	**163 (35)**	209 (45)
OC (10)	222 (86)	**173 (67)**	199 (77)	242 (94)	199 (77)
OA (15)	249 (74)	252 (75)	242 (72)	222 (66)	**211 (63)**
Period III					
CS (1)	484 (86)	514 (91)	258 (46)	**158 (28)**	203 (36)
CL (8)	202 (45)	273 (60)	**151 (33)**	192 (42)	159 (35)
OC (6)	230 (89)	231 (89)	258 (99)	**206 (79)**	211 (81)
OA (15)	242 (63)	280 (73)	208 (54)	196 (51)	**185 (48)**
Period IV					
CS (0)	NA	NA	NA	NA	NA
CL (11)	191 (46)	211 (50)	196 (47)	207 (50)	**183 (44)**
OC (4)	182 (59)	**167 (55)**	196 (64)	201 (65)	188 (61)
OA (15)	189 (49)	200 (51)	196 (50)	206 (53)	**184 (47)**
Annual					
CS (7)	220 (37)	317 (53)	213 (36)	217 (36)	**197 (33)**
CL (29)	221 (48)	250 (55)	194 (43)	210 (46)	**183 (40)**
OC (24)	228 (86)	226 (85)	240 (91)	222 (84)	**210 (79)**
OA (60)	224 (57)	250 (63)	216 (55)	216 (55)	**195 (49)**

CS = Clear Sky, CL = Cloudy, OC = Overcast, OA = Overall
RMSE values are stated in absolute values (W/m^2) and relative values (in the brackets, in percentage). The lowest RMSE for each category is in bold.

utilized. Comparison is done with those results of Perez et al. (2013) since it used the same persistence forecasting method as the one applied in this chapter (refer to the beginning section 7.4). The comparison of FS (mean of WRF and ETS) of station 305 and 500 of this chapter and those of WRF models from various locations in the world are shown in Table 7.5. The table shows that FS values for cities in Spain are relatively higher than the rests indicating better forecast in those locations. The lower FS values for Singapore can be explained as follows. Although the extraterrestrial solar irradiance is fairly constant troughout the year in the tropics, rapid cloud transformations cause the solar irradiance to vary considerably from

Table 7.4: GHI forecast errors for station 500

Sky Condition (No of Days)	Pers. RMSE	Seasonal ARIMA RMSE	ETS RMSE	WRF RMSE	mean (ETS, WRF) RMSE
Period I					
CS (1)	188 (33)	313 (53)	198 (35)	**135 (24)**	162 (29)
CL (9)	150 (31)	279 (59)	149 (31)	140 (29)	**117 (25)**
OC (5)	238 (96)	262 (106)	248 (100)	198 (80)	**194 (79)**
OA (15)	186 (46)	276 (68)	191 (47)	161 (40)	**150 (37)**
Period II					
CS (0)	NA	NA	NA	NA	NA
CL (6)	283 (60)	361 (77)	252 (54)	**158 (34)**	175 (37)
OC (9)	220 (85)	**195 (75)**	230 (89)	324 (125)	256 (98)
OA (15)	247 (72)	273 (79)	239 (69)	270 (79)	**227 (66)**
Period III					
CS (1)	204 (37)	387 (70)	**180 (33)**	344 (63)	251 (46)
CL (9)	237 (54)	238 (54)	174 (39)	160 (36)	**151 (34)**
OC (5)	249 (93)	**219 (82)**	258 (96)	271 (101)	243 (91)
OA (15)	239 (61)	245 (63)	206 (53)	219 (56)	**194 (50)**
Period IV					
CS (1)	237 (41)	328 (57)	263 (46)	238 (41)	**230 (40)**
CL (9)	**172 (43)**	226 (56)	185 (46)	219 (55)	179 (44)
OC (5)	179 (59)	289 (95)	205 (68)	148 (49)	**143 (47)**
OA (15)	179 (47)	256 (67)	198 (52)	200 (53)	**172 (45)**
Annual					
CS (3)	**211 (37)**	344 (61)	217 (38)	254 (45)	218 (39)
CL (33)	210 (47)	272 (61)	188 (42)	173 (39)	**156 (35)**
OC (24)	223 (83)	236 (88)	235 (88)	260 (97)	**221 (83)**
OA (60)	215 (57)	263 (69)	209 (55)	216 (57)	**188 (49)**

CS = Clear Sky, CL = Cloudy, OC = Overcast, OA = Overall
RMSE values are stated in absolute values (W/m^2) and relative values (in the brackets, in percentage). The lowest RMSE for each category is in bold.

one day to the other (Ye et al., 2013). This is expected from a tropical climate (according to Köppen-Geiger classification). As a consequence, of the WRF-modeled regions, Singapore is the cloudiest (Rubel and Kottek, 2010) and the most difficult to model. In addition, tropical area is dominated by the occurences of local and mesoscale weather phenomena (Laing and Evans, 2011). Another challenge comes from the sparse weather sensor network in this area (Laing and Evans, 2011) which affects the initial input parameters/boundary conditions of NWP model.

Table 7.5: FS comparison of various locations in the world

Location	FS
Station 305, Singapore	0.14
Station 500, Singapore	0.14
Córdoba, Spain	0.18
Granada, Spain	0.25
Huelva, Spain	0.27
Desert Rock, US	0.14
Goodwin Creek, US	0.12
Penn State, US	0.17
Fürstenzell, Germany	0.27
Stuttgart, Germany	0.19
Würzburg, Germany	0.07
Linz, Austria	0.10
Vienna, Austria	0.18

The forecast results of Spain, US, Germany, and Austria are taken from Perez et al. (2013). For Singapore, the results are obtained from the mean of WRF and ETS forecasting results.

7.5 Conclusion

This chapter has evaluated the performance of day-ahead GHI forecasting using the WRF model in Singapore. Observational data obtained from ground sensors at two stations are used as references. Persistence forecasting is adopted as a benchmark.

In general the WRF model performs better than the seasonal ARIMA, but does not outperform persistence and ETS methods. Averaging the forecast result of ETS and WRF yields the lowest annual rRMSE which is 49% in both meteorological stations.

The limitations of the forecasting methods described in this chapter point to the need for at least two main improvements in NWP forecasts in Singapore (and presumably similar tropical climate zones). One need is for greatly enhanced observational data to inform the models so that appropriate initial and boundary conditions can be selected. Singapore, being surrounded by ocean and developing countries with low density meteorological networks, is simply not able to inform the NWP with the level of detail available from the thousands of observational stations present in Europe and North America, where the NWP techniques originate and

are most effectively applied. This observation points to the need for novel data acquisition and data processing techniques to be developed if Singapore and other tropical sites are to have state-of-the-art NWP forecasts available operationally.

The second identified need is that the NWP models themselves are most likely not optimized for the - challenging - task of small scale cloud formation and hence irradiance forecasting. Nevertheless the work presented here is a step towards the optimal deployment of NWP for irradiance forecasting given the current state of meteorological networks and NWP algorithms.

Chapter 8

Ensemble techniques for probabilistic forecasts

8.1 Introduction

In this thesis, various methods of solar energy forecasting have been proposed and each method is suitable for a particular range of forecast horizons. Numerical weather prediction (NWP) is generally adopted for that beyond ~1 hour (Diagne et al., 2013). In operational forecasting, it has become customary to derive probabilistic forecasts from NWP results since predicted ranges of a meteorological variable are typically more useful than just point forecasts (Sloughter et al., 2010). To achieve this, there are at least two categories of probabilisitic forecasting methods, namely those which blend NWP outputs from a number of different models (Raftery et al., 2005; Sloughter et al., 2010; Gneiting et al., 2005) and those which employ forecasting results from one model (Alessandrini et al., 2015) which may

---

This chapter is partly based on: A. W. Aryaputera, H. Verbois, and W. M. Walsh, "Probabilistic accumulated solar irradiance forecast for Singapore using ensemble techniques," 43rd IEEE Photovoltaic Specialists Conference, Portland, Oregon, US, 2016 (oral presentation).

be evolved with different initial and/or boundary conditions. Blending several outputs of numerical models, whether from single or multiple NWPs (for single NWP, ensemble members can be gathered by utilizing past forecasts), has been observed to produce better probabilistic forecasts due to spread-skill relationships (Whitaker and Loughe, 1998) although this statement is not universally acknowledged (Zamo et al., 2014).

The work in this chapter attempts to evaluate intra-day probabilistic forecasts of accumulated solar irradiance in Singapore using several ensemble methods: Bayesian model averaging (BMA) (Raftery et al., 2005), ensemble model output statistics (EMOS) (Gneiting et al., 2005), and analog ensemble (AnEn) (Delle Monache et al., 2013). Originally, besides historical observation data, the inputs of the first two methods are forecast outputs from several NWP models while the latest one only requires those from single NWP model although in this chapter, outputs from several NWPs are also fed into the AnEn algorithm, as mentioned by Delle Monache et al. (2013). In line with Sloughter et al. (2010, 2007), for the BMA and EMOS methods, the effect of the selection of predictive probability density function (PDF), normal and skew-normal, on the forecasting results is investigated. The skew-normal PDF is chosen since it suits the PDF of the observation data.

The performance of the ensemble methods are compared with less sophisticated models, such as climatology (CLIM) (Gneiting et al., 2005), autoregressive integrated moving average (ARIMA) (Hyndman, 2016; Hyndman and Khandakar, 2008), exponential smoothing (ETS) (Hyndman, 2016; Hyndman and Khandakar, 2008), and bias-corrected ensemble (BCE) (Veenhuis, 2013).

The chapter is structured as follows. Section 8.1 provides a short introduction on the motivations and literature reviews. Section 8.2 describes the observation data, including their distribution, the NWP outputs used as inputs to the ensemble

methods, and the BMA, EMOS, and AnEn algorithms. Section 8.3 explains the list of methods tested in this work, choice on the ensemble size (for AnEn) and training period length (for the rests of the methods), and the assessment on the forecasting results. Finally, this chapter is closed with a conclusion in section 8.4.

8.2 Data and methods

8.2.1 Data

To validate the forecasting results, measurement data from an island-wide network of 22 silicon irradiance sensors (section 1.5.1.1; Fig. 1.3) is utilized. Stations 17 and 22 are excluded due to insufficient data during the chosen period. The data used in this chapter span from 27 February 2014 to 23 January 2016.

The GHI data of the 22 stations are spatially averaged and temporally accumulated for 0800-1400 and 0800-1800 Singapore time (SGT) for each day. In the latter parts of this chapter, the two types of accumulated data are called half- and full-day accumulated irradiance respectively. The spatial averaging of the 1-minute data is done only when at least 20 stations have available data in a particular time step, while the temporal accumulation is done only when there is 100% availability of spatially-averaged 1-minute data within the respective time range (0800-1400 or 0800-1800 SGT). This procedure results in 696 useable days.

As mentioned in section 1.5.2, The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE) is the source of the NWP outputs fed into the ensemble methods. The European Centre for Medium-Range Weather Forecast (ECMWF), Japan Meteorological Agency (JMA), and Korea Meteorological Administration (KMA) are selected based on the data availability and literature review on their deterministic forecast accuracies (Thorey et al., 2015). The numbers of the perturbed members of these

global models are 50, 26, and 23 for ECMWF, JMA, and KMA respectively. The particular NWP outputs which are used in this chapter are released every day at 0000 Coordinated Universal Time/UTC (0800 SGT) and have spatial resolution of $0.25° \times 0.25°$. Among some other meteorological parameters, the outputs comprise of accumulated irradiance within 0000-0600 and 0000-1200 UTC. In this work, these serve as the inputs for half- and full-day accumulated irradiance forecasting respectively. These raw forecast values are derived from weighted-average of forecast outputs from pixels containing the observation site. The weights are comparable with the number of sites lying on each of those pixels.

It is assumed here that the low spatial resolution of the NWP outputs will be compensated by the low temporal resolution of the forecasting products, which are half- and full-day accumulated irradiance.

8.2.2 Skew-normal distribution

In this section, the observation data PDF is investigated following Sloughter et al. (2010) as an effort to achieve probabilistic forecasts which are better calibrated. Four quarters of the observation data estimated kernel densities are shown in Fig. 8.1. Those are obtained using the function `density` of the R statistical software (R Core Team, 2015). The PDFs appear to be tilted although the tilts are less visually obvious for those of the full-day accumulated irradiance. Hence, as stated in section 8.1, it is decided to investigate the performances of BMA and EMOS with skew-normal PDF as their basis in addition to those of BMA and EMOS which adopt normal PDF. In this chapter, the first two methods are notated as BMA.s and EMOS.s, while the last two methods are notated as BMA.n and EMOS.n.

The PDF, mean (μ), and standard deviation (σ) of skew-normal PDF $SN\left(\xi, \mathfrak{w}^2, \alpha\right)$ can be expressed as

$$f(x) = \frac{2}{\mathfrak{w}} \varphi\left(\frac{x - \xi}{\mathfrak{w}}\right) \Phi\left(\alpha \frac{x - \xi}{\mathfrak{w}}\right) \tag{8.1}$$
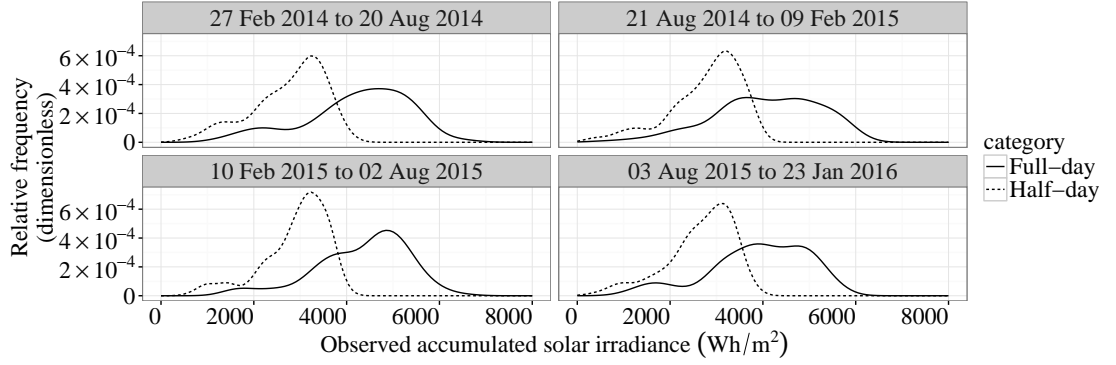
Fig. 8.1: Estimated kernel density of four quarters of the half-day and full-day accumulated solar irradiance derived from the SERIS sensor network data.

$$\mu = \xi + \mathfrak{w}\mho\sqrt{2/\pi} \tag{8.2}$$

$$\sigma = \sqrt{\mathfrak{w}^2\left(1 - 2\mho^2/\pi\right)} \tag{8.3}$$

$$\mho = \alpha/\sqrt{1 + \alpha^2} \tag{8.4}$$

where α, $\mathfrak{w}$, and ξ signify the slant, scale, and location parameters respectively, while $\Phi\left(\cdot\right)$ and $\varphi\left(\cdot\right)$ denote the cumulative density function (CDF) and PDF of normal distribution $N\left(0,1\right)$ respectively (Azzalini, 2014, 2015).

8.2.3 BMA

The BMA.s algorithm is explained as follows. Like in BMA.n (Raftery et al., 2005), a conditional PDF $g_k\left(y|\zeta_k\right)$ is assigned to each member forecast ζ_k. Thus, the BMA predictive PDF can be expressed as

$$p\left(y|\zeta_1,\cdots,\zeta_K\right) = \sum_{k=1}^{K} w_{\mathrm{BMA\_}k}g_k\left(y|\zeta_k\right) \tag{8.5}$$

where $g_k\left(y|\zeta_k\right)$ is the variable y conditional PDF given forecast ζ_k when ζ_k is the best forecast in the ensemble, $w_{\mathrm{BMA\_}k}$ is the weight for member forecast ζ_k, and K is the number of member forecasts (Sloughter et al., 2010). For BMA.s, skew-normal PDF is adopted for each conditional PDF whose mean is at the bias-corrected

forecast, $a_{\text{BMA\_}k} + b_{\text{BMA\_}k}\zeta_k$:

$$y|\zeta_k \sim SN\left(\xi_k, \mathfrak{w}^2, \alpha\right) \tag{8.6}$$

$$\xi_k = -\mathfrak{w}\mho\sqrt{2/\pi} + (a_{\text{BMA\_}k} + b_{\text{BMA\_}k}\zeta_k) \tag{8.7}$$

where variables $a_{\text{BMA\_}k}$ and $b_{\text{BMA\_}k}$ are estimated by fitting forecast values on the corresponding observation values in the learning data. Meanwhile, the slant variable α is chosen based on the fitting of skew-normal PDF on the estimated kernel density of the learning data, mentioned in section 8.2.2. The conditional PDF σ is optimized using the expectation-maximization (EM) algorithm as in Raftery et al. (2005) with some modifications in order to accomodate variables $\mathfrak{w}$ and ξ.

The E step of the EM algorithm is

$$\hat{z}_{kt}^{(j)} = \frac{w_{\text{BMA\_}k}^{(j-1)}g\left(y_t|\zeta_{kt}, \xi_{kt}^{(j-1)}, \mathfrak{w}^{2(j-1)}, \alpha\right)}{\sum_{i=1}^{K} w_{\text{BMA\_}i}^{(j-1)}g\left(y_t|\zeta_{it}, \xi_{it}^{(j-1)}, \mathfrak{w}^{2(j-1)}, \alpha\right)} \tag{8.8}$$

where $z_{kt}^{(j)}$ is the introduced unobserved quantity, j indicates the jth iteration in the EM algorithm, and $g\left(y_t|\zeta_{kt}, \xi_{kt}^{(j-1)}, \mathfrak{w}^{2(j-1)}, \alpha\right)$ is a skew-normal PDF with mean $a_{\text{BMA\_}k} + b_{\text{BMA\_}k}\zeta_{kt}$, scale parameter $\mathfrak{w}^{(j-1)}$, location parameter $\xi_{kt}^{(j-1)}$, and slant parameter α, evaluated at y_t. Variable t represents the time index. Estimation of the $w_{\text{BMA\_}k}$ and σ are performed in the M step as follows

$$w_{\text{BMA\_}k}^{(j)} = \frac{1}{n}\sum_t \hat{z}_{kt}^{(j)} \tag{8.9}$$

$$\sigma^{2(j)} = \frac{1}{n}\sum_t \sum_{k=1}^{K} \hat{z}_{kt}^{(j)}\left\{y_t - (a_{\text{BMA\_}k} + b_{\text{BMA\_}k}\zeta_{kt})\right\}^2 \tag{8.10}$$

where n is the length of the learning data. Finally, the slant and location parameters are adjusted as follows

$$\mathfrak{w}^{2(j)} = \sigma^{2(j)} + \frac{1}{nK}\sum_t \sum_{k=1}^{K}\left\{(a_{\text{BMA\_}k} + b_{\text{BMA\_}kt}\zeta_k) - \xi_{kt}^{(j-1)}\right\}^2 \tag{8.11}$$

$$\xi_{kt}^{(j)} = (a_{\text{BMA}\_k} + b_{\text{BMA}\_k}\zeta_{kt}) - \mathfrak{w}^{(j)}\mho\sqrt{2/\pi} \tag{8.12}$$

The EM algorithm is iterated to convergence, namely when there is little change in the log likelihood (Raftery et al., 2005) as follows

$$\ell\left(w_{\text{BMA}\_1}, \cdots, w_{\text{BMA}\_K}, \sigma^2\right) = \sum_t \log\left(\sum_{k=1}^{K} w_{\text{BMA}\_k}g_k\left(y_t|\zeta_{kt}, \xi_{kt}, \mathfrak{w}^2, \alpha\right)\right) \tag{8.13}$$

A moving training window is implemented to fit all the parameters: a_{BMA_1}, $\cdots$, a_{BMA_K}, b_{BMA_1}, $\cdots$, b_{BMA_K}, α, $\mathfrak{w}$, and $\xi_1, \cdots, \xi_K$. In the case that perturbed forecast members (mentioned in section 8.2.1) are fed into the ensemble model, exchangeability (Fraley et al., 2010) is applied, that is to say members from the same global model are forced to have the same weights (w_{BMA_k}). In addition, the fitting of variables a_{BMA_k} and b_{BMA_k} are executed only on the means of each global model to preserve the ensemble spread (Glahn et al., 2009).

To perform the BMA.n method, the code from Fraley et al. (2015) is adopted. Detailed description on this algorithm is documented by Raftery et al. (2005); Fraley et al. (2010).

8.2.4 EMOS

The EMOS.s algorithm is the same as that of EMOS.n (Gneiting et al., 2005), except that skew-normal PDF is implemented here. The details are as follows. Firstly, the biases of the forecasts are removed using linear regression as in the first step of the BMA method. Besides, in the case that perturbed forecasts members are fed in as inputs to the model, the fitting is done only to the means of each global model. In other words, all members from the same global model have the same a_{EMOS_k} and b_{EMOS_k} parameters (in Eq. (8.14)). Then, another regression is done in order to optimize

$$\lambda = \gamma + \theta_1(a_{\text{EMOS}\_1} + b_{\text{EMOS}\_1}\zeta_1) + \cdots + \theta_K(a_{\text{EMOS}\_K} + b_{\text{EMOS}\_K}\zeta_K) + \epsilon \tag{8.14}$$

$$\text{Var}\left(\epsilon\right) = c + dS^2, c \geq 0, d \geq 0 \tag{8.15}$$

$$y|\zeta_1, \cdots, \zeta_K \sim SN\left(\xi, \mathfrak{w}^2, \alpha\right) \tag{8.16}$$

$$\xi = -\mathfrak{w}\mho\sqrt{2/\pi} + \lambda \tag{8.17}$$

where λ is a univariate weather observation, γ, θ_1, $\cdots$, θ_K, c, and d are fitted parameters, ϵ represents the error with zero mean, and S^2 is the variance of the fed-in ensemble members. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm in the function `optim` of the R software (R Core Team, 2015) is implemented to find the optimum fitted parameters which minimize the fitting continuous ranked probability score (CRPS) of the learning data (Gneiting et al., 2005). The choice of the starting values fed into the optimization algorithm are made based on past experiences (Gneiting et al., 2005). In each iteration, ξ is adjusted to ensure that λ is always the mean of the predictive PDF (Eq. (8.17)). In order to guarantee that $c, d \geq 0$, $\mathfrak{k}^2 = c$ and $\Delta^2 = d$ are set, then $\mathfrak{k}$ and Δ are optimized. In addition, all perturbed forecast members coming from the same global model are guaranteed to have the same θ_k, in line with the principle of exchangeability mentioned in section 8.2.3. Besides, α is selected based on the fitting of skew-normal PDF on the estimated kernel density of the learning data. The CRPS, used in the optimization algorithm and the verification process (in section 8.3.3), is calculated as in Eq. (1.5). Like in section 8.2.3, a moving training window is implemented to adjust all the parameters: γ, θ_1, $\cdots$, θ_K, $a_{\text{EMOS\_1}}$, $\cdots$, $a_{\text{EMOS\_K}}$, $b_{\text{EMOS\_1}}$, $\cdots$, $b_{\text{EMOS\_K}}$, c, and d.

The code from Yuen et al. (2013) is implemented to perform the EMOS.n method with the additional linear regression step in the very beginning where the biases of the learning data are removed as in the EMOS.s algorithm. It is done especially since the KMA global models give forecast outputs differently compared to those of ECMWF and JMA (Thorey et al., 2015). The detail of the EMOS.n algorithm is reported in Gneiting et al. (2005).

8.2.5   AnEn

The AnEn method is a probabilistic forecasting method which originally intended to take in outputs from single NWP model (Delle Monache et al., 2013). This method derives the predictive PDF from past forecasting attempts whose NWP outputs are similar to those of the current one. It assumes that the NWP model is a frozen model. Otherwise, it may be difficult to find good analogs from past forecasts (Delle Monache et al., 2013). The degree of similarities can be computed as follows

$$||F_t, A_{t'}|| = \sum_{i=1}^{N_v} \frac{w_{\mathrm{AnEn}\_i}}{\sigma_{\mathrm{AnEn}\_i}} \sqrt{\sum_{j=-\tilde{t}}^{\tilde{t}} (F_{i,t+j} - A_{i,t'+j})^2} \qquad (8.18)$$

where $F_{i,t}$ represents the current forecasting output of physical variable i for future time t, $A_{i,t'}$ symbolizes that for time t' which has already passed, $\tilde{t}$ is the time window half-width in which the function is computed, σ_{AnEn_i} signifies the standard deviation of the past forecast time series of physical variable i, w_{AnEn_i} is the weight of each physical variable, and N_v is the number of the physical variables considered in the model. In this chapter, instead of various meteorological parameters, accumulated irradiance from various NWPs are utilized as inputs. Thus, N_v here signifies the number of NWP models. Apart from that, only $\tilde{t} = 0$ is considered in this chapter.

Following Junk et al. (2015), brute-force is implemented in order to optimize w_{AnEn_i}. This way is chosen since it is difficult to solve the problem using optimization algorithms. To facilitate this, the weights w_{AnEn_i} are limited by

$$\sum_{i=1}^{N_v} w_{\mathrm{AnEn}\_i} = 1 \text{ and } w_{\mathrm{AnEn}\_i} \in \{0, 0.1, 0.2, \cdots, 1\} \qquad (8.19)$$

Table 8.1: List of Methods

Method	NWP Inputs	Abbreviation
Climatology	-	CLIM
Autoregressive integrated moving average	-	ARIMA
Exponential smoothing	-	ETS
Bias-corrected ensemble	All perturbed forecasts	BCE.ALL
Bias-corrected ensemble	ECMWF perturbed forecasts	BCE.E
Bias-corrected ensemble	JMA perturbed forecasts	BCE.J
Bias-corrected ensemble	KMA perturbed forecasts	BCE.K
Normal BMA	All control forecasts	BMA.n.c
Normal BMA	All perturbed forecasts	BMA.n.p
Normal BMA	ECMWF perturbed forecasts	BMA.n.E.p
Skew-normal BMA	All control forecasts	BMA.s.c
Skew-normal BMA	All perturbed forecasts	BMA.s.p
Skew-normal BMA	ECMWF perturbed forecasts	BMA.s.E.p
Normal EMOS	All control forecasts	EMOS.n.c
Normal EMOS	All perturbed forecasts	EMOS.n.p
Normal EMOS	ECMWF perturbed forecasts	EMOS.n.E.p
Skew-normal EMOS	All control forecasts	EMOS.s.c
Skew-normal EMOS	ECMWF control forecasts	EMOS.s.E.c
Skew-normal EMOS	All perturbed forecasts	EMOS.s.p
Skew-normal EMOS	ECMWF perturbed forecasts	EMOS.s.E.p
AnEn	All control forecasts	AnEn.c
AnEn with weight optimization	All control forecasts	AnEn.c.opt
AnEn	ECMWF control forecasts	AnEn.E.c

8.3 Exploratory analysis, results, and discussion

8.3.1 List of methods

Several forecasting models are tested in this work and they are listed in Table 8.1. Besides various combinations of BMA, EMOS, and AnEn, as mentioned in section 8.1, some more basic models are tested for benchmarking purpose, namely CLIM (Gneiting et al., 2005), ARIMA (Hyndman, 2016; Hyndman and Khandakar, 2008), ETS (Hyndman, 2016; Hyndman and Khandakar, 2008), and BCE. The CLIM method is done by fitting an estimated kernel density on the historical observations and applying it as the predictive PDF. The BCE algorithm applies linear regression on the means of each global model in order to preserve the ensemble spread, as implemented in the BMA (section 8.2.3) and EMOS (section 8.2.4) algorithms. A moving training window is implemented in all of the forecasting methods.

For EMOS.s.E.c, the value of S in Eq. (8.15) is set as 0. Method AnEn.E.c has N_v (Eq. (8.18)) value of 1 since both methods only consider one type of meteorological parameter (accumulated irradiance) from single NWP. Methods which use solely JMA and KMA members, apart from BCE.J and BCE.K, are not included since typically ECMWF members give better results compared to those of the other two global models, as will be reflected in the BCE performances in Table 8.2, described in section 8.3.3.

8.3.2   Length of training period and AnEn ensemble size

As in Raftery et al. (2005); Gneiting et al. (2005), the optimum length of training period (τ_{opt}) needs to be chosen for each forecasting method other than the AnEn methods so that the probabilistic forecasts are well-calibrated. For the AnEn methods, the optimum ensemble size ($\mathrm{ES}_{\mathrm{opt}}$) needs to be chosen for each method instead. A way to evaluate the probabilistic forecasts calibratedness is by utilizing the probability integral transform (PIT) histogram whose squared bias (SB) can be calculated based on Eq. (1.6). The description of PIT has been written in section 1.6. In this chapter, the number of bins in PIT histograms is set as 20 and it is deemed to be sufficient (Gneiting et al., 2007a).

The selection of τ_{opt} for non-AnEn methods is done as follows. Proper forecasts are done to days ranging from 17 June 2014 to 23 January 2015 using moving training windows with the lengths of 20, 25, $\cdots$, 110 days. This span of dates will be subsequently called period I. Here, the number of forecast days is always 221 days regardless of the length of the moving training window. For each forecasting method, the length of training days with the minimum SB in period I is chosen as the corresponding τ_{opt}. This criterion may be improved in the future in order to achieve better results.

The selection of $\mathrm{ES}_{\mathrm{opt}}$ for the AnEn methods is done as follows. Proper

forecasts are done to days in period I using ensemble sizes of 10, 15, $\cdots$, 110 days. Exception is applied for the AnEn.c.opt method since the first 60 days of period I (17 June 2014 to 15 August 2014) are used to optimize w_{AnEn_i} in Eq. (8.18). The learning data, in which the ensemble can be constructed, are accumulative, namely they span from the very first date (27 February 2014) to the last available day before the forecasting day. For each forecasting method, the ensemble size with the minimum SB in period I (or in the last 161 days of period I for the AnEn.c.opt method) is chosen as the corresponding $\mathrm{ES}_{\mathrm{opt}}$. This criterion also subjects to limitations as that of τ_{opt}.

Table 8.2: Results of the Half- and Full-Day Accumulated Solar Irradiance Probabilistic Forecast of Period II

Forecast method	Half-Day Accumulated Solar Irradiance Forecasting							Full-Day Accumulated Solar Irradiance Forecasting						
	τ_{opt} (days)	ES_{opt} (days)	rRMSE (%)	CRPS (Wh/m²)	SB	r	p	τ_{opt} (days)	ES_{opt} (days)	rRMSE (%)	CRPS (Wh/m²)	SB	r	p
CLIM	90	NA	26.3	394	0.07	0.045	3.9e-01	35	NA	25.1	580	0.07	0.023	6.6e-01
ARIMA	100	NA	26.6	406	0.16	0.033	5.3e-01	95	NA	25.0	575	0.05	-0.125	1.7e-02
ETS	90	NA	26.4	406	0.18	0.012	8.2e-01	90	NA	25.4	585	0.06	-0.104	4.6e-02
BCE.ALL	25	NA	19.4	307	0.65	0.178	**6.5e-04**	30	NA	18.7	448	0.79	0.139	**7.9e-03**
BCE.E	40	NA	19.6	315	1.20	0.247	**1.7e-06**	50	NA	19.1	473	1.29	0.189	**2.9e-04**
BCE.J	25	NA	22.2	413	4.37	0.250	**1.3e-06**	30	NA	21.0	577	3.58	0.053	3.1e-01
BCE.K	60	NA	20.9	352	1.72	0.123	1.8e-02	30	NA	20.2	518	1.90	0.158	**2.4e-03**
BMA.n.c	40	NA	19.9	301	0.13	0.048	3.6e-01	70	NA	19.1	436	0.06	0.073	1.6e-01
BMA.n.p	70	NA	20.4	309	0.11	0.166	**1.5e-03**	60	NA	19.3	439	0.08	0.034	5.2e-01
BMA.n.E.p	65	NA	20.3	308	0.15	0.085	1.1e-01	35	NA	19.3	439	0.06	0.089	9.1e-02
BMA.s.c	55	NA	19.8	299	0.11	0.112	3.2e-02	75	NA	19.1	434	0.06	0.078	1.3e-01
BMA.s.p	80	NA	20.0	301	0.08	0.257	**6.6e-07**	35	NA	18.7	425	0.07	0.203	**9.1e-05**
BMA.s.E.p	80	NA	20.1	305	0.11	0.173	**9.3e-04**	30	NA	19.2	435	0.05	0.127	1.5e-02
EMOS.n.c	70	NA	20.2	310	0.15	0.087	9.7e-02	90	NA	19.3	446	0.13	0.018	7.3e-01
EMOS.n.p	20	NA	20.6	322	0.12	0.042	4.3e-01	25	NA	19.6	455	0.09	0.158	**2.4e-03**
EMOS.n.E.p	110	NA	20.4	307	0.25	0.274	**1.0e-07**	70	NA	19.5	447	0.08	0.127	1.5e-02
EMOS.s.c	60	NA	19.9	307	0.14	0.044	4.1e-01	90	NA	19.2	440	0.13	0.056	2.9e-01
EMOS.s.E.c	25	NA	20.0	303	0.12	0.001	9.9e-01	50	NA	19.5	448	0.09	0.040	4.4e-01
EMOS.s.p	105	NA	20.3	307	0.15	0.299	**5.7e-09**	30	NA	19.8	458	0.12	0.149	**4.3e-03**
EMOS.s.E.p	35	NA	19.5	293	0.10	0.178	**6.6e-04**	60	NA	19.4	443	0.09	0.143	**6.1e-03**
AnEn.c	acc.	90	20.7	312	0.17	0.295	**9.1e-09**	acc.	65	19.7	448	0.15	0.168	**1.3e-03**
AnEn.c.opt	acc.	65	20.3	306	0.13	0.255	**8.2e-07**	acc.	15	19.5	447	0.10	0.207	**6.9e-05**
AnEn.E.c	acc.	40	20.4	309	0.06	0.184	**4.3e-04**	acc.	45	19.7	452	0.09	0.161	**2.0e-03**

Abbreviation acc. stands for accumulative, as mentioned in section 8.3.2.
p values printed in bold are those which are < 0.010.

Table 8.3: Paired *t* Test of CRPS of Half-Day Accumulated Solar Irradiance Forecasting Results

Method I		BCE.ALL	BMA.n.c	BMA.n.p	BMA.n.E.p	BMA.s.c	BMA.s.p	BMA.s.E.p	EMOS.n.c	EMOS.n.p	EMOS.n.E.p	EMOS.s.c	EMOS.s.E.c	EMOS.s.p	EMOS.s.E.p	AnEn.c	AnEn.c.opt	AnEn.E.c
BMA.s.c	*p*	2e-01	7e-01	2e-02	7e-02	NA	7e-01	3e-01	**7e-03**	**6e-03**	3e-01	**7e-03**	6e-01	2e-01	3e-01	8e-02	4e-01	2e-01
	L99	-2e+01	-1e+01	-2e+01	-2e+01	NA	-1e+01	-2e+01	-2e+01	-4e+01	-2e+01	-2e+01	-2e+01	-2e+01	-9e+00	-3e+01	-3e+01	-3e+01
	H99	8e+00	9e+00	1e+00	4e+00	NA	9e+00	9e+00	-5e-01	-1e+00	1e+01	-3e-01	1e+01	6e+00	2e+01	6e+00	1e+01	1e+01
BMA.s.p	*p*	3e-01	9e-01	**9e-03**	7e-02	7e-01	NA	3e-01	5e-02	1e-02	2e-01	2e-01	7e-01	4e-02	1e-01	1e-01	5e-01	3e-01
	L99	-2e+01	-2e+01	-2e+01	-2e+01	-9e+00	NA	-3e+01	-2e+01	-4e+01	-2e+01	-2e+01	-2e+01	-1e+01	-6e+01	-3e+01	-2e+01	-3e+01
	H99	1e+01	1e+01	-9e-02	3e+00	1e+01	NA	5e+00	3e+00	9e-01	6e+00	5e+00	2e+01	2e+01	2e+01	7e+00	1e+01	1e+01
EMOS.s.E.p	*p*	2e-02	2e-01	**3e-03**	**1e-03**	3e-01	1e-01	**5e-03**	**1e-02**	**7e-04**	**9e-03**	2e-02	3e-02	3e-02	NA	3e-02	1e-01	4e-02
	L99	-3e+01	-3e+01	-3e+01	-3e+01	-2e+01	-2e+01	-2e+01	-3e+01	-5e+01	-3e+01	-3e+01	-2e+01	-3e+01	NA	-4e+01	-3e+01	-4e+01
	H99	2e+00	9e+00	-2e+00	-3e+00	9e+00	6e+00	-1e+00	-1e-01	-7e+00	-2e-01	2e+00	2e+00	2e+00	NA	4e+00	8e+00	4e+00

L99 and H99 are in Wh/m².
p values printed in bold are those whose H99 values are < 0.

Table 8.4: Paired *t* Test of CRPS of Full-Day Accumulated Solar Irradiance Forecasting Results

Method I		BCE.ALL	BMA.n.c	BMA.n.p	BMA.n.E.p	BMA.s.c	BMA.s.p	BMA.s.E.p	EMOS.n.c	EMOS.n.p	EMOS.n.E.p	EMOS.s.c	EMOS.s.E.c	EMOS.s.p	EMOS.s.E.p	AnEn.c	AnEn.c.opt	AnEn.E.c
BMA.s.c	*p*	1e-01	4e-01	5e-01	6e-01	NA	2e-01	9e-01	3e-02	7e-02	1e-01	2e-01	1e-01	3e-02	3e-01	9e-02	1e-01	7e-02
	L99	-4e+01	-9e+00	-2e+01	-3e+01	NA	-9e+00	-3e+01	-3e+01	-5e+01	-3e+01	-2e+01	-4e+01	-5e+01	-3e+01	-4e+01	-3e+01	-4e+01
	H99	8e+00	5e+00	1e+01	2e+01	NA	3e+01	3e+01	2e+00	9e+00	2e+00	6e+00	1e+01	4e+00	1e+01	8e+00	9e+00	7e+00
BMA.s.p	*p*	**3e-04**	1e-01	2e-02	4e-02	2e-01	NA	2e-01	3e-02	**8e-04**	**8e-03**	1e-01	1e-02	**2e-05**	2e-02	3e-02	5e-02	2e-02
	L99	-4e+01	-3e+01	-3e+01	-3e+01	-9e+00	NA	-3e+01	-5e+01	-5e+01	-4e+01	-4e+01	-5e+01	-5e+01	-4e+01	-5e+01	-5e+01	-6e+01
	H99	-7e+00	7e+00	1e+00	3e+00	3e+00	NA	9e+00	4e+00	-7e+00	-7e-01	9e+00	1e+00	-1e+01	1e+00	5e+00	7e+00	2e+00
BMA.s.E.p	*p*	1e-01	9e-01	7e-01	6e-01	9e-01	2e-01	NA	4e-01	6e-02	9e-02	7e-01	8e-02	2e-02	2e-01	3e-01	1e-01	1e-01
	L99	-3e+01	-3e+01	-3e+01	-2e+01	-9e+00	-3e+01	NA	-4e+01	-5e+01	-3e+01	-4e+01	-3e+01	-5e+01	-2e+01	-5e+01	-4e+01	-5e+01
	H99	8e+00	7e+00	3e+01	1e+01	9e+00	3e+01	NA	2e+01	6e+00	6e+00	3e+01	6e+00	2e+00	7e+00	2e+01	2e+01	1e+01

L99 and H99 are in Wh/m².
p values printed in bold are those whose H99 values are < 0.

8.3.3   Accumulated solar irradiance forecasting results

The forecasting results of period II with τ_{opt} for each non-AnEn method and with $\mathrm{ES}_{\mathrm{opt}}$ for each AnEn method are shown in Table 8.2. Period II spans from 24 January 2015 to 23 January 2016, and contains 365 days. Here, the deterministic forecasting values, used to compute the relative root mean square error (rRMSE), r, and p (r and p will be explained later), are taken from the expected values of the corresponding predictive PDFs. The lowest CRPS for half- and full-day accumulated irradiance forecasting are reached by the EMOS.s.E.p and BMA.s.p methods respectively. However, the BMA.s.p method also obtains the third lowest CRPS for the first forecast category.

Further investigation, nevertheless, shows that none of the methods is superior among the others in terms of CRPS. Tables 8.3 and 8.4 show the results of paired t tests for CRPS (Glahn et al., 2009) of half- and full-day accumulated irradiance forecasting results respectively. In both tables, method I column shows the best three methods for each forecasting category in terms of CRPS. The CRPS values of these methods are compared against those of the other methods. The CLIM method is excluded since it has very high CRPS. Furthermore, BCE.E, BCE.J, and BCE.K are excluded as well since their CRPS values are consistently worse than

those of BCE.ALL method. The tests are conducted using the function `t.test` (R Core Team, 2015). p value shows the statistical significance of the null hypothesis, namely that the corresponding pair of method I and method II have the same CRPS. Thus, smaller p value indicates more assurance that the corresponding method I has smaller CRPS compared to that of the corresponding method II. L99 and H99 show the low and high 99% confidence interval respectively of the difference between CRPS of the corresponding method I and that of the corresponding method II. Therefore, in this case, L99 and H99 are negatively-oriented.

Quantities r and p are computed following Veenhuis (2013) and describe the spread-skill relationships. Quantity r is the Pearson product-moment correlation coefficient between the series of square root of the absolute errors of deterministic forecast and the series of square root of the predictive PDFs standard deviations. The square root values are taken in order to diminish heteroscedasticity (Veenhuis, 2013). Higher r value indicates that the corresponding method is able to estimate its own error. Quantity p represents how probable the corresponding r value is obtained by chance given that the null hypothesis is true. p values in Table 8.2 which are smaller than 0.010 is printed in bold.

The r and p quantities in Table 8.2 shows that there is almost no spread-skill relationship in the ensemble forecasts. The highest r values are 0.299 and 0.207 for half- and full-day accumulated irradiance forecasting respectively. Higher r values are typically obtained for non-AnEn methods which utilize perturbed forecast members. The relatively small r values indicate that, in this case, spread-skill relationship does not help to construct sharper forecasts.

Fig. 8.2 depicts the PIT histograms of the best-three methods as in Tables 8.3 and 8.4, for half- and full-day accumulated irradiance forecasting. In addition, those of the CLIM and BCE.ALL methods are included as well. In general, the forecasting methods turn out to be better calibrated in the second forecasting category (Fig. 8.2 (bottom)) than in the first one (Fig. 8.2 (top)). Furthermore, consistent

with Raftery et al. (2005), the CLIM method appears to be well calibrated. However, it is not competitive enough in terms of CRPS due to its wide predictive PDFs. The predictive PDFs of BCE (represented by BCE.ALL in Fig. 8.2) are found to be underdispersive as many of the PITs fall in the first and last bins.
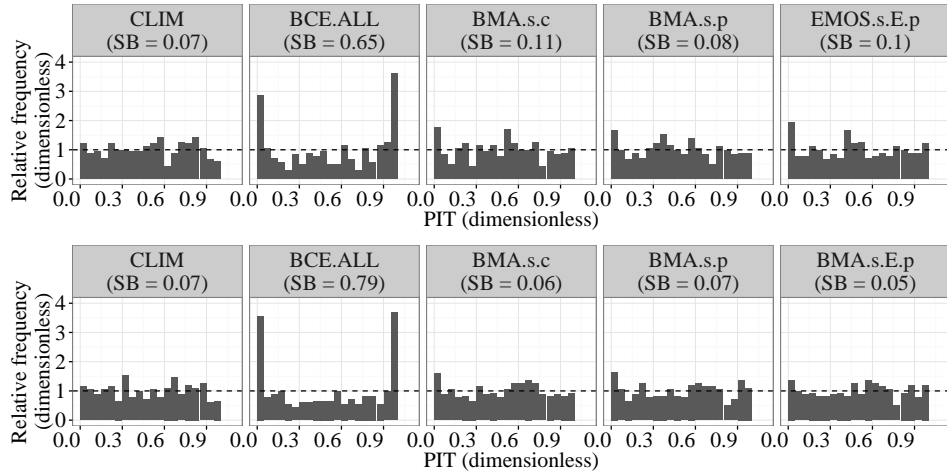
Fig. 8.2: PIT histograms of various methods for half- (top) and full-day (bottom) accumulated solar irradiance forecasting. The SB values are the same as those in Table 8.2 and included here for convenience of the readers.

Due to the limitations in the selection method of τ_{opt} and $\mathrm{ES}_{\mathrm{opt}}$, mentioned in section 8.3.2, Tables 8.5 - 8.7 are included for verification purpose. In Table 8.5, the τ and ES values which give the smallest CRPS in period II are chosen. Thus, these are just pseudo-forecasts. The results in Tables 8.2 - 8.7 and Fig. 8.2 show that, overall, BMA.s.p seems to be the most decent method although it is not significantly better than the rests.

Fig. 8.3 shows the BMA.s.p weights of the three global models over period II for half- and full-day accumulated irradiance forecasting. Those weights are the total weights of the perturbed members of each global model where members from the same global model are always assigned the same w_{BMA_k} (section 8.2.3; Eq. (8.5)). The figure shows that none of the models constantly has low weights, indicating the importance of each model in these forecasting cases. The ECMWF

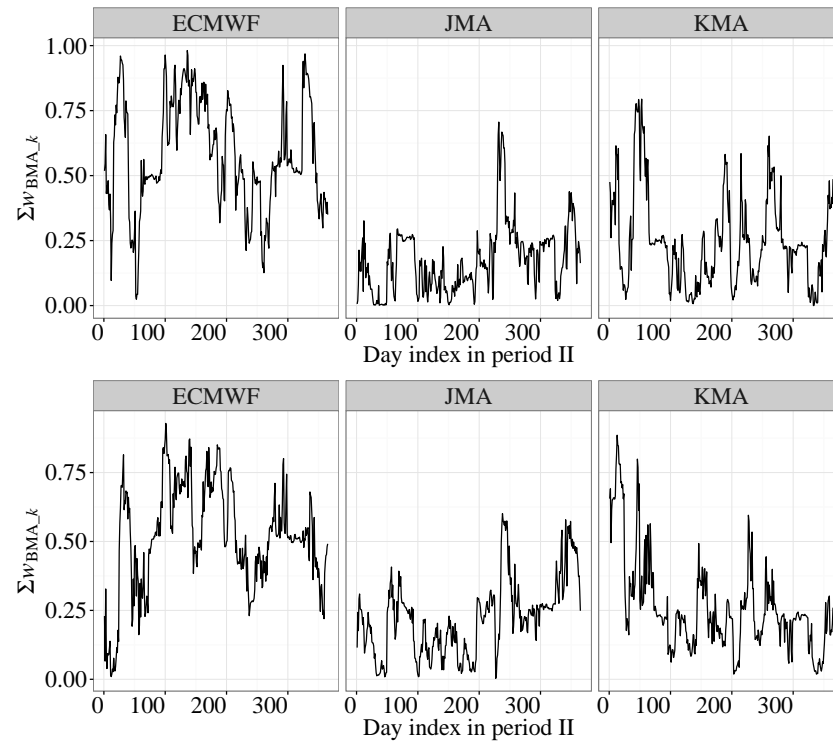Fig. 8.3: BMA.s.p weights for the three global models over period II for half- (top) and full-day (bottom) accumulated irradiance forecasting.

model seems dominating the ensemble forecasts, in line with the BCE.E accuracies compared to those of the BCE.J and BCE.K methods stated in Table 8.2.

Table 8.5: Results of the Half- and Full-Day Accumulated Solar Irradiance Probabilistic Pseudo-Forecast of Period II

Forecast method	Half-Day Accumulated Solar Irradiance Forecasting							Full-Day Accumulated Solar Irradiance Forecasting						
	τ_{opt} (days)	ES_{opt} (days)	rRMSE (%)	CRPS (Wh/m^2)	SB	r	p	τ_{opt} (days)	ES_{opt} (days)	rRMSE (%)	CRPS (Wh/m^2)	SB	r	p
CLIM	50	NA	25.7	386	0.07	0.006	9.1e-01	50	NA	25.0	574	0.08	-0.020	7.1e-01
ARIMA	110	NA	26.0	399	0.16	-0.035	5.0e-01	110	NA	24.2	556	0.07	-0.075	1.5e-01
ETS	55	NA	25.9	395	0.24	0.011	8.4e-01	110	NA	25.0	576	0.07	-0.039	4.5e-01
BCE.ALL	35	NA	19.3	305	0.72	0.196	**1.7e-04**	35	NA	18.5	444	0.76	0.140	**7.3e-03**
BCE.E	35	NA	19.5	312	1.19	0.241	**3.3e-06**	35	NA	18.9	463	1.19	0.216	**3.1e-05**
BCE.J	40	NA	21.9	403	4.28	0.205	**7.8e-05**	35	NA	20.6	567	3.57	0.047	3.7e-01
BCE.K	50	NA	20.8	349	1.59	0.134	1.1e-02	55	NA	19.9	512	1.93	0.155	**3.0e-03**
BMA.n.c	35	NA	19.9	301	0.13	0.066	2.1e-01	55	NA	18.8	432	0.08	0.056	2.8e-01
BMA.n.p	40	NA	20.2	305	0.17	0.143	**6.1e-03**	55	NA	19.3	439	0.07	0.066	2.1e-01
BMA.n.E.p	35	NA	20.0	302	0.17	0.081	1.2e-01	35	NA	19.3	439	0.06	0.089	9.1e-02
BMA.s.c	35	NA	19.7	297	0.09	0.126	1.6e-02	55	NA	18.7	429	0.07	0.059	2.6e-01
BMA.s.p	40	NA	19.5	292	0.08	0.213	**4.3e-05**	35	NA	18.7	425	0.07	0.203	**9.1e-05**
BMA.s.E.p	35	NA	19.5	294	0.11	0.112	3.3e-02	35	NA	18.9	431	0.06	0.138	**8.2e-03**
EMOS.n.c	40	NA	20.1	306	0.18	0.130	1.3e-02	55	NA	19.0	443	0.17	0.068	2.0e-01
EMOS.n.p	35	NA	19.9	303	0.12	0.120	2.1e-02	40	NA	19.1	440	0.06	0.118	2.4e-02
EMOS.n.E.p	35	NA	19.7	297	0.20	0.151	**3.8e-03**	35	NA	19.2	436	0.10	0.203	**9.4e-05**
EMOS.s.c	50	NA	19.8	304	0.10	0.093	7.6e-02	55	NA	19.0	437	0.13	0.093	7.7e-02
EMOS.s.E.c	35	NA	19.8	301	0.07	-0.032	5.4e-01	35	NA	19.4	445	0.12	0.008	8.8e-01
EMOS.s.p	40	NA	19.9	303	0.14	0.189	**2.8e-04**	50	NA	19.1	438	0.12	0.192	**2.2e-04**
EMOS.s.E.p	35	NA	19.5	293	0.10	0.178	**6.6e-04**	35	NA	19.1	434	0.08	0.185	**3.7e-04**
AnEn.c	acc.	50	20.3	307	0.12	0.275	**9.5e-08**	acc.	15	19.4	445	0.13	0.154	**3.2e-03**
AnEn.c.opt	acc.	45	20.1	304	0.14	0.286	**2.6e-08**	acc.	40	19.4	442	0.15	0.183	**4.4e-04**
AnEn.E.c	acc.	30	20.3	309	0.06	0.169	**1.2e-03**	acc.	45	19.7	452	0.09	0.161	**2.0e-03**

Abbreviation acc. stands for accumulative, as mentioned in section 8.3.2.
p values printed in bold are those which are < 0.010.
For the AnEn.c.opt method, the optimized weights are found to be: $w_{AnEn.ECMWF} = 0.4$, $w_{AnEn.JMA} = 0.2$, and $w_{AnEn.KMA} = 0.4$ for half-day accumulated irradiance forecast, and $w_{AnEn.ECMWF} = 0.5$, $w_{AnEn.JMA} = 0.2$, and $w_{AnEn.KMA} = 0.3$ for full-day accumulated irradiance forecast.

Table 8.6: Paired t Test of CRPS of Half-Day Accumulated Solar Irradiance Pseudo-Forecasting Results

Method I		BCE.ALL	BMA.n.c	BMA.n.p	BMA.n.E.p	BMA.s.c	BMA.s.p	BMA.s.E.p	EMOS.n.c	EMOS.n.p	EMOS.n.E.p	EMOS.s.c	EMOS.s.E.c	EMOS.s.p	EMOS.s.E.p	AnEn.c	AnEn.c.opt	AnEn.E.c
BMA.s.p	p	**3e-03**	9e-02	**9e-05**	3e-02	3e-01	NA	6e-01	**9e-03**	**3e-03**	3e-01	2e-02	1e-01	**2e-03**	9e-01	7e-02	1e-01	6e-02
	L99	-2e+01	-2e+01	-2e+01	-2e+01	-2e+01	NA	-1e+01	-3e+01	-2e+01	-2e+01	-3e+01	-2e+01	-2e+01	-1e+01	-4e+01	-3e+01	-4e+01
	H99	-2e+00	5e+00	-5e+00	2e+00	6e+00	NA	9e+00	-3e-01	-1e+00	7e+00	2e+00	5e+00	-2e+00	1e+01	7e+00	9e+00	6e+00
BMA.s.E.p	p	6e-02	3e-01	4e-02	4e-02	6e-01	6e-01	NA	1e-01	9e-02	3e-01	6e-02	6e-02	1e-01	3e-01	2e-01	3e-01	8e-02
	L99	-2e+01	-2e+01	-2e+01	-2e+01	-2e+01	-9e+00	NA	-3e+01	-2e+01	-7e+01	-3e+01	-2e+01	-2e+01	-3e+01	-4e+01	-3e+01	-3e+01
	H99	4e+00	1e+01	3e+00	2e+00	1e+01	1e+01	NA	6e+00	5e+00	3e+00	7e+00	3e+00	5e+00	6e+00	1e+01	1e+01	7e+00
EMOS.s.E.p	p	3e-02	2e-01	2e-02	2e-02	5e-01	9e-01	3e-01	6e-02	5e-02	**4e-04**	9e-02	3e-02	6e-02	NA	1e-01	2e-01	5e-02
	L99	-3e+01	-3e+01	-3e+01	-2e+01	-2e+01	-1e+01	-6e+00	-3e+01	-2e+01	-7e+00	-3e+01	-2e+01	-2e+01	NA	-4e+01	-3e+01	-4e+01
	H99	3e+00	1e+01	2e+00	1e+00	1e+01	1e+01	3e+00	5e+00	3e+00	-1e+00	6e+00	1e+00	4e+00	NA	9e+00	1e+01	5e+00

L99 and H99 are in Wh/m^2.
p values printed in bold are those whose H99 values are < 0.

Table 8.7: Paired t Test of CRPS of Full-Day Accumulated Solar Irradiance Pseudo-Forecasting Results

Method I		BCE.ALL	BMA.n.c	BMA.n.p	BMA.n.E.p	BMA.s.c	BMA.s.p	BMA.s.E.p	EMOS.n.c	EMOS.n.p	EMOS.n.E.p	EMOS.s.c	EMOS.s.E.c	EMOS.s.p	EMOS.s.E.p	AnEn.c	AnEn.c.opt	AnEn.E.c
BMA.s.c	p	2e-02	2e-01	5e-02	2e-01	NA	5e-01	8e-01	2e-02	1e-01	4e-01	1e-01	7e-02	1e-01	6e-01	1e-01	1e-01	2e-02
	L99	-3e+01	-8e+00	-2e+01	-3e+01	NA	-1e+01	-2e+01	-3e+01	-3e+01	-3e+01	-2e+01	-4e+01	-3e+01	-3e+01	-4e+01	-4e+01	-5e+01
	H99	2e+00	3e+00	3e+00	1e+01	NA	2e+01	2e+01	1e+00	7e+00	2e+01	5e+00	7e+00	7e+00	2e+01	9e+00	1e+01	2e+00
BMA.s.p	p	**5e-04**	3e-01	**9e-03**	4e-02	5e-01	NA	4e-01	4e-02	**6e-03**	1e-01	1e-01	2e-02	4e-02	2e-01	9e-02	1e-01	2e-02
	L99	-3e+01	-2e+01	-3e+01	-3e+01	-2e+01	NA	-2e+01	-4e+01	-3e+01	-3e+01	-3e+01	-4e+01	-3e+01	-3e+01	-5e+01	-4e+01	-6e+01
	H99	-5e+00	1e+01	-2e-01	3e+00	1e+01	NA	1e+01	4e+00	-1e+00	8e+00	9e+00	3e+00	3e+00	1e+01	1e+01	1e+01	2e+00
BMA.s.E.p	p	9e-02	9e-01	3e-01	1e-01	8e-01	4e-01	NA	2e-01	3e-01	4e-02	5e-01	**9e-03**	4e-01	2e-01	3e-01	3e-01	4e-02
	L99	-3e+01	-2e+01	-3e+01	-2e+01	-2e+01	-1e+01	NA	-4e+01	-1e+01	-1e+01	-3e+01	-3e+01	-3e+01	-9e+00	-5e+01	-4e+01	-5e+01
	H99	7e+00	2e+01	1e+01	6e+00	2e+01	2e+01	NA	1e+01	2e+00	2e+00	1e+01	1e+01	1e+01	3e+00	2e+01	2e+01	5e+00

L99 and H99 are in Wh/m^2.
p values printed in bold are those whose H99 values are < 0.

8.4 Conclusion

Probabilistic forecasting on half- and full-day accumulated irradiance in Singapore have been performed by utilizing global model NWP outputs. The results show that the BMA method which is based on skew-normal PDF and takes in perturbed forecast members is, overall, the most accurate ensemble method, but it is not very significantly better than the others considered. Moreoever, small spread-skill relationships are found among various models of ensemble forecasts, and they do not convincingly help to construct sharper probabilistic forecasts. Inclusion of perturbed forecast members results in better spread-skill relationships for methods other than analog ensemble.

It is also noted that the usage of the global model with low temporal resolution is able to produce acceptable deterministic forecasts which are better than climatological results. In the future, similar work may be executed on more-refined mesoscale NWP outputs, especially in order to increase the temporal resolutions.

Chapter 9

Summary and future work

9.1 Thesis summary

This thesis has presented several methods to forecast and model solar irradiance which may be implemented as part of operational procedures to design, implement and manage solar PV power. The work can be categorized into two categories: solar forecasting and solar resource assessment. The forecasting work can be further divided into point and spatial forecasts over differing timescales. The forecasting work presented here represents the state–of–the–art in irradiance forecasting for Singapore and comparable regions, and as such is indicative of what operational forecasting has to work with at the present time. The modeling work, on the other hand, corresponds to the quantification of the geographical smoothing effect which is useful in the planning stage of the building of solar power systems.

At present, the Solar Energy Research Institute of Singapore (SERIS) provides the Energy Market Authority (Singapore's national grid operator) with single point forecasts for 15-minute-ahead timescale computed using the the Auto-Regressive Integrated Moving Average (ARIMA) algorithm. Such forecasts are considered sufficient in 2016 when total PV generation capacity remains less than 60MW, compared with over 13 000MW of capacity registered in the National Elec-

tricity Market of Singapore. In the future, when the total photovoltaic (PV) generation in Singapore is expected to grow to become an appreciable fraction of the total, there shall be need to have high spatial resolution forecasts with various timescales.

The chapters entitled *Short-term forecasts using a network of PV systems*, *Very short-term forecasting using spatio-temporal kriging*, and *Modeling variability and geographic smoothing* represent the most directly applicable results of this thesis for grid operation today. The chapters *Ensemble techniques for probabilistic forecasts* and *Numerical weather prediction applied to Singapore irradiance forecasts* represent state-of-the-art research on NWP in Singapore, but it is very likely that this work will be superceded in the future, as new codes are developed, new datasets are assimilated into tropical NWP, and once the new Singapore National Supercomputing Centre capacity comes online, allowing much larger ensembles to be processed.

The chapter *Short-term forecasts using a network of PV systems* is significant as it offers the prospect of applying the forecasting method in places where there is no network of irradiance sensors present. Indeed such sensor networks are the exception rather than the rule, and this is likely to remain the case in developing cities. By using the output of PV systems themselves in forecasting, this chapter presents a method applicable anywhere that a large number of PV systems occur. The method employed is the least absolute shrinkage and selection operator (lasso) which enables the usage of the neighbouring stations historical data to forecast the output of a particular system. In this chapter, the lasso method appears to be the best for forecasting tasks of individual and total power output of the PV systems.

The chapter *Very short-term forecasting using spatio-temporal kriging* is one of the most important results of this thesis as it shows, for the first time, how kriging can be used to make forecasts at arbitrary locations within a region of interest. This
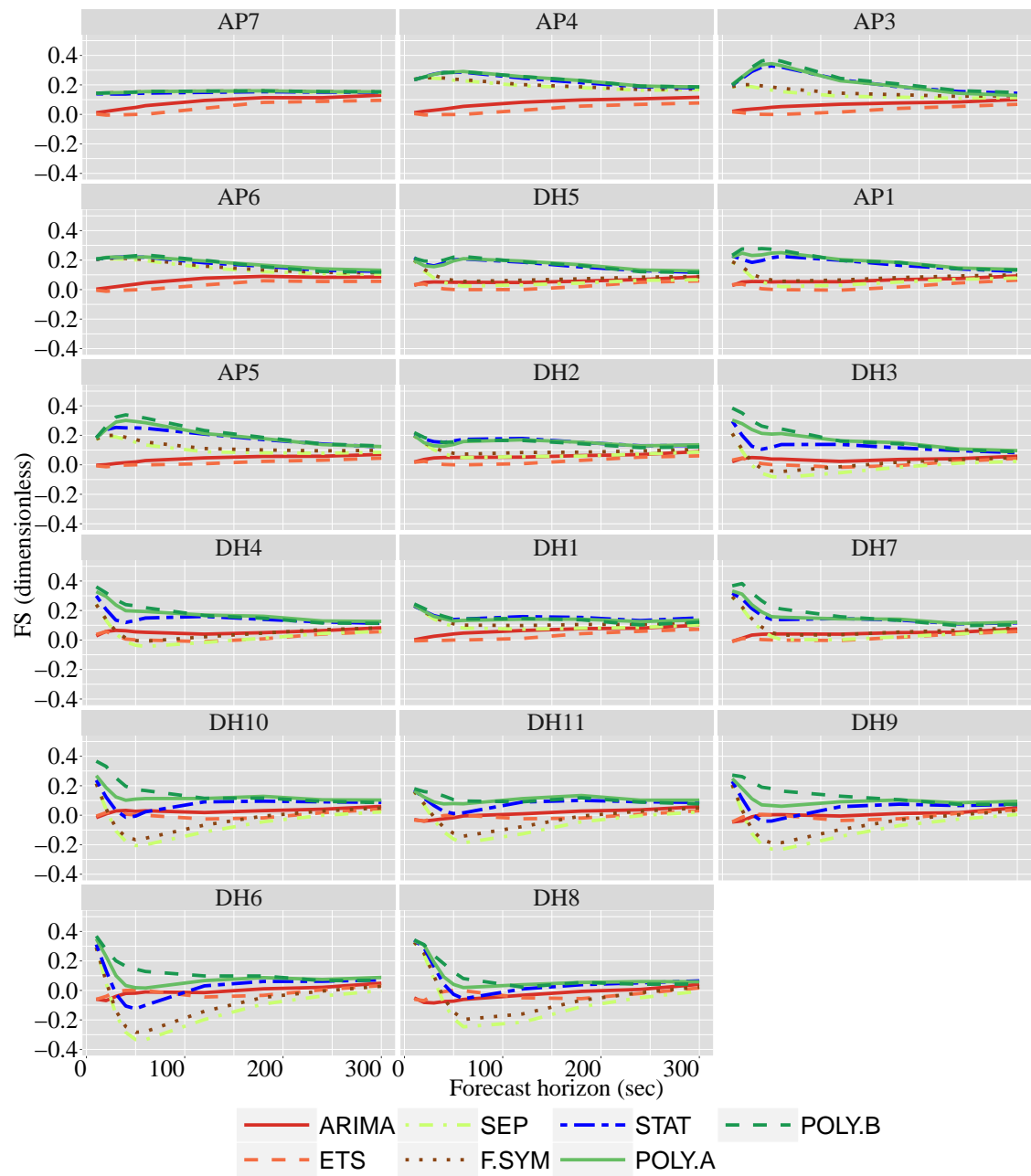
Fig. 9.1: FS for irradiance prediction at unobserved locations in Oahu, Hawaii, for various forecast horizons (averaged over the 13 days dominated by broken clouds in 2010)

ability was lacking, for example, in Yang (2014), who was the first to apply kriging to irradiance forecasts. The method presented in this chapter requires an appropriate design of the sensor network according to the dominant wind direction and speed. Fig. 9.1 shows the forecast skill (FS) of the various forecasting methods tested in this chapter. Forecasts are performed based on the measurement data of a sensor network in Oahu, Hawaii, maintained by the USA's National Renewable Energy Laboratory (NREL). This network has irradiance characteristics comparable with Singapore. The smart persistence method of the average measurement of all sensors is taken as the benchmark. The figure shows that it is possible to even reach an FS value of 0.4 when wind direction is included in the kriging model (for station AP3, using method POLY.B).

In the chapter *Modeling variability and geographic smoothing*, the kriging method is utilized to model the geographic smoothing effect of irradiance over an area, by making use of measurement data from irradiance sensors. This is particularly useful for resource assessment prior to the building of a large solar power plant. By integrating the information of wind direction for days with significant wind, the kriging method is able to estimate the geographic smoothing effect better than the existing methods in the literatures, at least for timescale of 20-120 seconds. However, deployment of more sensors is the price to get more accurate estimation.

The chapter *Numerical weather prediction applied to Singapore irradiance forecasts* presents work on the day-ahead irradiance forecast in Singapore using the Weather Research and Forecasting (WRF) model for Singapore. The result shows that combining the WRF and exponential smoothing (ETS) forecasts is able to produce day-ahead predictions with rRMSE of 49%. This level of forecast skill is not likely to be of use for operational purposes, a fact which motivates the following chapter.

In the chapter *Ensemble techniques for probabilistic forecasts*, the results of 6-

and 12-hour-ahead probabilistic accumulated irradiance forecast for Singapore are reported. The procedure to use multiple NWP runs (an "ensemble") is introduced as a method to obtain an improved forecast. The ensemble global numerical weather prediction (NWP) data are utilized. Here, the Bayesian model averaging (BMA), ensemble model output statistics (EMOS), and analog ensemble methods are shown to produce probabilistic forecasts which are more competitive than climatology and have better coverages than simple bias-corrected ensemble. However, among many variants of BMA, EMOS, and analog ensemble, there is no significantly superior method. Futhermore, the spread-skill relationships observed in the results are very weak and do not help in increasing the competitiveness of probabilistic forecasts. Nevertheless, it is noted that the bias-corrected global NWP outputs, despite the low spatial resolution, are good enough to reduce the relative root mean square error (rRMSE) of deterministic forecasts from $\sim$ 26 to 20% for 6-hour-ahead forecasts and from $\sim$ 25 to 19% for 12-hour-ahead forecasts.

9.2   Proposed future work

Future work on the geographic smoothing effect in Singapore (as discussed in the chapter *Singapore irradiance characteristics*) may be conducted by utilizing solar power plant data. Such data will give useful results since when a power plant (or a collection of plants) covers a large enough area, the output power will be smoother than the irradiance measurement of a single sensor. The work in chapter *Modeling variability and geographic smoothing* can also be repeated by using real data of one or more solar power plants *and* irradiance sensors on the same area, following Lave et al. (2013). This work can readily be applied to other cities and regions, and be used to facilitate PV power generation forecasting in those places.

The work in the chapters *Short-term forecasts using a network of PV systems*, *Very short-term forecasting using spatio-temporal kriging*, and *Ensemble techniques*

for probabilistic forecasts can be extended by way of simulations of PV and battery system (Hanna et al., 2014), and for grid integration, and/or for hybrid systems with other renewable energy sources. Financial impacts due to the implementation of solar forecasting can also be investigated (Martinez-Anido et al., 2016). In order to achieve this, the temporal resolution of the probabilistic forecasts in chapter *Ensemble techniques for probabilistic forecasts* need to be increased, to match the local electricity trading system. Sky camera and satellite images may also be explored in the future to improve sensor-based short-term forecasts.

The irradiance forecasts using the WRF model in the chapter *Numerical weather prediction applied to Singapore irradiance forecasts* can be improved in the future using data assimilation techniques. These techniques attempt to provide much more accurate boundary conditions for numerical weather prediction (NWP) models. Assimilable data can be taken from ground measurements or satellite images. It has been proposed that GPS data can be used to estimate atmospheric precipitable water vapour content. The datasets used for assimilation should, perhaps, cover areas inside and around the Singapore island so that the dynamic of the weather can be properly modeled. Finally, the ensemble work should be extended to include a range of physically meaningful initial and boundary conditions motivated by meteorological observations and theory. Larger and more frequent ensembles should be created, and their results processed as described in this thesis in order to obtain the best possible, and most rapidly-updated irradiance forecasts.

Bibliography

Achleitner, S., Kamthe, A., Liu, T., and Cerpa, A. (2014). SIPS: Solar irradiance prediction system. In *Information Processing in Sensor Networks, IPSN-14 Proceedings of the 13th International Symposium on*, pages 225–236. Available from: `http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6846755&url=` `http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%` `3D6846755`.

Ackerman, T. P. and Stokes, G. M. (2003). The atmospheric radiation measurement program. *Physics Today*, 56:38.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716 – 723. Available from: `http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1100705&url=` `http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F9%2F24140%2F01100705`.

Alessandrini, S., Delle Monache, L., Sperati, S., and Cervone, G. (2015). An analog ensemble for short-term probabilistic solar power forecast. *Applied Energy*, 157:95 – 110. Available from: `http://www.sciencedirect.com/science/article/` `pii/S0306261915009368`.

Andreas, A. and Stoffel, T. (1981). NREL solar radiation research laboratory (SRRL): Baseline measurement system (BMS); Golden, Colorado (data); NREL report no. DA-5500-56488. Technical report, National Renewable Energy Labo-

ratory. (accessed 1 March 2016). Available from: `http://dx.doi.org/10.5439/1052221`.

Andreas, A. and Wilcox, S. (2011). Solar technology acceleration center (Solar-TAC); Aurora, Colorado (data); NREL report no. DA-5500-56491. Technical report, National Renewable Energy Laboratory. (accessed 1 March 2016). Available from: `http://dx.doi.org/10.5439/1052224`.

Arias-Castro, E., Kleissl, J., and Lave, M. (2014). A Poisson model for anisotropic solar ramp rate correlations. *Solar Energy*, 101:192 – 202. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X13005549`.

Arnold, T. A. and Emerson, J. W. (2011). Nonparametric goodness-of-fit tests for discrete null distributions. *The R Journal*, 3(2):34–39. Available from: `http://journal.r-project.org/archive/2011-2/RJournal_2011-2_Arnold+Emerson.pdf`.

Aryaputera, A., Yang, D., Zhao, L., and Walsh, W. (2015a). Very short-term irradiance forecasting at unobserved locations using spatio-temporal kriging with polynomial anisotropy fitting. In *25th International Photovoltaic Science and Engineering Conference*. (oral presentation).

Aryaputera, A. W., Engerer, N., Zhao, L., and Walsh, W. M. (2016a). Output power forecast of a photovoltaic network. In *26th International Photovoltaic Science and Engineering Conference*. (poster presentation).

Aryaputera, A. W., Verbois, H., and Walsh, W. M. (2016b). Probabilistic accumulated solar irradiance forecast for Singapore using ensemble techniques. In *43rd IEEE Photovoltaic Specialists Conference*. (oral presentation).

Aryaputera, A. W., Yang, D., and Walsh, W. M. (2015b). Day-ahead solar irradi-

ance forecasting in a tropical environment. *Journal of Solar Energy Engineering*, 137(5):051009. Available from: `http://dx.doi.org/10.1115/1.4030231`.

Aryaputera, A. W., Yang, D., Zhao, L., and Walsh, W. M. (2015c). Very short-term irradiance forecasting at unobserved locations using spatio-temporal kriging. *Solar Energy*, 122:1266 – 1278. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X15005745`.

Aryaputera, A. W., Zhao, L., and Walsh, W. M. (2015d). Solar irradiance variability in Singapore. In *PV Asia Scientific Conference*. (poster presentation).

Ashraphijuo, M., Fattahi, S., Lavaei, J., and Atamtürk, A. (1999). A strong semidefinite programming relaxation of the unit commitment problem. (accessed 5 July 2015). Available from: `http://www.ieor.berkeley.edu/~lavaei/UC_2016.pdf`.

Asiaone (2016). Singapore haze. (accessed 31 March 2016). Available from: `http://weather.asiaone.com.sg/weather-air.php`.

Azzalini, A. (2014). *The Skew-Normal and Related Families*. Cambridge University Press, New York.

Azzalini, A. (2015). *The R package sn: The Skew-Normal and Skew-t distributions (version 1.3-0)*. Università di Padova, Italia. Available from: `http://azzalini.stat.unipd.it/SN`.

Bank, J., Mather, B., Keller, J., and Coddington, M. (2013). High penetration photovoltaic case study report; NREL report no. TP-5500-54742. Technical report, National Renewable Energy Laboratory. (accessed 4 July 2016). Available from: `http://www.nrel.gov/docs/fy13osti/54742.pdf`.

Beyer, H. G., Martinez, J. P., Suri, M., Martinez, J. L., Lorenz, E., Müller, S. C., Hoyer-Klick, C., and Ineichen, P. (2009). Report on benchmarking of radiation

products. Technical Report 038665 of MESOR, Management and Exploitation of Solar Resource Knowledge. Available from: `http://www.mesor.org/docs/ MESoR_Benchmarking_of_radiation_products.pdf`.

Bieri, M., Baker, R. S., Tay, S., and Reindl, T. (2016). Unsubsidised PV markets: How a national public-private financing platform could reduce the impact from low oil prices. In *32nd European Photovoltaic Solar Energy Conference and Exhibition*.

Bjerknes, V. (1904). Das problem der wettervorhersage, betrachtet vom standpunkte der mechanik und der physik. *Meteorologische Zeitschrift*, 21:1 – 7.

Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., Chen, D. H., Ebert, B., Fuentes, M., Hamill, T. M., Mylne, K., Nicolau, J., Paccagnella, T., Park, Y.-Y., Parsons, D., Raoult, B., Schuster, D., Dia, P. S., Swinbank, R., Takeuchi, Y., Tennant, W., Wilson, L., and Worley, S. (2010). The THORPEX interactive grand global ensemble. *Bull. Amer. Meteor. Soc.*, 91:10591072.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2011). *Time Series Analysis, Forecasting and Control*. Wiley.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.*, 16(5):1190–1208.

Chen, F. (2007). The Noah land surface model in WRF. `http://www.atmos. illinois.edu/~snesbitt/ATMS597R/notes/noahLSM-tutorial.pdf`. accessed: 06.05.2014.

Chotamonsak, C., Salathé, E. P., Kreasuwan, J., Chantara, S., and Siriwitayakorn, K. (2011). Projected climate change over Southeast Asia simulated using a WRF regional climate model. *Atmospheric Science Letters*, 12(2):213–219. Available from: `http://dx.doi.org/10.1002/asl.313`.

Chow, C. W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., and Washom, B. (2011). Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Solar Energy*, 85:2881–2893. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X11002982`.

Christou, N. (2014). Statistics c173/c273: Applied geostatistics (lecture notes). `http://www.stat.ucla.edu/~nchristo/statistics_c173_c273/`. (accessed: 26.08.2014).

Chu, Y., Pedro, H. T., Li, M., and Coimbra, C. F. (2015a). Real-time forecasting of solar irradiance ramps with smart image processing. *Solar Energy*, 114(0):91 – 104. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X15000389`.

Chu, Y., Urquhart, B., Gohari, S. M. I., C., H. T., Kleissl, J., and Coimbra, C. F. M. (2015b). Short-term reforecasting of power output from a 48 MWe solar PV plant. *Solar Energy*, 112:68–77. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X14005611`.

Cornaro, C., Bucci, F., Pierro, M., Del Frate, F., Peronaci, S., and Taravat, A. (2015). Twenty-four hour solar irradiance forecast based on neural networks and numerical weather prediction. *Journal of Solar Energy Engineering*, 137(3):031011.

Corripio, J. G. (2014). *insol: Solar Radiation*. R package version 1.1.1. Available from: `https://CRAN.R-project.org/package=insol`.

Cressie, N. (1990). The origins of kriging. *Mathematical Geology*, 22(3):239–252. Available from: `http://link.springer.com/article/10.1007%2FBF00889887`.

Cressie, N. (1993). *Statistics for Spatial Data.* USA: John Wiley & Sons, Inc. Available from: `http://as.wiley.com/WileyCDA/WileyTitle/productCd-0471002550.html`.

Cressie, N. and Huang, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94(448):1330–1339. Available from: `http://www.jstor.org/stable/2669946`.

Cressie, N. and Wikle, C. (2011). *Statistics for spatio-temporal data.* Wiley. Available from: `http://as.wiley.com/WileyCDA/WileyTitle/productCd-EHEP002348.html`.

Dambreville, R., Blanc, P., Chanussot, J., and Boldo, D. (2014). Very short term forecasting of the global horizontal irradiance using a spatio-temporal autoregressive model. *Renewable Energy*, 72(0):291 – 300. Available from: `http://www.sciencedirect.com/science/article/pii/S096014811400398X`.

Delle Monache, L., Eckel, F. A., Rife, D. L., Nagarajan, B., and Searight, K. (2013). Probabilistic weather prediction with an analog ensemble. *Mon. Wea. Rev.*, 141:3498–3516. Available from: `http://doi.org/10.1175/MWR-D-12-00281.1`.

Diagne, M., David, M., Boland, J., Schmutz, N., and Lauret, P. (2014). Post-processing of solar irradiance forecasts from WRF model at Reunion Island. *Solar Energy*, 105(0):99 – 108. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X14001509`.

Diagne, M., David, M., Lauret, P., Boland, J., and Schmutz, N. (2013). Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews*, 27:65 – 76. Available from: `http://www.sciencedirect.com/science/article/pii/S1364032113004334`.

Dong, Z., Yang, D., Reindl, T., and Walsh, W. M. (2013). Short-term solar irradiance forecasting using exponential smoothing state space model. *Energy*, 55(0):1104 – 1113. Available from: `http://www.sciencedirect.com/science/article/pii/S0360544213003381`.

Dong, Z., Yang, D., Reindl, T., and Walsh, W. M. (2014). Satellite image analysis and a hybrid ESSS/ANN model to forecast solar irradiance in the tropics. *Energy Conversion and Management*, 79(0):66 – 73. Available from: `http://www.sciencedirect.com/science/article/pii/S0196890413007644`.

Dudhia, J. (1989). Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences*, 46(20):3077–3107. Available from: `http://dx.doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2`.

Dudhia, J. (2014). Overview of WRF physics. Available from: `http://www2.mmm.ucar.edu/wrf/users/tutorial/201401/Physics_full.pdf`.

Duffie, J. A. and Beckman, W. A. (2013). *Solar Engineering of Thermal Processesw*. John Wiley & Sons, Inc. Available from: `http://dx.doi.org/10.1002/9781118671603`.

ECMWF (2015). Forecasts. (accessed 10 June 2015). Available from: `http://www.ecmwf.int/en/forecasts`.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32:407–499. Available from: `https://projecteuclid.org/euclid.aos/1083178935`.

Ela, E., Milligan, M., Bloom, A., Botterud, A., Townsend, A., and Levin, T. (2014). Evolution of wholesale electricity market design with increasing levels of renewable generation; NREL report no. TP-5D00-61765. Technical report,

National Renewable Energy Laboratory. (accessed 5 July 2016). Available from: `http://www.nrel.gov/docs/fy14osti/61765.pdf`.

Ela, E., Milligan, M., and Kirby, B. (2011). Operating reserves and variable generation; NREL report no. TP–5500–51978. Technical report, National Renewable Energy Laboratory. (accessed 5 July 2015). Available from: `http://www.nrel.gov/docs/fy11osti/51978.pdf`.

Energy Market Authority (2010). Introduction to the National Electricity Market of Singapore. (accessed 4 February 2017). Available from: `https://www.ema.gov.sg/cmsmedia/Handbook/NEMS_111010.pdf`.

Energy Market Authority (2016). Introducing demand response to enhance competition in the National Electricity Market of Singapore. (accessed 4 February 2017). Available from: `https://www.ema.gov.sg/cmsmedia/Electricity/Demand_Response/13jun16%20DR%20Factsheet.pdf`.

Engerer, N. and Mills, F. (2014). K_{PV}: A clear-sky index for photovoltaics. *Solar Energy*, 105(0):679 – 693. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X14002151`.

Engerer, N. A. and Hansard, J. (2015). Real-time simulations of 15,000+ distributed PV arrays at sub-grid level using the Regional PV Simulation System (RPSS). In *Solar World Congress*.

Fraley, C., Raftery, A. E., and Gneiting, T. (2010). Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging. *Mon. Wea. Rev.*, 138:190–202.

Fraley, C., Raftery, A. E., Sloughter, J. M., Gneiting, T., and University of Washington. (2015). *ensembleBMA: Probabilistic Forecasting using Ensembles*

and Bayesian Model Averaging. R package version 5.1.2. Available from: `http://CRAN.R-project.org/package=ensembleBMA`.

GE (2010). Western wind and solar integration study. Available from: `http://www.nrel.gov/docs/fy10osti/47781.pdf`.

Glahn, B., Peroutka, M., Wiedenfeld, J., Wagner, J., Zylstra, G., Schuknecht, B., and Jackson, B. (2009). MOS uncertainty estimates in an ensemble framework. *Mon. Wea. Rev.*, 137:246–268.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97(458):590–600. Available from: `http://www.jstor.org/stable/3085674`.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007a). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 69(2):243–268.

Gneiting, T., Genton, M. G., and Guttorp, P. (2007b). Geostatistical space-time models, stationarity, separability and full symmetry. In Finkenstaedt, B., Held, L., and Isham, V., editors, *Statistical Methods for Spatio-Temporal Systems*, chapter 4, pages 151–175. Chapman & Hall/CRC.

Gneiting, T., Raftery, A. E., Westveld III, A. H., and Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, 113:1098–1118.

Gould, S. (2013). Short–term machine–learning–based forecasting of distributed solar energy production. (accessed 8 June 2015). Available from: `http://www.uspvmc.org/proceedings/2ndAnnualCSiWorkshop0710/12.%20Forecasting%20Distributed%20PV%20Production.pdf`.

Gregory, K. J., Simmons, I. G., Brazel, A. J., Day, J. W., Keller, E. A., Sylvester, A. G., and Yáñez, A. (2009). *Environmental Sciences*. SAGE Publications Ltd.

Grell, E. D., Grell, G. A., Bao, J. W., and Freitas, S. R. (2013). Experiments with a convective parameterization suitable for high-resolution mesoscale models in tropical cylone simulations. 14th Annual WRF Users Workshop.

Grell, G. A. and Freitas, S. R. (2013). A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmospheric Chemistry and Physics Discussions*, 13(9):23845–23893. Available from: http://www.atmos-chem-phys-discuss.net/13/23845/2013/.

Gueymard, C. A. (2009). Direct and indirect uncertainties in the prediction of tilted irradiance for solar engineering applications. *Sol. Energy*, 83(3):432444. Available from: http://www.sciencedirect.com/science/article/pii/S0038092X08002983.

Hanna, R., Kleissl, J., Nottrott, A., and Ferry, M. (2014). Energy dispatch schedule optimization for demand charge reduction using a photovoltaic-battery storage system with solar forecasting. *Solar Energy*, 103:269 – 287. Available from: http://www.sciencedirect.com/science/article/pii/S0038092X1400098X.

Hastie, T. and Efron, B. (2013). `lars: Least Angle Regression, Lasso and Forward Stagewise`. R package version 1.2. Available from: https://CRAN.R-project.org/package=lars.

He, S., Raghavan, S. V., Nguyen, N. S., and Liong, S.-Y. (2013). Ensemble rainfall forecasting with numerical weather prediction and radar-based nowcasting models. *Hydrological Processes*, 27(11):1560 – 1571. Available from: http://dx.doi.org/10.1002/hyp.9254.

Hiemstra, P. H., Pebesma, E. J., Twenhöfel, C. J. W., and Heuvelink, G. B. M. (2008). Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Computers & Geosciences*.

Hinkelman, L. M. (2013). Differences between along-wind and cross-wind solar irradiance variability on small spatial scales. *Solar Energy*, 88:192–203. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X12004021`.

Hoff, T. E. and Perez, R. (2012). Modeling PV fleet output variability. *Solar Energy*, 86(8):2177 – 2189. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X11004154`.

Hong, S.-Y. and Lim, J.-O. J. (2006). The WRF single-moment 6-class microphysics scheme (WSM6). *Journal of the Korean Meteorological Society*, 42(2):129 – 151.

Hong, S.-Y., Noh, Y., and Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, 134(9):2318 – 2341. Available from: `http://dx.doi.org/10.1175/MWR3199.1`.

Hong, S.-Y. and Pan, H.-L. (1996). Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Monthly Weather Review*, 124:3233–3229. Available from: `http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1996)124%3C2322%3ANBLVDI%3E2.0.CO%3B2`.

Hummon, M. (2014). Solar power data for integration studies. Available from: `http://www.nrel.gov/electricity/transmission/solar_integration_methodology.html?print`.

Hyndman, R. J. (2016). `forecast`: *Forecasting functions for time series and linear models*. R package version 7.1. Available from: `http://github.com/robjhyndman/forecast`.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., and Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis*, 55(9):2579 – 2589. Available from: `http://www.sciencedirect.com/science/article/pii/S0167947311000971`.

Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3):1–22. Available from: `http://ideas.repec.org/a/jss/jstsof/27i03.html`.

Hyndman, R. J., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2008). *Forecasting with Exponential Smoothing*. Springer, Deblik, Berlin, Germany.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., and Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3):439 – 454. Available from: `http://www.sciencedirect.com/science/article/pii/S0169207001001108`.

Ineichen, P. (2008). Conversion function between the Linke turbidity and the atmospheric water vapor and aerosol content. *Solar Energy*, 82(11):1095 – 1097. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X08001035`.

Inman, R. H., Pedro, H. T., and Coimbra, C. F. (2013). Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, 39(6):535 – 576. Available from: `http://www.sciencedirect.com/science/article/pii/S0360128513000294`.

Jager, D. and Andreas, A. (1996). NREL national wind technology center (NWTC): M2 tower; Boulder, Colorado (data); NREL report no. DA-5500-56489. Technical report, National Renewable Energy Laboratory. (accessed 1 March 2016). Available from: `http://dx.doi.org/10.5439/1052222`.

Junk, C., Delle Monache, L., Alessandrini, S., Cervone, G., and von Bremen, L. (2015). Predictor-weighting strategies for probabilistic wind power forecasting with an analog ensemble. *Meteorologische Zeitschrift*, 24(4):361–379. Available from: `http://doi.org/10.1127/metz/2015/0659`.

Kahle, D. and Wickham, H. (2013). `ggmap`: Spatial visualization with ggplot2. *The R Journal*, 5(1):144–161. Available from: `http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf`.

Kasten, F. (1965). A new table and approximation formula for the relative optical air mass. *Archiv für Meteorologie, Geophysik und Bioklimatologie Serie B*, 14:206–223.

Katzenstein, W., Fertig, E., and Apt, J. (2010). The variability of interconnected wind plants. *Energy Policy*, 38(8):4400 – 4410. Available from: `http://www.sciencedirect.com/science/article/pii/S0301421510002594`.

Khoo, Y. S., Nobre, A., Malhotra, R., Yang, D., Rüther, R., Reindl, T., and Aberle, A. G. (2014). Optimal orientation and tilt angle for maximizing in-plane solar irradiation for PV applications in Singapore. *IEEE J. Photovolt.*, 4(2):647–653. Available from: `http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6683004&queryText%3DOptimal+Orientation+and+Tilt+Angle+for+Maximizing+in-Plane+Solar+Irradiation+for+PV+Applications+in+Singapore`.

Klima, K. and Apt, J. (2015). Geographic smoothing of solar PV: results from Gujarat. *Environmental Research Letters*, 10(10):104001. Available from: `http://stacks.iop.org/1748-9326/10/i=10/a=104001`.

Kostylev, V. and Pavlovski, A. (2011). Solar power forecasting performance - towards industry standards. In *1st International Work-*

shop on the Integration of Solar Power into Power Systems. Available from: `https://ams.confex.com/ams/92Annual/webprogram/Manuscript/Paper203131/AMS_VK_%20AP_Paper%202011%20submitted.pdf`.

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *J. Chem. Metal. Min. Soc. South Africa*, 52(6):119–139.

Kubis, M. and Nobre, A. (2014). Large-area real-time solar irradiance mapping for photovoltaic applications. (accessed 1 March 2016). Available from: `http://sine.ni.com/cs/app/doc/p/id/cs-16302#`.

Laing, A. and Evans, J.-L. (2011). *Introduction to Tropical Meteorology.* University Corporation for Atmospheric Research. Available from: `http://www.goes-r.gov/users/comet/tropical/textbook_2nd_edition/index.htm`.

Lara-Fanego, V., Ruiz-Arias, J., Pozo-Vázquez, D., Santos-Alamillos, F., and Tovar-Pescador, J. (2012). Evaluation of the WRF model solar irradiance forecasts in andalusia (southern spain). *Solar Energy*, 86(8):2200 – 2217. Progress in Solar Energy 3. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X11000582`.

Laux, P., Phan, V., Lorenz, C., Thuc, T., Ribbe, L., and Kunstmann, H. (2013). Setting up regional climate simulations for southeast asia. In Nagel, W. E., Kroner, D. H., and Resch, M. M., editors, *High Performance Computing in Science and Engineering 12*, pages 391–406. Springer Berlin Heidelberg. Available from: `http://dx.doi.org/10.1007/978-3-642-33374-3_29`.

Lave, M. and Kleissl, J. (2010). Solar variability of four sites across the state of Colorado. *Renewable Energy*, 35(12):2867 – 2873.

Lave, M. and Kleissl, J. (2011). Optimum fixed orientations and benefits of tracking for capturing solar radiation in the continental United States. *Renewable Energy*,

36(3):1145 – 1152. Available from: `http://www.sciencedirect.com/science/article/pii/S0960148110003964`.

Lave, M., Kleissl, J., and Arias-Castro, E. (2012). High-frequency irradiance fluctuations and geographic smoothing. *Solar Energy*, 86(8):2190 – 2199. Progress in Solar Energy 3. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X11002611`.

Lave, M., Kleissl, J., and Stein, J. (2013). A wavelet-based variability model (WVM) for solar PV power plants. *Sustainable Energy, IEEE Transactions on*, 4(2):501–509. Available from: `http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6269913`.

Lave, M., Reno, M. J., and Broderick, R. J. (2015). Characterizing local high-frequency solar variability and its impact to distribution studies. *Solar Energy*, 118:327 – 337. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X15002881`.

Li, X.-X., Koh, T.-Y., Entekhabi, D., Roth, M., Panda, J., and Norford, L. K. (2013). A multi-resolution ensemble study of a tropical urban environment and its interactions with the background regional atmosphere. *Journal of Geophysical Research: Atmospheres*, 118(17):9804 – 9818. Available from: `http://dx.doi.org/10.1002/jgrd.50795`.

Lonij, V. P., Brooks, A. E., Cronin, A. D., Leuthold, M., and Koch, K. (2013). Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors. *Solar Energy*, 97(0):58 – 66. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X13003125`.

Lorenz, E., Hurka, J., Heinemann, D., and Beyer, H. (2009). Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. *Selected Topics*

in Applied Earth Observations and Remote Sensing, IEEE Journal of, 2(1):2 –
10.

Lubitz, W. D. (2011). Effect of manual tilt adjustments on incident irradiance on
fixed and tracking. *Appl. Energ.*, 88:1710–1719. Available from: `http://www.sciencedirect.com/science/article/pii/S030626191000471X`.

Lustbader, J. and Andreas, A. (2012). NREL vehicle testing and integration
facility (VTIF): Rotating shadowband radiometer (RSR); Golden, Colorado
(data); NREL report no. DA-5500-56490. Technical report, National Renewable Energy Laboratory. (accessed 1 March 2016). Available from: `http://dx.doi.org/10.5439/1052223`.

Luther, J. and Reindl, T. (2013). Solar photovoltaic (PV) roadmap for Singapore
(a summary). Available from: `https://www.nccs.gov.sg/sites/nccs/files/Roadmap_Solar_20140729.pdf`.

Marcos, J., Marroyo, L., Lorenzo, E., and Garca, M. (2012). Smoothing of PV power fluctuations by geographical dispersion. *Progress in Photovoltaics: Research and Applications*, 20(2):226–237. Available from: `http://onlinelibrary.wiley.com/doi/10.1002/pip.1127/abstract;jsessionid=89A477B5EB63791E611AE9F7C7C35CC4.f04t03?deniedAccessCustomisedMessage=&userIsAuthenticated=false`.

Martinez-Anido, C. B., Botor, B., Florita, A. R., Draxl, C., Lu, S., Hamann,
H. F., and Hodge, B.-M. (2016). The value of day-ahead solar power forecasting improvement. *Solar Energy*, 129:192 – 203. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X16000736`.

Matheron, G. (1963). Principles of geostatistics. *Economic Geol.*, 58(8):1246–

1266. Available from: `http://econgeol.geoscienceworld.org/content/58/8/1246.abstract`.

Mathiesen, P. and Kleissl, J. (2011). Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. *Solar Energy*, 85(5):967 – 977. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X11000570`.

McMahan, A. C., Grover, C. N., and Vignola, F. E. (2013). Evaluation of resource risk in solar-project financing. In Kleissl, J., editor, *Solar Energy Forecasting and Resource Assessment*, chapter 3, pages 49–79. Academic Press, United States of America.

Miller, S. D., Heidinger, A. K., and Sengupta, M. (2013). Physically based satellite methods. In Kleissl, J., editor, *Solar Energy Forecasting and Resource Assessment*, chapter 3, pages 49–79. Academic Press, United States of America.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, 102(D14):16663 – 16682. Available from: `http://dx.doi.org/10.1029/97JD00237`.

Monger, S. H., Morgan, E. R., Dyreson, A. R., and Acker, T. L. (2016). Applying the kriging method to predicting irradiance variability at a potential PV power plant. *Renewable Energy*, 86:602 – 610. Available from: `http://www.sciencedirect.com/science/article/pii/S0960148115302585`.

MyPower (2017a). Eligibility criteria. (accessed 4 February 2017). Available from: `https://www.mypower.com.sg/eligibility/eligibility-criteria`.

MyPower (2017b). General FAQs. (accessed 4 February 2017). Available from: `https://www.mypower.com.sg/faqs/general#q1`.

National Environment Agency (2009). WEATHERWise - Singapore. Available from: `http://www.nea.gov.sg/docs/default-source/training-knowledge-hub/guidebook-on-climate-of-singapore.pdf?sfvrsn=0)p`.

National Renewable Energy Laboratory (2014). Lowry Range solar station. Technical report. (accessed 1 March 2016). Available from: `http://dx.doi.org/10.5439/1052550`.

NCCS (2016). Singapore's approach to alternative energy. (accessed 8 August 2016). Available from: `https://www.nccs.gov.sg/climate-change-and-singapore/national-circumstances/singapores-approach-alternative-energy`.

NEA (2016). Historical PSI reading. (accessed 31 March 2016). Available from: `http://www.haze.gov.sg/haze-updates/historical-psi-readings/`.

NOAA (2015a). Global forecast system (GFS). (accessed 10 June 2015). Available from: `http://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs`.

NOAA (2015b). North American mesoscale forecast system (NAM). (accessed 10 June 2015). Available from: `http://www.ncdc.noaa.gov/data-access/model-data/model-datasets/north-american-mesoscale-forecast-system-nam`.

Nobre, A. M., Jr., C. A. S., Karthik, S., Kubis, M., Zhao, L., Martins, F. R., Pereira, E. B., Rther, R., and Reindl, T. (2016). {PV} power conversion and short-term forecasting in a tropical, densely-built environment in singapore. *Renewable Energy*, 94:496 – 509. Available from: `//www.sciencedirect.com/science/article/pii/S0960148116302592`.

Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization.* Springer. Available from: `http://link.springer.com/book/10.1007/b98874`.

Noorian, A. M., Moradi, I., and Kamali, G. A. (2008). Evaluation of 12 models to estimate hourly diffuse irradiation on inclined surfaces. *Renew. Energ.*, 33(6):1406–1412. Available from: `http://www.sciencedirect.com/science/article/pii/S0960148107002509`.

Nottrott, A. and Kleissl, J. (2010). Validation of the NSRDB-SUNY global horizontal irradiance in California. *Solar Energy*, 84:1816 – 1827. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X10002410`.

Nottrott, A., Kleissl, J., and Washom, B. (2013). Energy dispatch schedule optimization and cost benefit analysis forgrid-connected, photovoltaic-battery storage systems. *Renewable Energy*, 55:230 – 240. Available from: `http://www.sciencedirect.com/science/article/pii/S0960148112008026`.

NREL (2014a). Measurement and instrumentation data center (MIDC). Available from: `http://www.nrel.gov/midc`.

NREL (2014b). PVWatts calculator. Available from: `http://pvwatts.nrel.gov/`.

Paulson, C. A. (1970). The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer. *Journal of Applied Meteorology*, 9(6):857 – 861.

Pebesma, E. J. (2001). `gstat` user's manual. Available from: `https://www.bayceer.uni-bayreuth.de/bophy/de/top/dl/91739/gstat_doku.pdf`.

Pebesma, E. J. (2004). Multivariable geostatistics in S: the `gstat` package. *Computers & Geosciences*, 30:683–691.

Pedro, H. T. C. and Coimbra, C. F. M. (2012). Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, 86:2017–2028. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X12001429`.

Pelland, S., Galanis, G., and Kallos, G. (2013a). Solar and photovoltaic forecasting through post-processing of the global environmental multiscale numerical weather prediction model. *Progress in Photovoltaics: Research and Applications*, 21(3):284–296. Available from: `http://dx.doi.org/10.1002/pip.1180`.

Pelland, S., Remund, J., Kleissl, J., Oozeki, T., and Brabandere, K. D. (2013b). Photovoltaic and solar forecasting: State of the art. Technical report, International Energy Agency. Available from: `http://iea-pvps.org/index.php?id=278`.

Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., and Vignola, F. (2002). A new operational model for satellite-derived irradiances: description and validation. *Solar Energy*, 73(5):307 – 317. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X02001226`.

Perez, R., Ineichen, P., Seals, R., Michalsky, J., and Stewart, R. (1990a). Modeling daylight availability and irradiance components from direct and global irradiance. *Sol. Energy*, 44(5):271289. Available from: `http://www.sciencedirect.com/science/article/pii/0038092X9090055H`.

Perez, R., Ineichen, P., Seals, R., and Zelenka, A. (1990b). Making full use of the clearness index for parameterizing hourly insolation conditions. *Solar Energy*, 45(2):111 – 114. Available from: `http://www.sciencedirect.com/science/article/pii/0038092X9090036C`.

Perez, R., Kivalov, S., Schlemmer, J., Jr., K. H., and Hoff, T. (2011). Parameteri-

zation of site-specific short-term irradiance variability. *Solar Energy*, 85(7):1343 – 1353. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X11000995`.

Perez, R., Lorenz, E., Pelland, S., Beauharnois, M., Knowe, G. V., Jr., K. H., Heinemann, D., Remund, J., Muller, S. C., Traunmuller, W., Steinmauer, G., Pozo, D., Ruiz-Arias, J. A., Lara-Fanego, V., Ramirez-Santigosa, L., Gaston-Romero, M., and Pomares, L. M. (2013). Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. *Solar Energy*, 94(0):305 – 326. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X13001886`.

Perez, R., Seals, R., Ineichen, P., Stewart, R., and Menicucci, D. (1987). A new simplified version of the Perez diffuse irradiance model for tilted surfaces. *Sol. Energy*, 39(3):221–231. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X87800312`.

Perez, R., Stewart, R., Seals, R., and Guertin, T. (1988). The development and verification of the Perez diffuse radiation model. Technical report, Atmospheric Sciences Research Center, SUNY at Albany, Albany, NY. Available from: `http://prod.sandia.gov/techlib/access-control.cgi/1988/887030.pdf`.

Qian, P. Z. G. and Wu, F. J. (2008). Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, 50:192 – 204. Available from: `http://www.tandfonline.com/doi/abs/10.1198/004017008000000082#.VHLColeUdPo`.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from: `https://www.R-project.org/`.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, 113:1155–1174.

Remund, J., Calhau, C., Perret, L., and Marcel, D. (2015). Characterization of the spatio-temporal variations and ramp rates of solar radiation and PV. Technical report. (accessed 1 March 2016). Available from: `http://iea-pvps.org/index.php?id=336`.

Ricci, V. (2005). R functions for regression analysis. `<http://cran.r-project.org/doc/contrib/Ricci-refcard-regression.pdf>`. (accessed 29.06.2011).

Rubel, F. and Kottek, M. (2010). Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen–Geiger climate classification. *Meteorol. Z.*, 19:135 – 141. Available from: `http://koeppen-geiger.vu-wien.ac.at/pdf/Paper_2010.pdf`.

Ruf, T. (1999). The lomb-scargle periodogram in biological rhythm research: Analysis of incomplete and unequally spaced time-series. *Biological Rhythm Research*, 30:178–201.

Ruiz-Arias, J. A., Pozo-Vázquez, D., Sánchez-Sánchez, N., Montávez, J. P., Hayas-Barru, A., and Tovar-Pescador, J. (2008). Evaluation of two MM5-PBL parameterization for solar radiation and temperature estimation in the South-Eastern area of the Iberian peninsula. *Il Nuovo Cimento*, 31(5-6):825 – 842.

Sengupta, M. and Andreas, A. (2010). Oahu solar measurement grid (1–year archive); NREL report no. DA–5500–56506. Technical report, National Renewable Energy Laboratory. (accessed 8 June 2015). Available from: `http://www.nrel.gov/midc/oahu_archive/`.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Duda, M., Huang, X. Y., and Wang, W. (2008). A description of the advanced research WRF version 3. Technical Report NCAR/TN-475+STR, National Center for Atmospheric Research, Boulder, CO. Available from: `http://www2.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf`.

Sloughter, J. M., Gneiting, T., and Raftery, A. E. (2010). Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *Journal of the American Statistical Association*, 105(489):25–35. Available from: `http://dx.doi.org/10.1198/jasa.2009.ap08615`.

Sloughter, J. M., Raftery, A. E., Gneiting, T., and Fraley, C. (2007). Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Monthly Weather Review*, 135(9):3209–3220. Available from: `doi.org/10.1175/MWR3441.1`.

The R Core Team (2014). R: A language and environment for statistical computing (reference index). Available from: `http://cran.r-project.org/doc/manuals/r-devel/fullrefman.pdf`.

Thorey, J., Mallet, V., Chaussin, C., Descamps, L., and Blanc, P. (2015). Ensemble forecast of solar radiation using TIGGE weather forecasts and HelioClim database. *Solar Energy*, 120:232 – 243. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X15003576`.

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73:273–282. Available from: `http://econpapers.repec.org/article/blajorssb/v_3a73_3ay_3a2011_3ai_3a3_3ap_3a273-282.htm`.

Tursilowati, L., Sumantyo, J. T. S., Kuze, H., and Adiningsih, E. S. (2011). The

integrated WRF urban modeling system and its application to monitoring urban heat island in Jakarta, Indonesia. *Journal of Urban and Environmental Engineering*, 6(1):1 – 9.

UCAR (2011). Challenges of tropical weather forecasting. (accessed 31 March 2016). Available from: `http://www.goes-r.gov/users/comet/tropical/textbook_2nd_edition/navmenu.php_tab_10_page_1.0.0.htm`.

Vaid, B. H. (2013). Numerical simulations and analysis of June 16, 2010 heavy rainfall event over Singapore using the WRFV3 model. *International Journal of Atmospheric Sciences*, 2013(2013). Available from: `http://dx.doi.org/10.1155/2013/825395`.

van Haaren, R., Morjaria, M., and Fthenakis, V. (2014). Empirical assessment of short-term variability from utility-scale solar PV plants. *Progress in Photovoltaics: Research and Applications*, 22(5):548–559. Available from: `http://dx.doi.org/10.1002/pip.2302`.

Veenhuis, B. A. (2013). Spread calibration of ensemble MOS forecasts. *Mon. Wea. Rev.*, 141:2467–2482.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York: Springer.

Wang, W., Bruyére, C., Duda, M., Dudhia, J., Gill, D., Kavulich, M., Keene, K., Lin, H.-C., Michalakes, J., Rizvi, S., Zhang, X., Beezley, J. D., Coen, J. L., Mandel, J., Chuang, H.-Y., McKee, N., Slovacek, T., and Wolff, J. (2008). ARW version 3 modeling system user's guide. Technical report, National Center for Atmospheric Research, Boulder, CO. Available from: `http://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3/ARWUsersGuideV3.pdf`.

Wellby, S. J. and Engerer, N. A. (2016). Categorizing the meteorological origins of critical ramp events in collective photovoltaic array output. *Journal of Applied Meteorology and Climatology*, 55(6):1323–1344. Available from: `http://dx.doi.org/10.1175/JAMC-D-15-0107.1`.

West, S. R., Rowe, D., Sayeef, S., and Berry, A. (2014). Short-term irradiance forecasting using skycams: Motivation and development. *Solar Energy*, 110:188 – 207. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X14004150`.

Whitaker, J. S. and Loughe, A. F. (1998). The Relationship between Ensemble Spread and Ensemble Mean Skill. *Monthly Weather Review*, 126(12):3292–3302.

Wickham, H. (2009). `ggplot2`: *Elegant Graphics for Data Analysis*. Springer-Verlag New York. Available from: `http://ggplot2.org`.

Wilcox, S. (2012). National solar radiation database 1991-2010 update: User's manual. Technical report, National Renewable Energy Laboratory, Golden, CO. Available from: `http://www.nrel.gov/docs/fy12osti/54824.pdf`.

Wirth, H. (2017). Recent facts about photovoltaics in germany. Available from: `http://www.nrel.gov/docs/fy10osti/47781.pdf`.

Xiang, Y., Gubian, S., Suomela, B., and Hoeng, J. (2013). Generalized simulated annealing for efficient global optimization: the `GenSA` package for R. *The R Journal Volume 5/1, June 2013*. Available from: `http://journal.r-project.org/`.

Yan, R., Saha, T. K., Meredith, P., and Goodwin, S. (2013). Analysis of year-long performance of differently tilted photovoltaic systems in Brisbane, Australia. *Energy Conversion and Management*, 74:102–108. Available from: `http://www.sciencedirect.com/science/article/pii/S0196890413002549`.

Yang, D. (2014). *Solar Irradiance Modeling and Forecasting using Novel Statistical Techniques.* PhD thesis, Department of Electrical and Computer Engineering, National University of Singapore.

Yang, D. (2015). Space-time kriging of solar irradiance: A review of methods and practices. unpublished.

Yang, D., Dong, Z., Reindl, T., Jirutitijaroen, P., and Walsh, W. M. (2014a). Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage. *Solar Energy*, 103(0):550 – 562. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X14000425`.

Yang, D., Gu, C., Dong, Z., Jirutitijaroen, P., and Chen, N. (2013). Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging. *Renewable Energy*, 60:235–245. Available from: `http://www.sciencedirect.com/science/article/pii/S0960148113002759`.

Yang, D., Jirutitijaroen, P., and Walsh, W. M. (2012). Hourly solar irradiance time series forecasting using cloud cover index. *Solar Energy*, 86(12):3531 – 3543. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X12003039`.

Yang, D., Sharma, V., Ye, Z., Lim, L. I., Zhao, L., and Aryaputera, A. W. (2015a). Forecasting of global horizontal irradiance by exponential smoothing, using decompositions. *Energy*, 81(0):111 – 119. Available from: `http://www.sciencedirect.com/science/article/pii/S0360544214013528`.

Yang, D., Walsh, W. M., and Jirutitijaroen, P. (2014b). Estimation and applications of clear sky global horizontal irradiance at the equator. *Journal of Solar Energy Engineering*, 136(August):1–4.

Yang, D., Ye, Z., Lim, L. H. I., and Dong, Z. (2015b). Very short term irradiance forecasting using the lasso. *Solar Energy*, 114:314 – 326. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X15000304`.

Yang, H., Kurtz, B., Nguyen, D., Urquhart, B., Chow, C. W., Ghonima, M., and Kleissl, J. (2014c). Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego. *Solar Energy*, 103:502–524. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X14001327`.

Ye, J.-Y., Ding, K., Reindl, T., and Aberle, A. G. (2013). Outdoor PV module performance under fluctuating irradiance conditions in tropical climates. *Energy Procedia*, 33:135 – 141. Available from: `http://www.sciencedirect.com/science/article/pii/S187661021300074X`.

Yuen, R., Gneiting, T., Thorarinsdottir, T., and Fraley, C. (2013). `ensembleMOS`: *Ensemble Model Output Statistics*. R package version 0.7. Available from: `http://CRAN.R-project.org/package=ensembleMOS`.

Zamo, M., Mestre, O., Arbogast, P., and Pannekoucke, O. (2014). A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production. Part II: Probabilistic forecast of daily production. *Solar Energy*, 105:804 – 816. Available from: `http://www.sciencedirect.com/science/article/pii/S0038092X14001601`.

Zamora, R. J., Dutton, E. G., Trainer, M., McKeen, S. A., Wilczak, J. M., and Hou, Y.-T. (2005). The accuracy of solar irradiance calculations used in mesoscale numerical weather prediction. *Monthly Weather Review*, 133(4):783 – 792. Available from: `http://dx.doi.org/10.1175/MWR2886.1`.

Zhang, Q., Deng, X., Qian, P. Z. G., and Wang, X. (2013). Spatial modeling for refining and predicting surface potential mapping with enhanced resolution.

Nanoscale, 5:921 – 926. Available from: `http://pubs.rsc.org/en/Content/` `ArticleLanding/2013/NR/c2nr33603k#!divAbstract`.

Zhao, L., Aryaputera, A. W., Yang, D., Walsh, W. M., and Reindl, T. (2015). Modeling of solar variability and geographic smoothing effect. In *PV Asia Scientific Conference.* (oral presentation). Available from: `http://www.asiacleanenergysummit.com/events/asia-clean-energy-` `summit-2015/custom-151-ef92883b35e14a42b65a73522cc02443.aspx`.

Appendix A

Detailed forecast results of chapter 3

This chapter displays graphically the results of the power output forecasts of the Australian Capital Territory (ACT) photovoltaic (PV) network of chapter 3. Figs. A.1 - A.13 are the plots of the forecast and observed collective power output time series for the days listed in Table 3.1. Five forecasting methods used are persistence (Pers), independent ARIMA and lasso (ARIMA.IND and lasso.IND), and bottom-up ARIMA and lasso (ARIMA.BU and lasso.BU). The forecast variable is the power output (in kW/kW_p). The clear sky power output time series are modeled based on Engerer and Mills (2014).
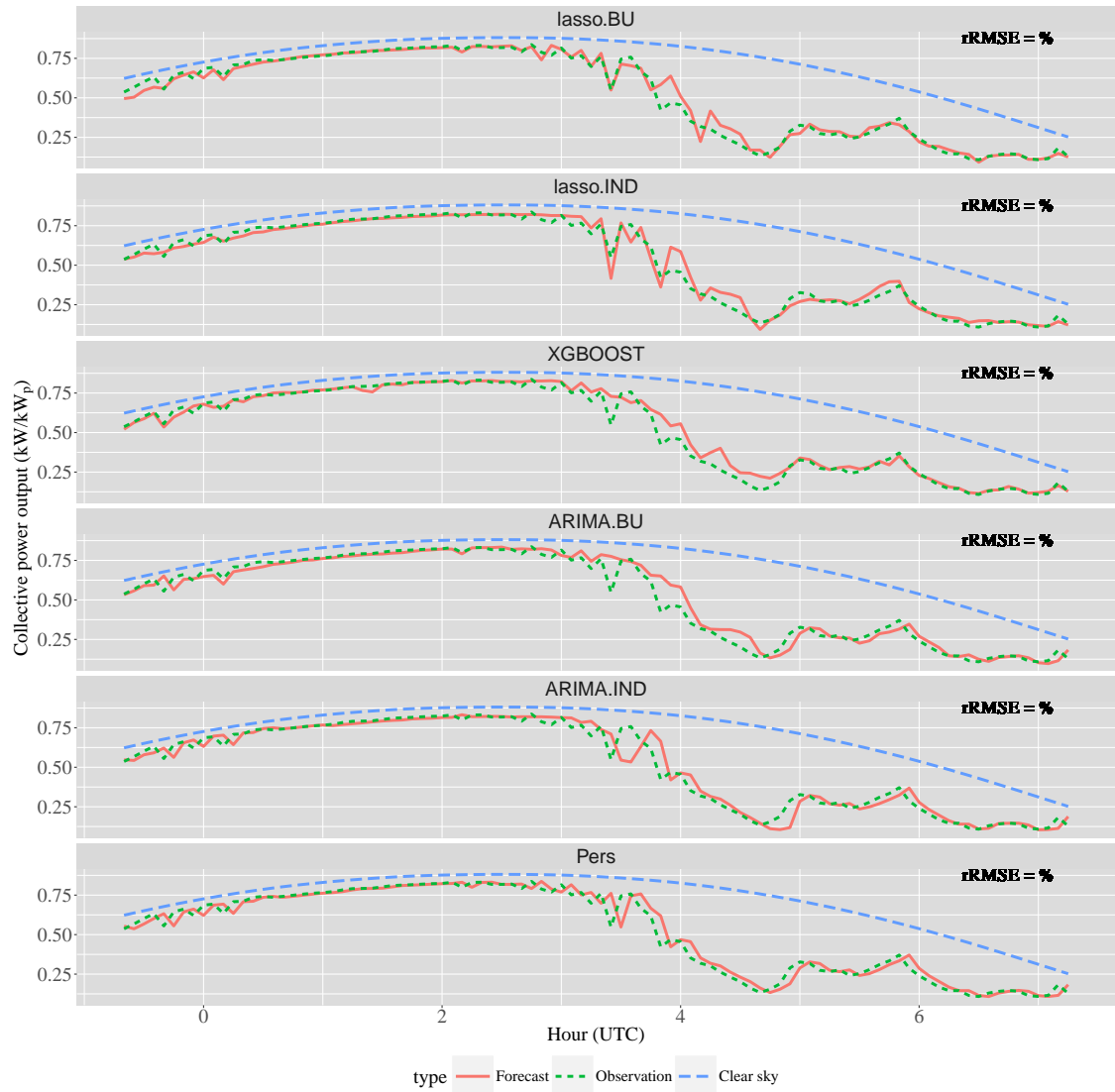
Fig. A.1: Plot of forecast and observed collective power output time series of the ACT PV network of 9 February 2013 (20% learning data). The time is stated in the Universal Coordinated Time (UTC).
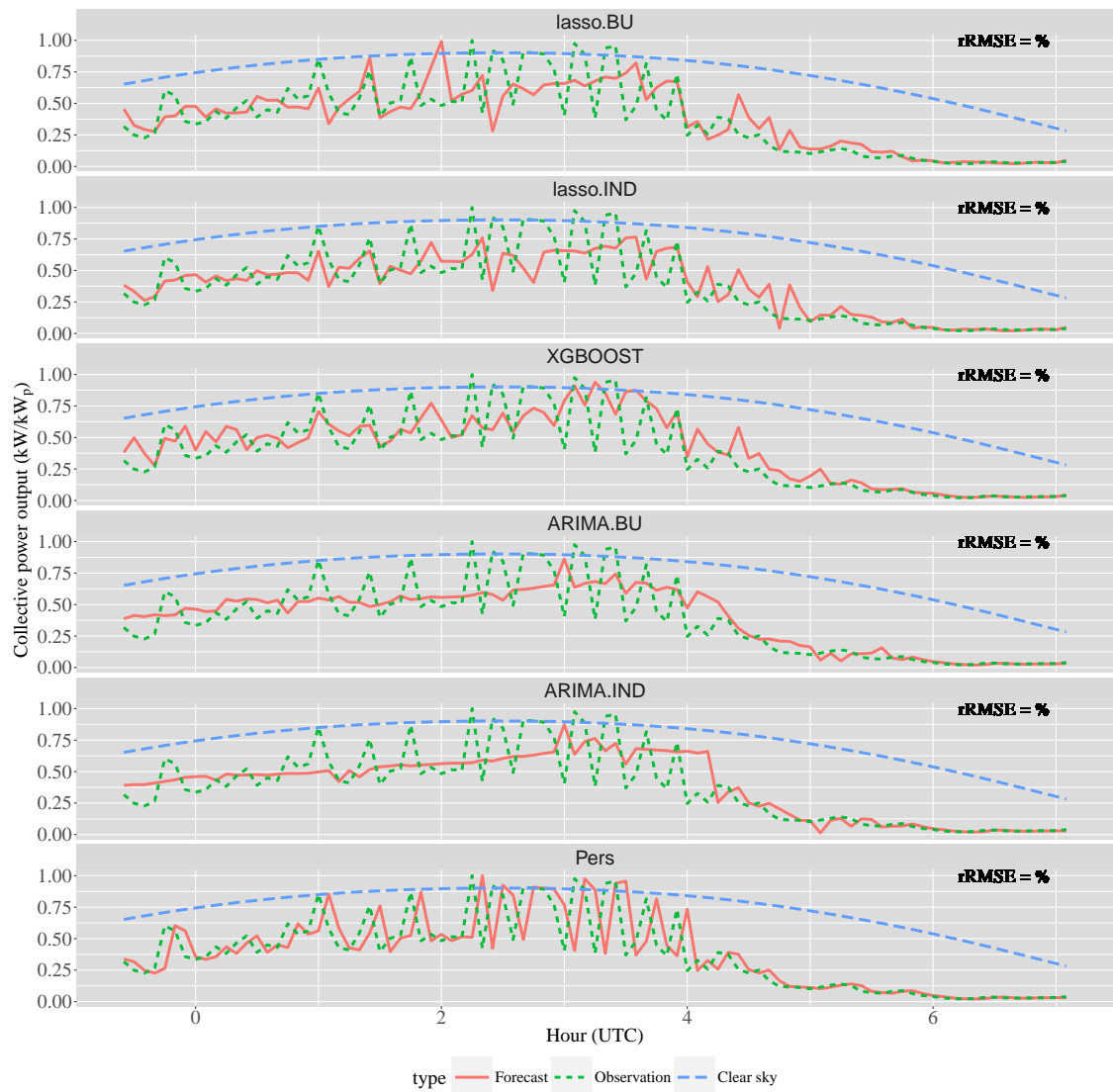
Fig. A.2: As in Fig. A.1, but for 20 February 2013.
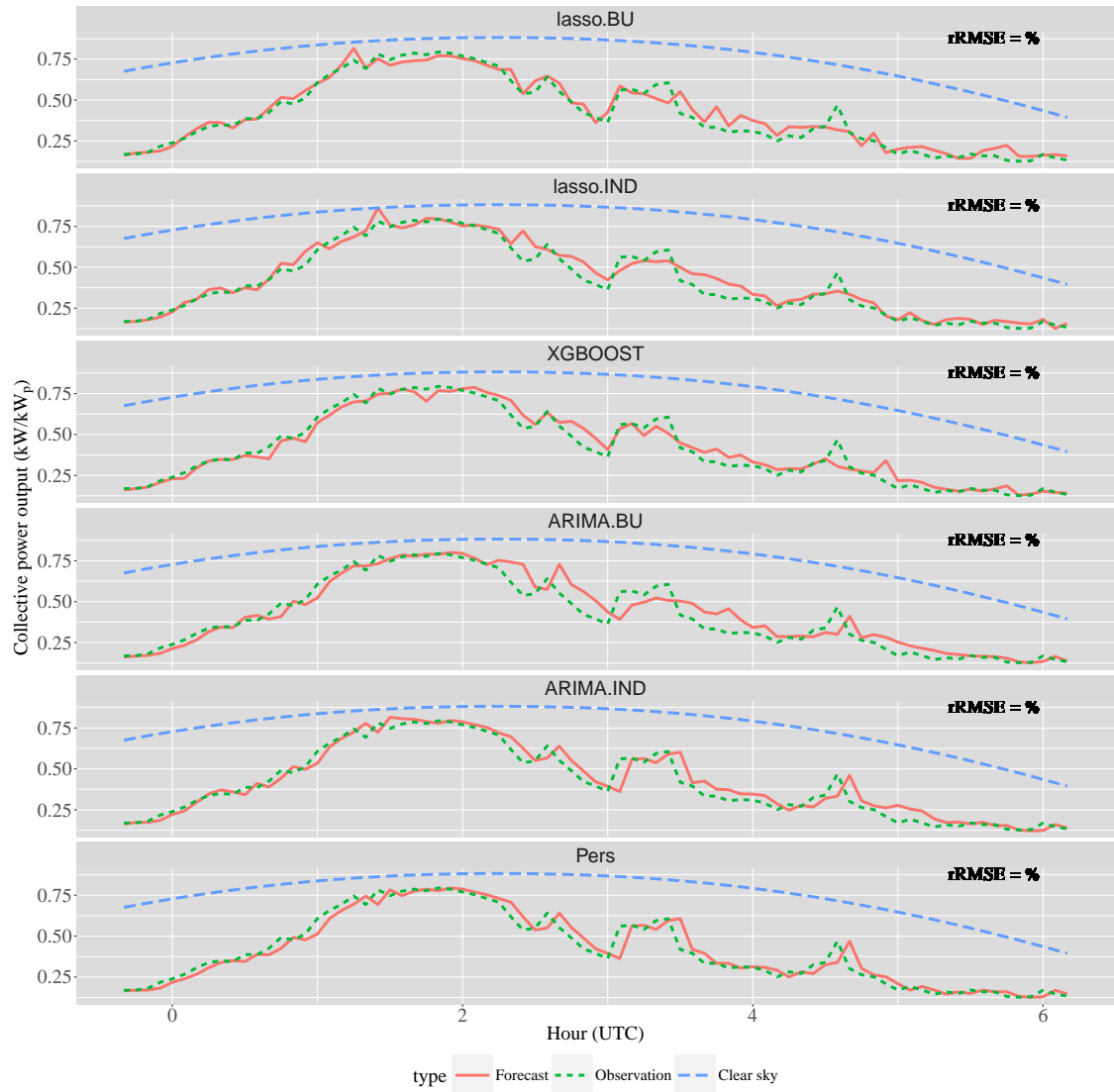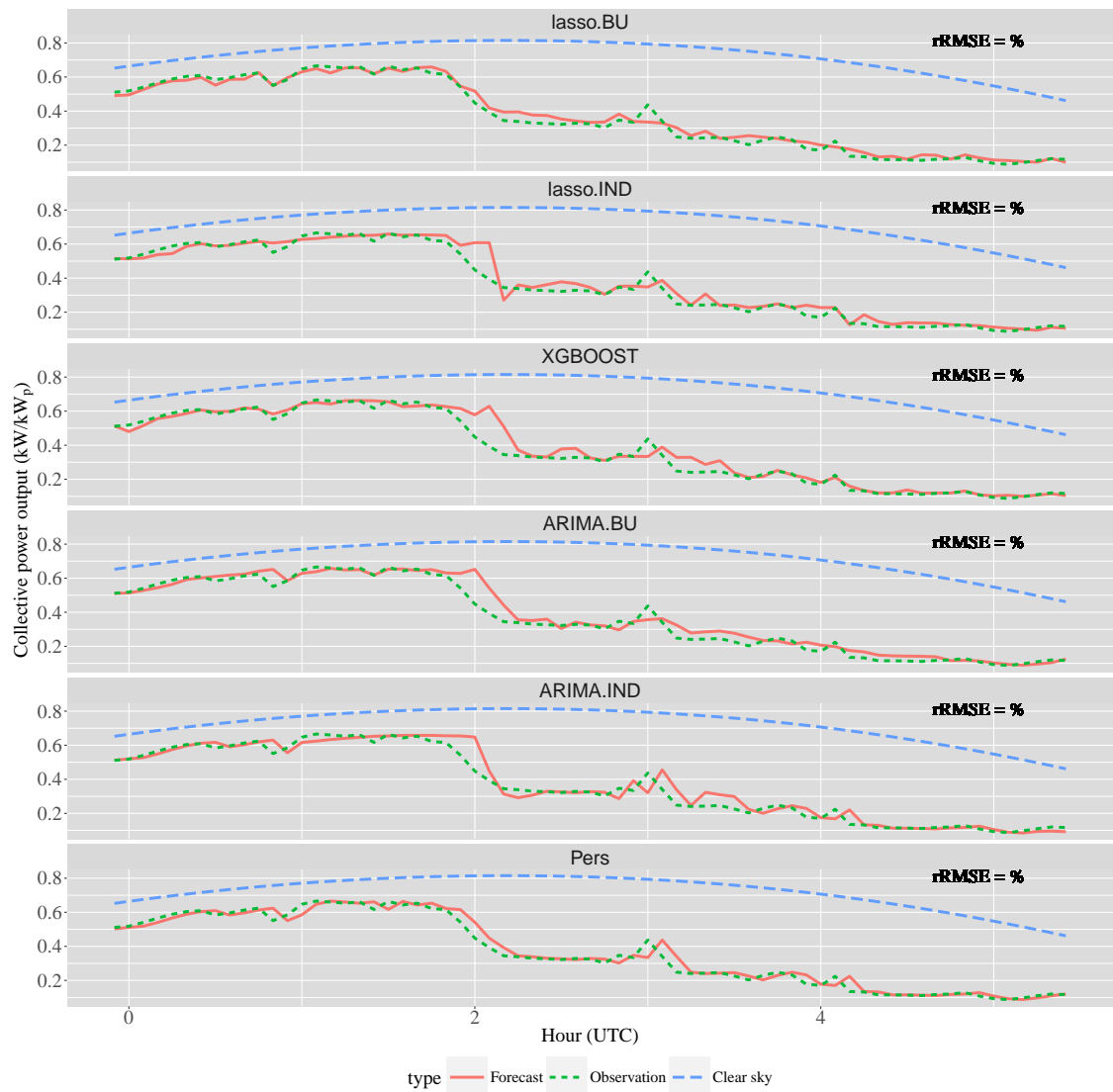
Fig. A.3: As in Fig. A.1, but for 31 March 2013.
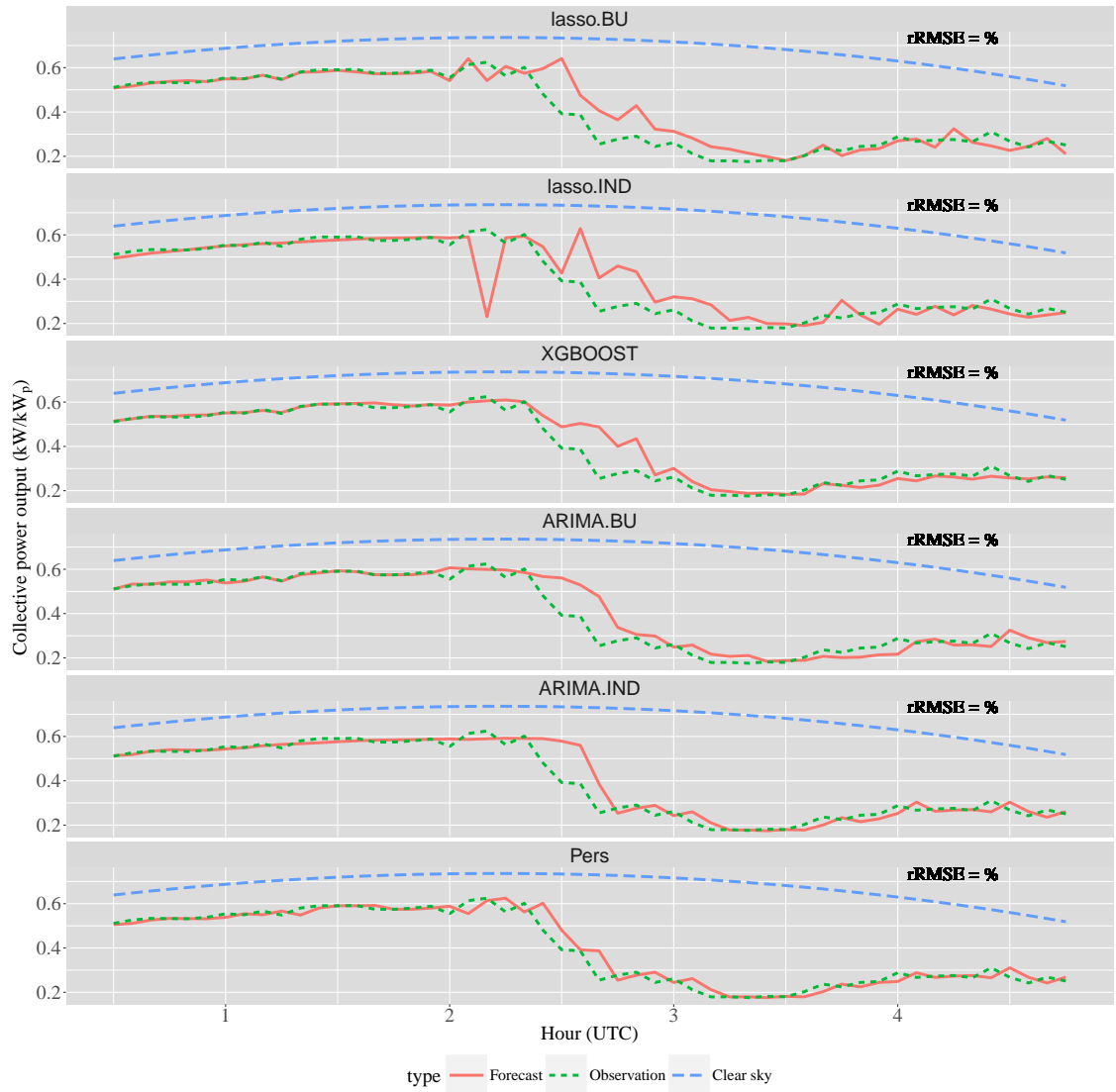
Fig. A.4: As in Fig. A.1, but for 29 April 2013.
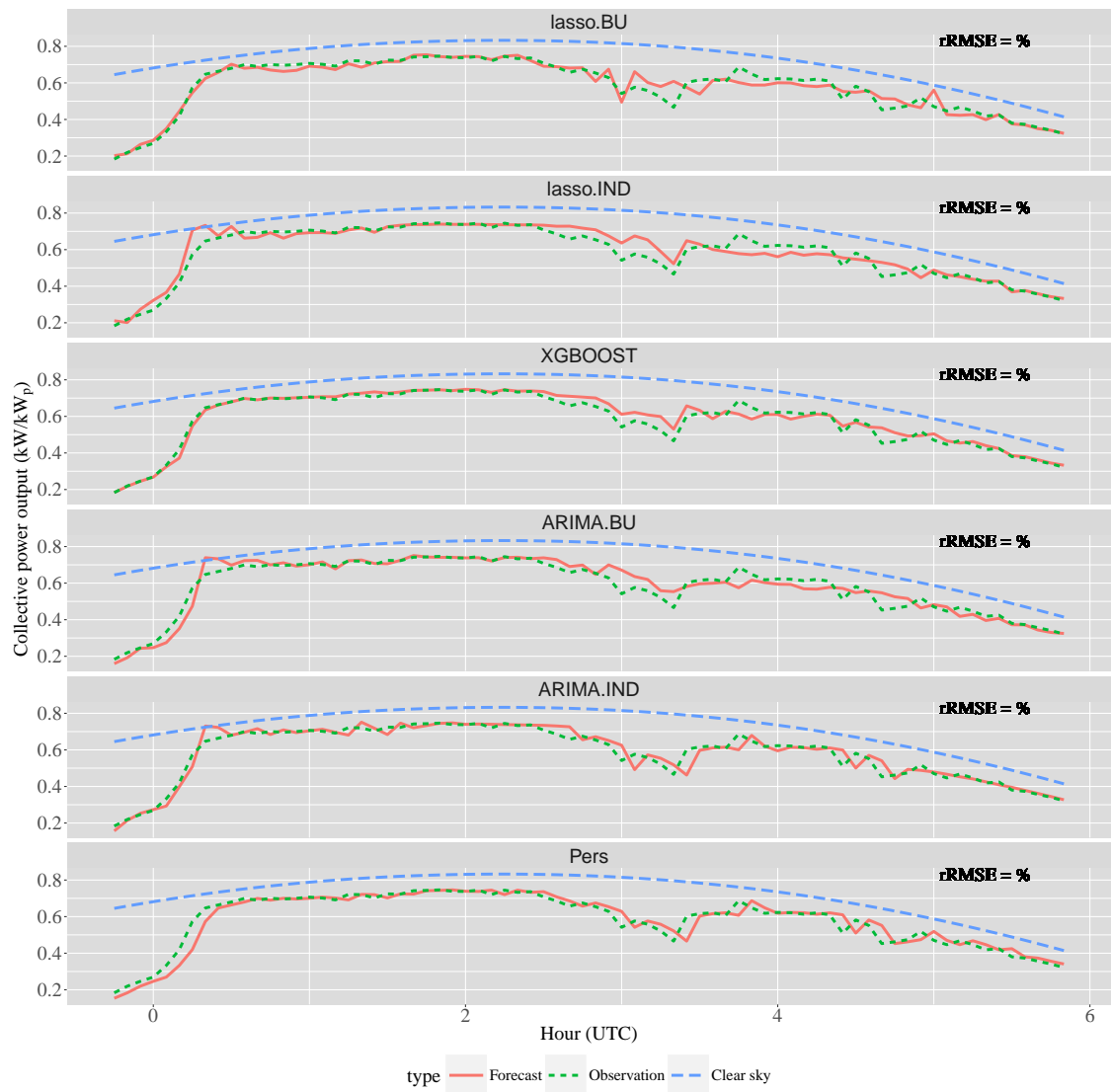
Fig. A.5: As in Fig. A.1, but for 22 June 2013.

Fig. A.6: As in Fig. A.1, but for 31 August 2013.

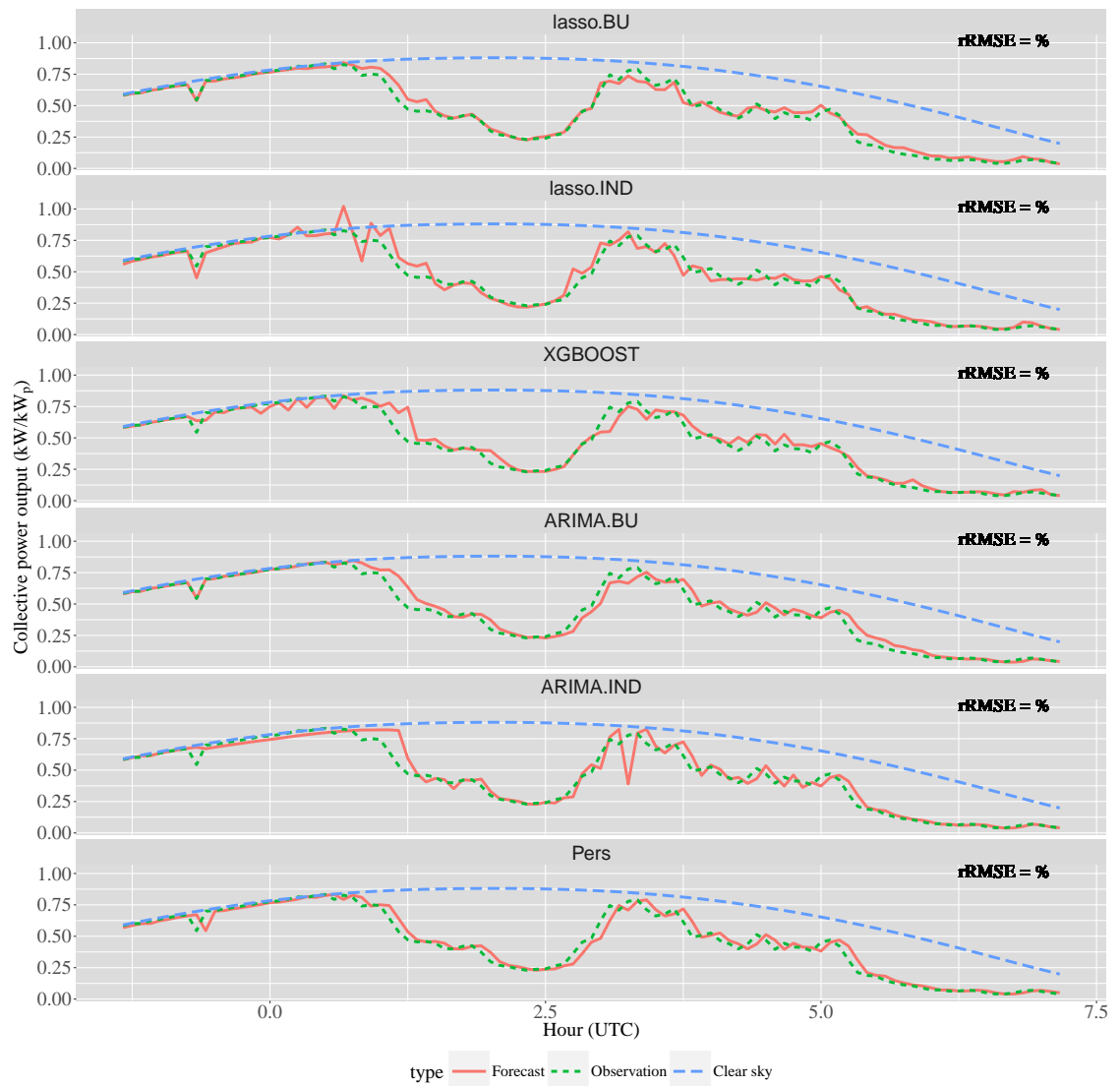186

Fig. A.7: As in Fig. A.1, but for 13 October 2013.

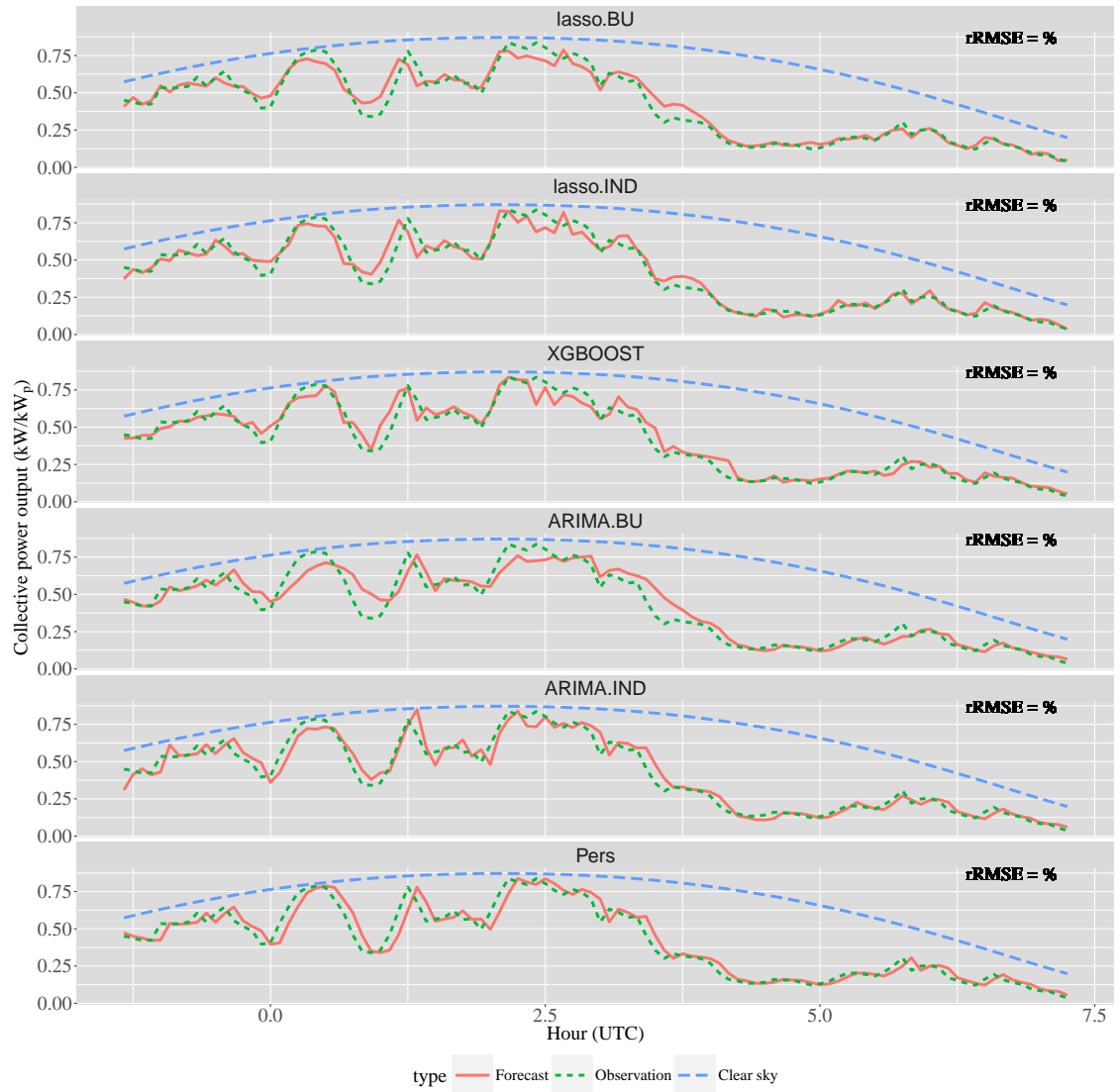Fig. A.8: As in Fig. A.1, but for 28 November 2013.
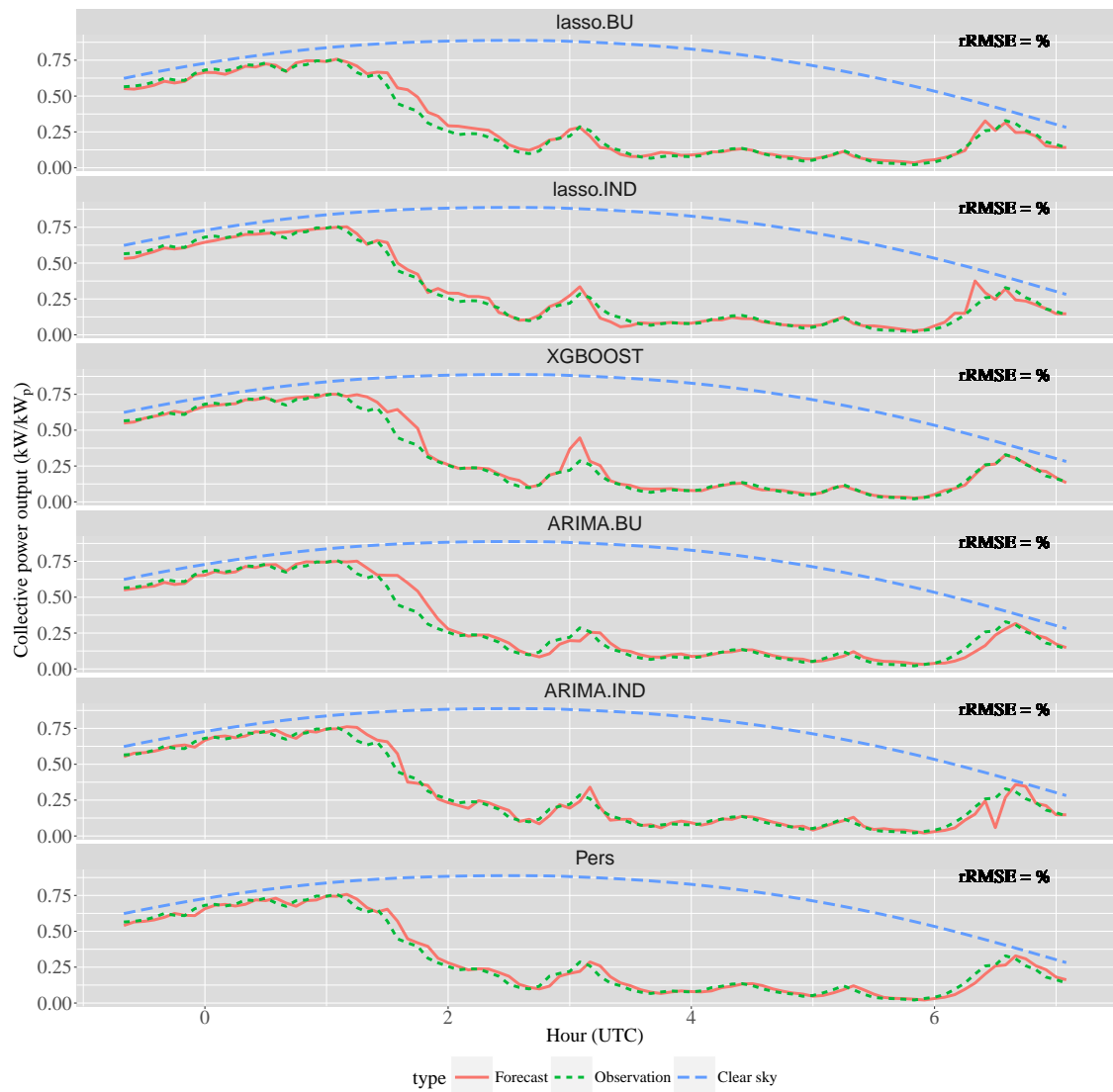
Fig. A.9: As in Fig. A.1, but for 9 December 2013.

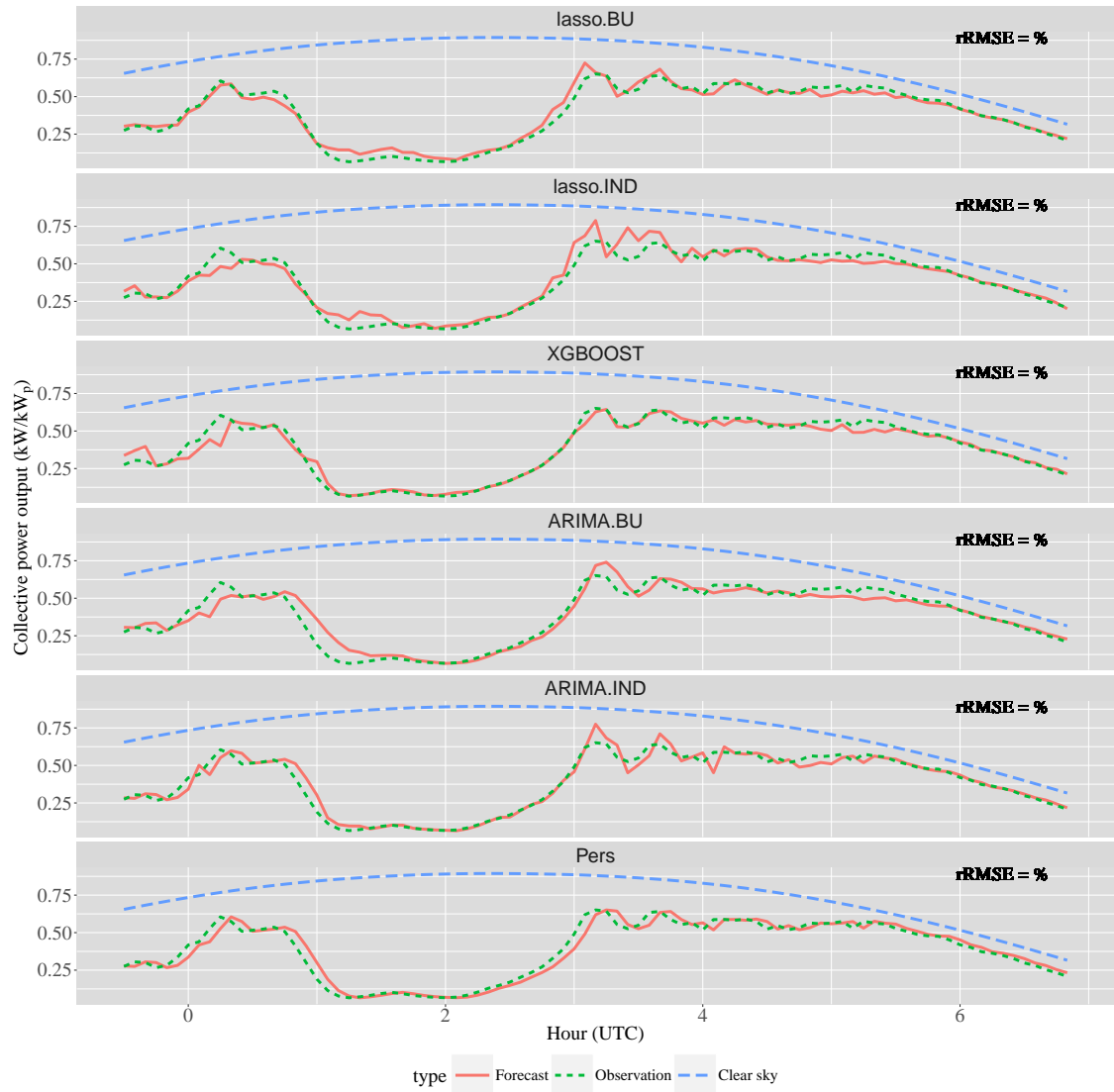Fig. A.10: As in Fig. A.1, but for 19 February 2014.
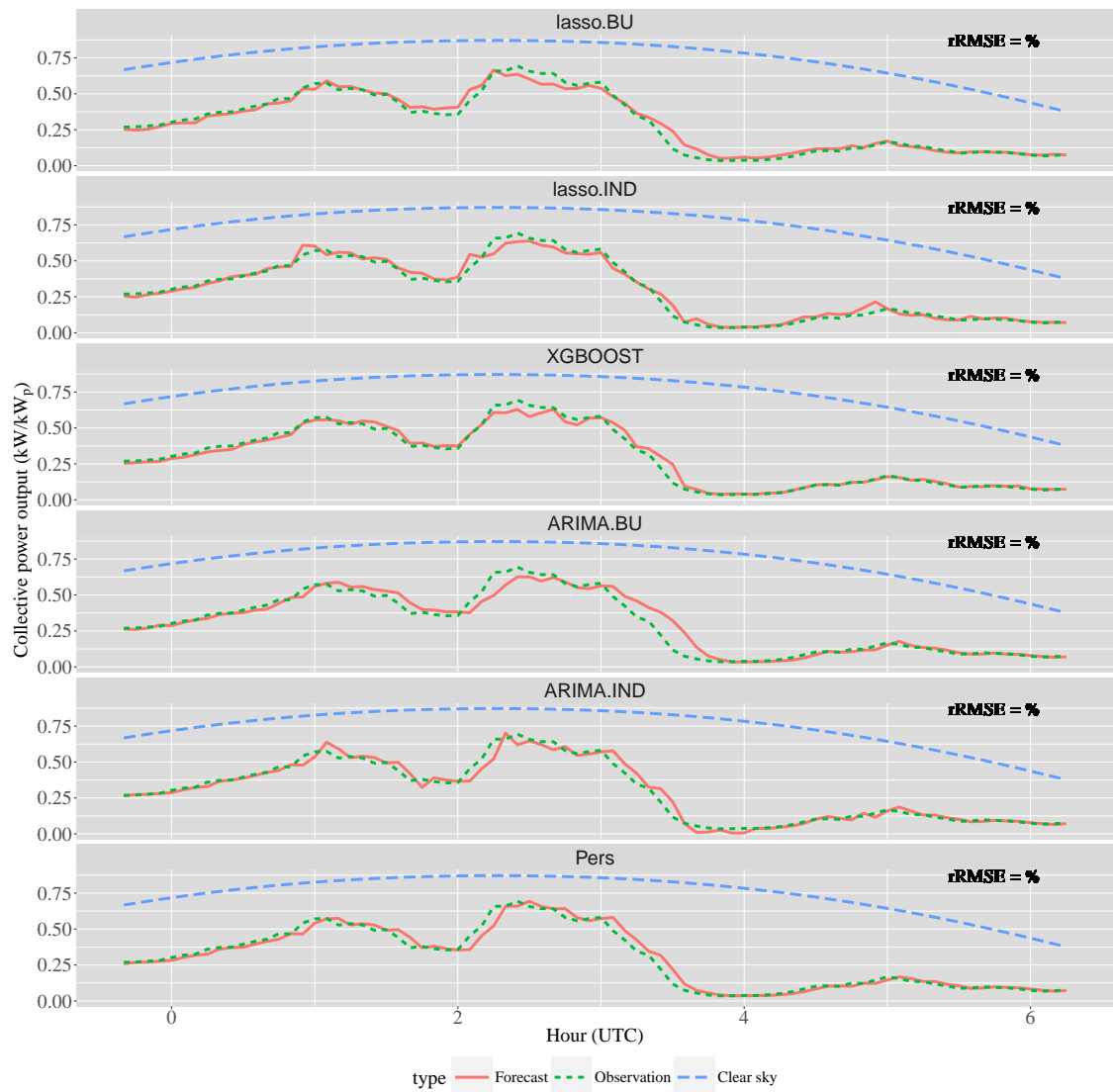
Fig. A.11: As in Fig. A.1, but for 5 March 2014.

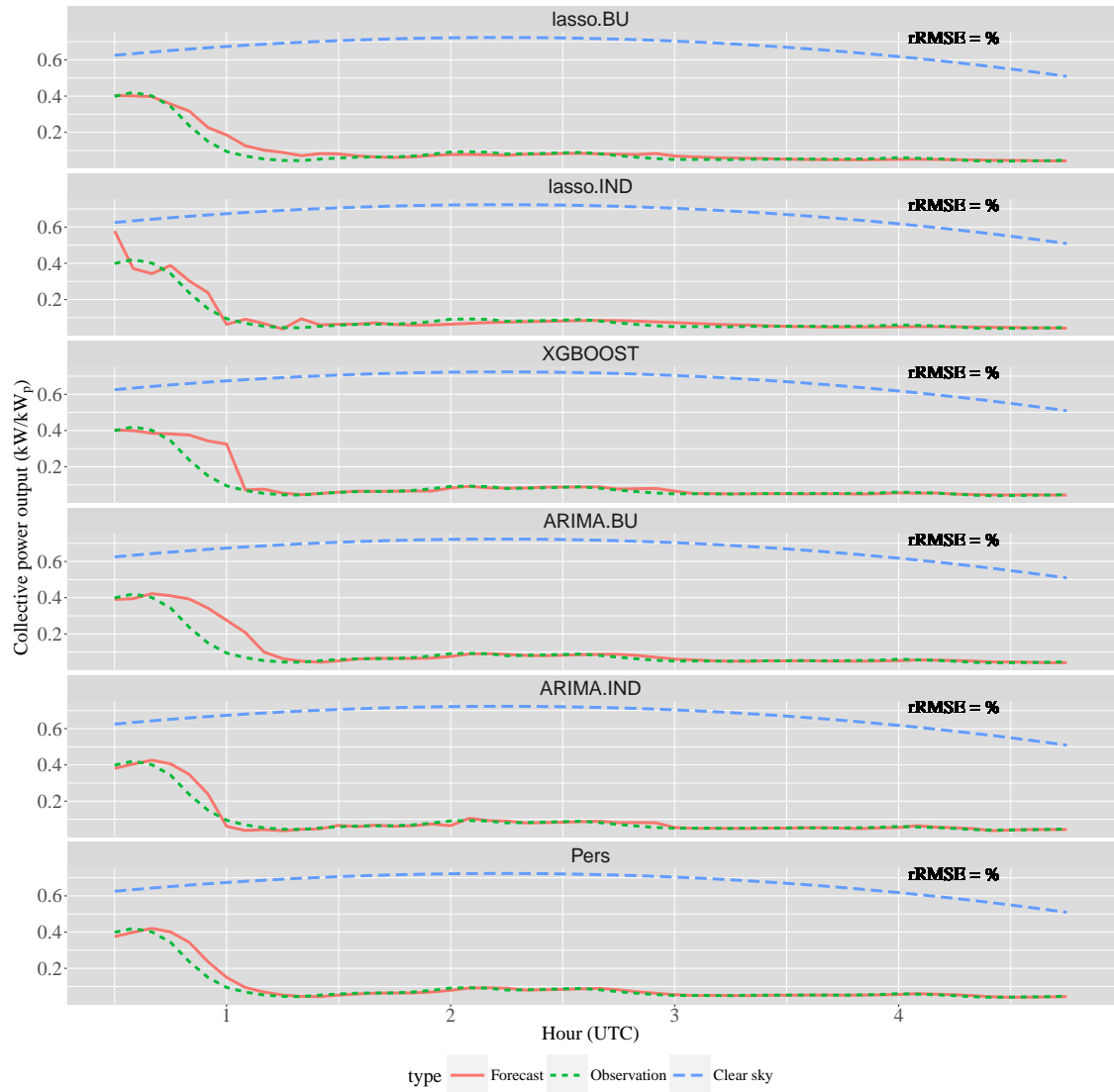Fig. A.12: As in Fig. A.1, but for 30 March 2014.

Fig. A.13: As in Fig. A.1, but for 28 June 2014.

Appendix B

Transposition model

This chapter presents the irradiance transposition model that is used in chapter 6 of this thesis.

B.1 Modeling global solar irradiance on a tilted plane

I_t on a tilted plane could be defined in Eq. (B.1):

$$I_t = I_{\text{Dir}} \cos\theta_{\text{inc}} + I_{\text{Dif}} R_d + \mathfrak{P} I_{\text{Glo}} R_r \tag{B.1}$$

where I_{Dir} represents the direct normal irradiance (DNI), θ_{inc} symbolizes the incidence angle of sun rays on the tilted plane, I_{Dif} is the diffuse horizontal irradiance (DHI), R_d represents the diffuse transposition factor, $\mathfrak{P}$ is the albedo of the foreground, I_{Glo} symbolizes the global horizontal irradiance (GHI), and R_r is the transposition factor for ground reflection (Gueymard, 2009).

θ_{inc} could be computed using purely geometrical relations. In chapter 6, θ_{inc} values are computed using `insol` package (Corripio, 2014) in the R statistical software (The R Core Team, 2014) . The value of $\mathfrak{P}$ is assumed to be 0.2 for the whole year (Gueymard, 2009) due to absence of albedo measurement data. Based

on Yang et al. (2013), R_r is considered to be isotropic as shown in Eq. (B.2):

$$R_r = \frac{(1 - \cos \omega)}{2} \tag{B.2}$$

where ω is plane tilt angle.

To calculate R_d, Perez model (Perez et al., 1990a) is adopted. Perez model is an anisotropic model. Anisotropic model tends to allocate more diffuse radiation on a tilted plane facing the sun than that of the opposite direction (Yang et al., 2013).

B.2   Perez diffuse irradiance model

In order to calculate R_d according to Perez model, firstly extraterrestrial direct normal irradiance (I_o) needs to be calculated. The formula is shown in Eq. (B.3):

$$I_o = I_{sc} \times r_e \tag{B.3}$$

where r_e is the earth eccentricity correction factor, calculated according to Eq. (2.5), and I_{sc} is the solar constant ($= 1362$ W/m^2).

The next step is to calculate the sky clearness index ($\mathfrak{e}$), sky brightness (κ), and air mass (m_{air}) as shown in Eqs. (B.4) - (B.6) respectively:

$$\mathfrak{e} = \left[\frac{I_{Dif} + I_{Dir}}{I_{Dif}} + 1.041 \left(\frac{\pi \times \theta_z}{180} \right)^3 \right] \div \left[1 + 1.041 \left(\frac{\pi \times \theta_z}{180} \right)^3 \right] \tag{B.4}$$

$$\kappa = I_{Dif} \times \frac{m_{air}}{I_o} \tag{B.5}$$

$$m_{air} = \frac{1}{\cos \left[\frac{\theta_z \times \pi}{180} + 0.15 \left(90 - \theta_z + 3.885 \right)^{-1.235} \right]} \tag{B.6}$$

where θ_z is solar zenith angle in degree. The detailed calculation of m_{air} can be found in Kasten (1965).

Finally, the circumsolar brightening coefficient (F_1), horizon brightening coefficient (F_2), and R_d can be computed according to Eqs. (B.7) - (B.9) respectively.

$$F_1 = \max\left\{0, F_{11} + F_{12} \times \kappa + F_{13} \times \left(\frac{\pi \times \theta_z}{180}\right)\right\} \tag{B.7}$$

$$F_2 = F_{21} + F_{22} \times \kappa + F_{23} \times \left(\frac{\pi \times \theta_z}{180}\right) \tag{B.8}$$

$$R_d = (1 - F_1) \times \frac{1 + \cos\omega}{2} + F_1 \times \frac{a}{b} + F_2 \times \sin\omega \tag{B.9}$$

where F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, and F_{23} are determined according to Table B.1. The value of a is $\max(0, \cos\theta_{\text{inc}})$ while that of b is $\max(0.087, \cos\theta_z)$.

Table B.1: Perez Coefficients

ε	F_{11}	F_{12}	F_{13}	F_{21}	F_{22}	F_{23}
$[1, 1.065)$	-0.0083	0.5877	-0.0621	-0.0596	0.0721	-0.022
$[1.065, 1.23)$	0.1299	0.6826	-0.1514	-0.0189	0.066	-0.0289
$[1.23, 1.5)$	0.3297	0.4869	-0.2211	0.0554	-0.064	-0.0261
$[1.5, 1.95)$	0.5682	0.1875	-0.2951	0.1089	-0.1519	-0.014
$[1.95, 2.8)$	0.873	-0.392	-0.3616	0.2256	-0.462	0.0012
$[2.8, 4.5)$	1.1326	-1.2367	-0.4118	0.2878	-0.823	0.0559
$[4.5, 6.2)$	1.0602	-1.5999	-0.3589	0.2642	-1.1272	0.1311
$[6.2, +\infty)$	0.6777	-0.3273	-0.2504	0.1516	-1.3765	0.2506