

Department of Electrical Engineering and Automation

# Statistical Methods for Variable Renewable Energy Generation Modelling

---

Jussi Ekström

# Statistical Methods for Variable Renewable Energy Generation Modelling

**Jussi Ekström**

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held at the lecture hall AS1 of the school on 11 May 2018 at 12.

**Aalto University**  
**School of Electrical Engineering**  
**Department of Electrical Engineering and Automation**  
**Power Systems and High Voltage Engineering**

**Supervising professor**

Prof. Matti Lehtonen, Aalto University, Finland

**Thesis advisors**

University Teacher Emeritus Ilkka Mellin, Aalto University, Finland

Dr. Matti Koivisto, Technical University of Denmark, Denmark

**Preliminary examiners**

Prof. Mikko Kolehmainen, University of Eastern Finland, Finland

Prof. Rimantas Deksnys, Kaunas University of Technology, Lithuania

**Opponent**

Prof. Peter Schegner, Technische Universität Dresden, Germany

Aalto University publication series

**DOCTORAL DISSERTATIONS 42/2018**

© 2018 Jussi Ekström

ISBN 978-952-60-7881-6 (printed)

ISBN 978-952-60-7882-3 (pdf)

ISSN-L 1799-4934

ISSN 1799-4934 (printed)

ISSN 1799-4942 (pdf)

<http://urn.fi/URN:ISBN:978-952-60-7882-3>

Unigrafia Oy

Helsinki 2018

Finland



**Author**

Jussi Ekström

**Name of the doctoral dissertation**

Statistical Methods for Variable Renewable Energy Generation Modelling

**Publisher** School of Electrical Engineering

**Unit** Department of Electrical Engineering and Automation

**Series** Aalto University publication series DOCTORAL DISSERTATIONS 42/2018

**Field of research** Power Systems and High Voltage Engineering

**Manuscript submitted** 19 December 2017

**Date of the defence** 11 May 2018

**Permission to publish granted (date)** 15 February 2018

**Language** English

**Monograph**

**Article dissertation**

**Essay dissertation**

**Abstract**

One significant way to mitigate global warming is to replace energy generation based on fossil fuels with CO<sub>2</sub> emission free energy sources. In electricity generation, these kinds of energy sources are, e.g., hydro, nuclear and variable renewable energy (VRE) sources, such as wind and solar energy. The installed capacities of wind power and photovoltaic panels are growing globally at an increasing pace. However, the increase of the installed variable generation capacity can cause problems for the power systems as it also increases the variability of the power generation. To ensure reliable operation of the power systems and large scale integration of VRE generation, the behaviour of VRE generation has to be modelled and understood both in short and long term.

This thesis focuses on developing statistical modelling methodologies applicable in Monte Carlo simulations for the long term modelling of wind and solar power generation in new non-measured generation locations. Modelling methodologies are developed for the modelling of new wind power plants (WPPs) and new photovoltaic power plants (PVPs) in non-measured locations, and for the joint modelling of both generation types. Furthermore, an approach to model wind direction in new WPPs is developed and a methodology for the modelling of changing installed capacity of wind generation in large geographical areas is proposed.

The proposed methodologies are able to model both temporal and spatial dependency structures in multiple new generation locations and can be applied in the analysis of the variability of the generation and power ramps. The main modelling approach in the proposed methodologies is based on time series models combined with data transformations comparable to copula modelling. The developed methodologies are verified against out-of-sample test data, consisting of measurements which were left out from the model selection and estimation processes.

In addition, the effects of the geographical distribution of the VRE generation locations to the aggregated power generation are studied with WPPs and PVPs. The observed and quantified impact of the negative correlation between WPPs and PVPs is analysed jointly with geographical distributions of the locations.

**Keywords** Copula, Monte Carlo simulation, photovoltaic power generation, renewable energy, solar energy, solar irradiance, statistical modelling, vector autoregressive model, VRE generation, wind power generation, wind speed

**ISBN (printed)** 978-952-60-7881-6

**ISBN (pdf)** 978-952-60-7882-3

**ISSN-L** 1799-4934

**ISSN (printed)** 1799-4934

**ISSN (pdf)** 1799-4942

**Location of publisher** Helsinki

**Location of printing** Helsinki

**Year** 2018

**Pages** 199

**urn** <http://urn.fi/URN:ISBN:978-952-60-7882-3>



**Tekijä**

Jussi Ekström

**Väitöskirjan nimi**

Tilastollisia menetelmiä vaihtelevan uusiutuvan energiantuotannon mallintamiseen

**Julkaisija** Sähkötekniikan korkeakoulu**Yksikkö** Sähkötekniikan ja automaation laitos**Sarja** Aalto University publication series DOCTORAL DISSERTATIONS 42/2018**Tutkimusala** Sähköverkot ja suurjännitetekniikka**Käsikirjoituksen pvm** 19.12.2017**Väitöspäivä** 11.05.2018**Julkaisuluvan myöntämispäivä** 15.02.2018**Kieli** Englanti **Monografia** **Artikkeliväitöskirja** **Esseeväitöskirja****Tiivistelmä**

Yksi merkittävä keino ilmastonmuutoksen hillitsemiseksi on fossiilisiin polttoaineisiin perustuvan energiantuotannon korvaaminen CO<sub>2</sub>-vapailta tuotantomuodoilla. Sähköntuotannossa tällaisia tuotantomuotoja ovat mm. vesivoima, ydinvoima sekä vaihtelevat uusiutuvat energianlähteet kuten tuuli- ja aurinkoenergia. Tuuli- ja aurinkoenergian tuotannon asennetun kapasiteetin määrä kasvaa globaalisti nopealla tahdilla. Kasvava vaihtelevan energiantuotannon määrä voi kuitenkin aiheuttaa ongelmia voimajärjestelmille, sillä se lisää myös tehotuotannon vaihteluita. Jotta voimajärjestelmän luotettava toiminta voidaan taata ja uusiutuvan vaihtelevan energiantuotannon laajamittainen käyttöönotto mahdollistaa, tulee vaihtelevan uusiutuvan tuotannon käyttäytymistä pystyä ymmärtämään ja mallintamaan tarkasti niin lyhyellä kuin pitkällä aikavälillä.

Tämä työ keskittyy kehittämään Monte Carlo simulaatioihin soveltuvia tilastollisia mallinusmenetelmiä tuuli- ja aurinkoenergian tuotannon pitkän aikavälin mallinnukseen uusissa tuotantokohteissa, joista ei ole saatavilla mittausdataa. Mallinusmenetelmiä kehitetään uusille tuulipuistoille, uusille aurinkoenergian tuotantolaitoksille (sisältäen myös kuluttajien PV-paneelit) sekä näiden kahden tuotantotyyppin samanaikaiseen yhteismallinnukseen. Lisäksi työssä kehitetään menetelmä tuulen suunnan tilastolliseen mallinnukseen uusissa tuulipuistoissa sekä menetelmä muuttuvan tuulituotantokapasiteetin mallinnukseen suuremmilla maantieteellisillä alueilla.

Kehitetyt menetelmät mahdollistavat niin ajallisten kuin maantieteellisten riippuvuusrakenteiden tarkan mallinnuksen useiden tuotantolaitosten välillä ja niitä voidaan hyödyntää tuotannon vaihteluiden ja nopeiden tehomuutosten mallinnuksessa. Valitussa mallinnusstrategiassa yhdistetään tilastolliset aikasarjamallit ja kopula-mallinnuksen kaltaiset datamuunnokset. Kehitettyjen menetelmien toimivuus on vahvistettu vertaamalla mallien tuottamia simulaatiotuloksia testaukseen varattuun vertailudataan.

Lisäksi tutkitaan tuotantolaitosten maantieteellisen hajautuksen vaikutuksia tuuli- ja aurinkoenergian aggregoituun tuotantoon ja tuotannon vaihteluihin. Tuuli- ja aurinkoenergiantuotannon havaitaan olevan negatiivisesti korreloituneita yhden tunnin aikaresoluutiolla ja tämän negatiivisen korrelaation vaikutuksia tutkitaan myös yhdessä tuotantolaitosten maantieteellisen hajautuksen kanssa.

**Avainsanat** Auringon irradianssi, aurinkoenergia, aurinkosähkö, kopula, Monte Carlo simulaatiot, tilastollinen mallinnus, tuulen nopeus, tuulivoima, usean muuttujan autoregressiivinen malli, uusiutuva energia

**ISBN (painettu)** 978-952-60-7881-6**ISBN (pdf)** 978-952-60-7882-3**ISSN-L** 1799-4934**ISSN (painettu)** 1799-4934**ISSN (pdf)** 1799-4942**Julkaisupaikka** Helsinki**Painopaikka** Helsinki**Vuosi** 2018**Sivumäärä** 199**urn** <http://urn.fi/URN:ISBN:978-952-60-7882-3>



# Preface

The research work was carried out in the Research Group of Power Systems and High Voltage Engineering at Aalto University School of Electrical Engineering, Department of Electrical Engineering and Automation, between April 2014 and February 2018. The work was supported by Aalto University as a part of the Sustainable Transition of European Energy Markets (STEEM) project and through a funded position in Aalto ELEC Doctoral School for three and half years. The additional financial support from the Emil Aaltonen Foundation, KAUTE Foundation, Walter Ahlström Foundation and Kansallis-Osake Pankki Fund is also gratefully acknowledged.

First, I would like to thank my supervisor Prof. Matti Lehtonen for providing me the opportunity to do research under his guidance, and for his help and supervision. The pre-examination of my thesis was performed by Prof. Mikko Kolehmainen from University of Eastern Finland and Prof. Rimantas Deksnys from Kaunas University of Technology whom I would also like to thank for their constructive feedback and suggestions.

In addition, I would like to express my gratitude for my advisors University Teacher Emeritus Ilkka Mellin and Dr. Matti Koivisto for their invaluable contribution to my research work and for our fruitful co-operation, which has yielded many joint publications. Without your input this thesis wouldn't have been possible.

I'm also grateful for my colleagues and co-authors Prof. Liisa Haarla, Dr. Antti Alahäivälä, Dr. Mubbashir Ali, Dr. Eero Saarijärvi, Dr. Janne Seppänne, Toni Tukia, Semen Uimonen, and Dr. Antti Lehikoinen. Thank you all! Furthermore, special thanks to Dr. John Millar for both proofreading my writings and for all the great philosophical discussions we had.

During the nine and half years I have spent in Aalto University and Otaniemi from the beginning of my Bachelor's studies, through my Master's degree all the way to the finalization of my doctoral dissertation, I have had the greatest privilege to meet so many awesome individuals and form life long friendships. There are too many of you to mention here by name, but thank you all for these unforgettable years! In addition, I would like to express my gratitude for Aalto University's Guild of Electrical Engineering for everything, and for the

extraordinary dwellers of the depths of JMT 7, you have brought invaluable camaraderie and laughter into my days sometimes a bit too filled with quite tedious research work.

Last but also most importantly, I would like to express my endless gratitude for my parents. Thank you for all your encouragement and support! And thank you Sanni for your love and for having the patience to listen my endless ramblings about all the problems related to my research and this thesis.

Espoo, February 26, 2018,

Jussi Ekström

# Contents

<b>Preface</b>	<b>1</b>
<b>List of Publications</b>	<b>5</b>
<b>Author's Contribution</b>	<b>7</b>
<b>List of Symbols and Abbreviations</b>	<b>8</b>
<b>1. Introduction</b>	<b>13</b>
1.1 Background . . . . .	13
1.2 Motivation, Research Questions and Scope of the Thesis . . . . .	14
1.3 Scientific Contribution . . . . .	16
1.4 Structure of the Thesis . . . . .	18
<b>2. Review of Relevant Research</b>	<b>19</b>
2.1 Long Term Wind Power Modelling . . . . .	19
2.2 Long Term Solar and Combined Wind and Solar Power Modelling	22
<b>3. Modelling Methods</b>	<b>25</b>
3.1 Copula Theory . . . . .	25
3.2 Modelling of Temporal and Spatial Dependency Structures . . . . .	26
3.2.1 Univariate Autoregressive Approach . . . . .	27
3.2.2 Simplified Vector Autoregressive Approach . . . . .	28
3.2.3 Full Vector Autoregressive Approach . . . . .	30
3.2.4 Error Term Modelling . . . . .	32
3.3 Modelling of New Wind Power Plant Locations . . . . .	33
3.3.1 Estimation Data and Marginal Distributions . . . . .	33
3.3.2 Model Selection and Parameter Estimation . . . . .	35
3.3.3 Estimation of the Temporal and Spatial Correlations	35
3.3.4 Wind Turbine Power Generation Models . . . . .	36
3.3.5 Simulation Procedure . . . . .	37
3.4 Modelling of New Solar Power Plant Locations . . . . .	39
3.4.1 Estimation Data and Marginal Distributions . . . . .	39
3.4.2 Model Selection and Parameter Estimation . . . . .	41
3.4.3 Estimation of Spatial Correlations . . . . .	42
3.4.4 Photovoltaic System Power Generation Model . . . . .	43
3.4.5 Simulation Procedure . . . . .	44
3.5 Combined Modelling of New Wind and Solar Power Plant Locations . . . . .	44
3.5.1 Marginal Distributions . . . . .	44
3.5.2 Model Selection and Parameter Estimation . . . . .	46

3.5.3	Estimation of Spatial Correlations . . . . .	47
3.5.4	Simulation Procedure . . . . .	47
3.6	Wind Direction Modelling . . . . .	48
3.6.1	Estimation and Simulation Procedure . . . . .	48
3.6.2	Dynamic Wake Effect Model . . . . .	50
3.7	Regional Wind Generation Modelling . . . . .	51
<b>4.</b>	<b>Simulation Results</b>	<b>53</b>
4.1	New Wind Power Plant Locations . . . . .	53
4.2	New Solar Power Plant Locations . . . . .	58
4.3	Combined New Wind and Solar Power Plant Locations . . . . .	59
4.4	Effect of Geographical Distribution of Generation Locations . . . . .	61
4.4.1	Wind Power Plants . . . . .	62
4.4.2	Photovoltaic Power Plants . . . . .	65
4.4.3	Wind and Photovoltaic Power Plants . . . . .	66
<b>5.</b>	<b>Discussion and Conclusions</b>	<b>69</b>
5.1	Discussion of the Modelling Methods and Simulation Results . . . . .	69
5.1.1	Summary of the Findings . . . . .	69
5.1.2	Practical Applications of the Work . . . . .	72
5.2	Recommendations for Future Work . . . . .	73
5.3	Conclusions . . . . .	74
	<b>References</b>	<b>75</b>
	<b>Publications</b>	<b>83</b>

# List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

- I** Matti Koivisto, Jussi Ekström, Ilkka Mellin, Janne Seppänen, Eero Saarijärvi, Liisa Haarla. Statistical analysis of large scale wind power generation using Monte Carlo simulations. In *Power Systems Computation Conference (PSCC)*, Wroclaw, Poland, pages 1-7, August 2014.
- II** Jussi Ekström, Matti Koivisto, Ilkka Mellin, Robert John Millar, Eero Saarijärvi, Liisa Haarla. Assessment of large scale wind power generation with new generation locations without measurement data. *Renewable Energy*, Volume 83, pages 362-374, November 2015.
- III** Jussi Ekström, Matti Koivisto, Ilkka Mellin, Robert John Millar, Matti Lehtonen. A Vector Autoregressive Based Methodology for Wind Generation and Power Ramp Analysis in New Locations. Submitted to a *peer reviewed journal*, August 2017.
- IV** Jussi Ekström, Matti Koivisto, Robert John Millar, Ilkka Mellin, Matti Lehtonen. A Statistical Approach for Hourly Photovoltaic Power Generation Modeling with Generation Locations without Measured Data. *Solar Energy*, Volume 132, pages 173-187, July 2016.
- V** Jussi Ekström, Matti Koivisto, Ilkka Mellin, Robert John Millar, Matti Lehtonen. A statistical model for hourly large-scale wind and photovoltaic generation in new locations. *IEEE Transactions on Sustainable Energy*, Volume 8, Issue 4, pages 1383-1393, October 2017.
- VI** Matti Koivisto, Jussi Ekström, Ilkka Mellin, Robert John Millar, Matti Lehtonen. Statistical wind direction modeling for the analysis of large scale

wind power generation. *Wind Energy*, Volume 20, pages 677-694, 2017.

**VII** Matti Koivisto, Jussi Ekström, Janne Seppänen, Ilkka Mellin, Robert John Millar, Liisa Haarla. A Statistical Model for Comparing Future Wind Power Scenarios with Varying Geographical Distribution of Installed Generation Capacity. *Wind Energy*, Volume 19, pages 665-679, 2016.

# Author's Contribution

## **Publication I: “Statistical analysis of large scale wind power generation using Monte Carlo simulations”**

The main idea and the methodology was developed by Matti Koivisto. Jussi Ekström and Matti Koivisto performed the simulations and analyzed the results in co-operation. Matti Koivisto wrote the paper. Eero Saarijärvi assisted with the geographical modeling. Ilkka Mellin, Janne Seppänen and Liisa Haarla contributed through comments and discussion.

## **Publication II: “Assessment of large scale wind power generation with new generation locations without measurement data”**

The main idea and the methodology was developed by Jussi Ekström and Matti Koivisto. Jussi Ekström performed the simulations, analyzed the results and wrote the paper. Eero Saarijärvi assisted with the geographical modeling. Ilkka Mellin, John Millar and Liisa Haarla contributed through comments and discussion.

## **Publication III: “A Vector Autoregressive Based Methodology for Wind Generation and Power Ramp Analysis in New Locations”**

The main idea was developed by Jussi Ekström and Matti Koivisto. Jussi Ekström developed the methodology, performed the simulations, analyzed the results and wrote the paper. Ilkka Mellin, John Millar and Matti Lehtonen contributed through comments and discussion.

**Publication IV: “A Statistical Approach for Hourly Photovoltaic Power Generation Modeling with Generation Locations without Measured Data”**

The main idea was developed by Jussi Ekström and Matti Koivisto. Jussi Ekström developed the methodology, performed the simulations, analyzed the results and wrote the paper. Ilkka Mellin, John Millar and Matti Lehtonen contributed through comments and discussion.

**Publication V: “A statistical model for hourly large-scale wind and photovoltaic generation in new locations”**

The main idea was developed by Jussi Ekström and Matti Koivisto. Jussi Ekström developed the methodology, performed the simulations, analyzed the results and wrote the paper. Ilkka Mellin, John Millar and Matti Lehtonen contributed through comments and discussion.

**Publication VI: “Statistical wind direction modeling for the analysis of large scale wind power generation”**

The main idea was developed by Jussi Ekström and Matti Koivisto. Jussi Ekström and Matti Koivisto developed the methodology, performed the simulations and analyzed the results in co-operation. Matti Koivisto wrote the paper. Ilkka Mellin, John Millar and Matti Lehtonen contributed through comments and discussion.

**Publication VII: “A Statistical Model for Comparing Future Wind Power Scenarios with Varying Geographical Distribution of Installed Generation Capacity”**

The main idea was developed by Jussi Ekström and Matti Koivisto. Jussi Ekström and Matti Koivisto performed the simulations and analyzed the results in co-operation. Matti Koivisto developed the methodology and wrote the paper. Janne Seppänen, Ilkka Mellin, John Millar and Liisa Haarla contributed through comments and discussion.

# List of Symbols and Abbreviations

## Abbreviations

ACF	Autocorrelation function
AR	Autoregressive
ARC	Univariate AR models combined via spatial correlation matrix $\mathbf{C}$
ARMA	Autoregressive-moving-average
BRL	Boland-Ridley-Laurent
CDF	Cumulative distribution function
Cov	Covariance
CSI	Clear-sky index
DSO	Distribution system operator
ECDF	Empirical cumulative distribution function
EWEA	European Wind Energy Association
FMI	Finnish Meteorological Institute
i.i.d.	Independent and identically distributed
MC	Monte Carlo
MERRA	Modern Era Retrospective-analysis for Research and Applications dataset
mod	Modulo operator
NWP	Numerical weather prediction
OLS	Ordinary least squares

PACF	Partial autocorrelation function
PDF	Probability density function
PIC	Proportion of installed capacity
PV	Photovoltaic
PVP	Photovoltaic power plant
TSO	Transmission system operator
VAR	Vector autoregressive
VRE	Variable renewable energy
WPP	Wind power plant
XCF	Cross-correlation function

## Symbols

$\mathbf{A}$	State-space form of vector autoregressive coefficient matrix
$\mathbf{A}_h$	Vector autoregressive coefficient matrix at lag $h$
$a_h$	Autoregressive coefficient at lag $h$
$\mathbf{C}_u$	Correlation matrix of the error term $\mathbf{u}$
$\mathbf{C}$	Correlation matrix
$C$	Copula
$\mathbf{c}$	Vector of intercept terms
$c$	Intercept term
$cap_{i,t}$	Installed generation capacity in area $i$ at time $t$
$\mathbf{D}$	Diagonal matrix with standard deviations on the diagonal
$\tilde{\mathbf{E}}_t$	Simulated global irradiance data vector at time $t$
$\mathbf{E}_t^{CS}$	Clear-sky irradiance at time $t$
$\mathbf{E}_t$	Global irradiance at time $t$
$F_i$	Cumulative distribution function of a random variable (usually margin) $i$
$F_N$	Cumulative distribution function of standard normal distribution

$\mathbf{H}$	Cumulative distribution function of a multivariate distribution
$h$	Lag in hours
$\mathbf{I}_k$	Identity matrix of size $k$
$\tilde{\mathbf{K}}_t$	Simulated clear-sky index data vector at time $t$
$\mathbf{K}_t$	Clear-sky index at time $t$
$k$	Dimension of a vector, matrix or model (usually denotes the number of locations)
$\mathbf{L}$	Lower triangular matrix
$P_r$	Rated power
$P_{i,t}$	Average power generation data in $i$ at time $t$
$p$	Order of autoregressive or vector autoregressive model
$p_{i,t}^{pic}$	Proportion of installed capacity value in area $i$ at time $t$
$\mathbf{R}_x(h)$	Autocorrelation matrix of multivariate time series $\mathbf{x}_t$ at lag $h$
$s$	Value of gradient in inflection point
$T$	Number of observations in time series
$t$	Time index in hours
$\mathbf{u}_t$	Error terms of a model at time $t$
$\mathbf{x}_t$	Value of time series $\mathbf{x}$ at time $t$
$\mathbf{Y}$	Random variable $\mathbf{Y}$
$\tilde{\mathbf{y}}_t$	Simulated wind speed data vector at time $t$
$\mathbf{y}_t$	Wind speed at time $t$
$y_{ip}$	Inflection point
$\tilde{\mathbf{z}}_t$	Simulated transformed data vector at time $t$
$\mathbf{z}_t$	Estimation transformed data vector at time $t$
$\tilde{\mathbf{z}}_t^{\text{wd}}$	Simulated transformed data vector with day structures included in the data at time $t$
$\mathbf{z}_t^{\text{wd}}$	Transformed data vector with day structures included in the data at time $t$

$\Gamma_X(0)$	State-space form of autocovariance matrix of process $\mathbf{x}_t$
$\Gamma_x(h)$	Autocovariance matrix of multivariate time series $\mathbf{x}_t$ at lag $h$
$\boldsymbol{\mu}$	Mean vector
$\rho$	Pearson's product-moment correlation coefficient
$\rho_T$	Circular correlation coefficient
$\boldsymbol{\Sigma}$	Covariance matrix
$\boldsymbol{\Sigma}_U$	State-space form of covariance matrix of error term $\mathbf{u}$
$\boldsymbol{\Sigma}_u$	Covariance matrix of error term $\mathbf{u}$
$\sigma$	Standard deviation
$\sigma^2$	Variance
$\omega_t$	Wind direction at time $t$
$\tilde{\omega}_t$	Simulated wind direction data vector at time $t$

# 1. Introduction

## 1.1 Background

Global warming is arguably one of the greatest of today's challenges and green house gases, such as CO<sub>2</sub> emissions, are one of its main causes. The most notable human activities causing CO<sub>2</sub> emissions are burning of fossil fuels and deforestation. From the fossil fuels, which account for two thirds of the CO<sub>2</sub> emissions [1], coal accounted for 43% in 2011, and is the most used in heat and electricity generation [2]. Thus, one prominent way to mitigate global warming is to replace coal based energy generation with cleaner energy sources. In electricity generation the most notable of such generation types are nuclear power plants and renewable energy such as hydro, wind and solar energy. Of these, wind and solar energy are variable renewables, which are in the focus of this work.

The installed wind and solar generation capacities have been increasing rapidly during the recent years. At the end of 2016 the total installed wind generation capacity was globally 487 GW [3]. In 2016, the installed wind capacity was 150 GW in the EU area and approximately 12.5 GW of new wind generation was installed [4]. The installed photovoltaic (PV) capacity by the end of 2016 was globally at least 303 GW and approximately 75 GW of new PV generation was installed in 2016 [5].

The rapid growth of the installed variable renewable energy (VRE) generation has been enabled through several factors including technological development and decreased costs of wind and solar power generation technologies. In addition, major contributing factors have been the policy driven renewable energy targets and related incentives, such as subsidies for renewable generation. Other supporting factors have been, e.g., the global trend of decentralization of energy generation, especially affecting the popularity of consumer PV systems.

Wind and solar power generation is stochastic by nature, and thus, the generation varies in time depending on the wind conditions and cloudiness. The increased variations in power generation, caused by the growing penetration

of VRE, require also increased flexibility for the system to maintain the power balance, i.e., the balance between generation and consumption of energy, and to ensure the reliable operation of the power system [6]. Also, the increasing variations in generation, and especially times of excess generation, can cause notable wind and solar curtailment, and thus, wasting of generated energy [7]. Consequently, the analysis and modelling of VRE generation and prediction of the power generation in short and long term have been topics with increasing interest and importance in recent years.

The analysis and modelling of VRE generation can be roughly divided into short term forecasts and medium to long term simulations. Short term forecasts aim to predict the output of wind power plants (WPPs) and photovoltaic power plants (PVPs) for the following hours, and thus, are critical for maintaining the power balance through regulation power reserves and balancing markets.

Medium to long term simulations focus on the modelling of the effects of the increasing VRE generation capacity on power and energy systems. The long term simulation studies enable the analysis of the effects of new installed or planned capacity on energy generation and its volatility in a certain geographical area. This kind of knowledge is crucial for the planning of future power and energy systems, and thus, alongside the accurate forecasting methods, to enable the large scale integration of VRE generation.

## 1.2 Motivation, Research Questions and Scope of the Thesis

As mentioned, the mitigation of CO<sub>2</sub> emissions is paramount in order to slow down the global warming. In electricity generation, one of the key approaches to reduce CO<sub>2</sub> emissions is to replace coal with large scale deployment of VRE generation. However, this increases the overall variability of the power generation, which can cause problems for the operation of the system.

The installed capacities of VRE generation have been growing fast and the trend can be expected to continue also in the future. However, a significant problem is how to model the planned large scale VRE generation and its effects. The aim of this thesis is to answer that particular question. Medium to long term modelling methods have been studied extensively in existing WPPs and PVPs with measurement data available, however, easily applicable simulation models for new generation locations without any measurement data, i.e., new non-measured locations, are not widely available.

Some meteorological reanalysis datasets are available for such purposes [8], but these datasets, or the reanalysis models, do not enable Monte Carlo (MC) simulations, as they combine historical measurements to numerical weather prediction models [9]. In addition, the reanalysis models underestimate the variability of wind speeds in time resolutions less than few hours [10]. The applicability of the models for MC simulation studies, and time resolutions less than few hours in the simulations, are becoming increasingly important when

wind and solar generation are modelled together and further combined with the modelling of electricity consumption. Thus, this work develops accurate modelling methods applicable for MC simulations for versatile analysis of new planned WPPs and PVPs to quantify their effect on the system. This is to optimize and enable the integration of large scale VRE generation to power systems. Reflecting this goal, the main research question this work aims to answer is

- How future VRE generation and its variability can be modelled in new non-measured generation locations to ease the large scale integration of VRE generation to the power system?

The main research question can be further divided into the following more detailed research questions

- How to model the power generation and variability of non-measured WPPs and PVPs separately and jointly in detail, including the modelling of wind speed, wind direction and solar irradiance?
- How the geographical distribution of large scale VRE generation locations affect the aggregated power generation and its variability and the need for balancing power?

To narrow down the vast topic of VRE modelling, the main focus is on the detailed modelling of new WPPs and PVPs. Autoregressive (AR) time series models are chosen for the main statistical modelling approach in this work. Markov chain based approaches [11] or other common machine learning methods such as artificial neural networks, support vector machines and different tree based approaches are not considered. Such methods are widely used in various regression and classification problems in many domains [12], however most of these approaches are not designed for time series modelling, which is in the main focus in this thesis. AR models, on the other hand, are specifically designed for time series analysis. From the most common machine learning approaches recursive neural networks are also applicable for time series modelling [13], however, in this work approaches that have good interpretability are favoured over block box modelling, which is usually the case with neural networks [12]. Thus, as approaches based on AR models have good interpretability, are more straightforward compared with recursive neural networks and, most importantly, have good performance in the studied modelling tasks, they are chosen for the main modelling approach.

Application wise, a particular interest is placed on the effects of the geographical distribution of the VRE generation locations on the aggregated power generation and its volatility. Furthermore, the overall focus is on the long term modelling and simulation of VRE generation, and thus, short term simulations

and forecasting are left out from this work, even though the developed models can also be applied in such modelling.

The main research methods applied in this thesis are data analysis and mathematical modelling based on data. First, measurement data (e.g. wind speeds or solar irradiances) is analysed and pre-processed. This work utilizes hourly data, however, the proposed methodologies are also applicable for sub-hourly time resolutions, if sufficient data is available. Then, appropriate statistical models are selected and full modelling methodologies are developed. The statistical models are estimated using the measurement data and the modelling methodologies verified against measurements reserved for testing using MC simulations. Finally, the proposed methods are utilized in the simulation of cases and future scenarios and the obtained results are further interpreted and analysed.

Last, in this work WPPs refer to large scale wind power generation in wind farms. PVPs can be considered either as large scale PV power plants, i.e., solar farms or as aggregates of consumer PV systems on the rooftops in a suitable geographical areas. Furthermore, even though the abbreviation PVP is used to address the solar power plants, the developed modelling methods are also applicable for the modelling of other types of solar generation in addition to PV such as concentrated solar power.

### 1.3 Scientific Contribution

The main novel scientific contributions of the seven publications included in this thesis are summarized in this section.

- Two statistical methodologies for the modelling of new non-measured WPPs are developed in Publications I–III. The methodologies are developed further in each consecutive publication from the transformed ARC model (Univariate AR models combined via spatial correlation matrix  $\mathbf{C}$ ) to a complete modular modelling methodology based on transformed full VAR model with  $t$ -distributed errors. In Publication III, the full VAR based methodology and an approach to estimate the model parameters for new WPP locations is developed and the applicability of the model for aggregated power ramp modelling is demonstrated. All methodologies are verified against out-of-sample data with good accuracy and used to model new scenarios using MC simulations. Modelling methodologies are presented in Section 3.3 and simulation results in Sections 4.1 and 4.4.1.
- A modular statistical modelling methodology for new non-measured PVPs (i.e. centralized PV plants or aggregated consumer PV systems) is developed in Publication IV. The methodology is based on a transformed time-varying ARC model with  $t$ -distributed errors, which is verified against out-of-sample measurements with good accuracy. Furthermore, the methodology is used

to model new PVPs in future scenarios using MC simulations. Modelling methodology is presented in Section 3.4 and simulation results in Sections 4.2 and 4.4.2.

- A joint modelling methodology for a system with both WPPs and PVPs is developed in Publication V. The developed modelling methodology is based on a transformed simplified time-varying VAR model with  $t$ -distributed errors, of which estimation process without model order limitations for new generation locations is demonstrated. The methodology is verified against out-of-sample data with good accuracy and the need for joint modelling of both WPPs and PVPs is also justified. In addition, a notable negative correlation between WPPs and PVPs is observed. This observation implies that both generation types should be included in the power system to decrease the volatility of the generation, which is also assessed in case studies through MC simulations. Modelling methodology is presented in Section 3.5 and simulation results in Sections 4.3 and 4.4.3.
- A statistical methodology for the modelling of wind direction in new non-measured locations is developed in Publication VI. A simplified VAR model combined with an approach to estimate the model from circular wind direction data is developed and the accuracy of the model is verified against out-of-sample measurements. Wind direction modelling using MC simulations allows more accurate wake effect modelling in new WPP locations as demonstrated in case studies. Modelling methodology is presented in Section 3.6 and simulation results in Section 4.4.1.
- A transformed VAR model with  $t$ -distributed errors is developed for statistical modelling of aggregated wind generation in large geographical areas in Publication VII. It is demonstrated that the model is capable of modelling new scenarios with changing installed wind power capacities in the areas. An approach to minimize the variance of the aggregated wind generation is also proposed. Modelling methodology is presented in Section 3.7 and simulation results in Section 4.4.1.
- The effect of the geographical distribution of the VRE generation locations is analysed in Publications I–VII using long term MC simulation studies. It is assessed how much dispersed geographical distribution decreases the probabilities of very low or high generation events and the volatility of the aggregated generation compared with geographically concentrated distribution. The effect of the geographical distribution is also quantified for aggregated wind power ramps. The effect of the geographical distribution to the aggregated generation is found to be notably smaller with PVPs compared with WPPs in hourly time resolution. Furthermore, the impact of the negative correlation between WPPs and PVPs is analysed jointly with different geographical distributions and

optimal proportional annual generated energies (and thus also proportional installed capacities) for WPPs and PVPs are presented. The simulation results related to the geographical distribution of the generation locations are presented in Section 4.4.

## **1.4 Structure of the Thesis**

The thesis is organized as follows. Chapter 2 provides a review of the prior work on wind and solar power generation modelling with an emphasis on long term simulations. Chapter 3 presents the statistical modelling methods developed in this thesis for the modelling of wind and solar power generation. Chapter 4 presents the main simulation results obtained with the developed models. Finally, Chapter 5 concludes the thesis by giving a discussion and recommendations for future work on the developed models and obtained simulation results.

## 2. Review of Relevant Research

This chapter presents a review of relevant prior research related to the medium to long term modelling of wind speeds, solar irradiances, WPPs and PVPs. First, research related to the modelling of wind speeds and WPPs is discussed, and after that the modelling approaches for PVPs, and also shortly for the joint modelling of both VRE types, is presented.

### 2.1 Long Term Wind Power Modelling

When analysing wind power generation and WPPs in long term (usually a time span of months or years), there are two viable modelling approaches to choose from. These are either statistical or physical models such as meteorological reanalysis models. The reanalysis models combine years of wind speed (and many kinds of other weather data such as solar irradiance and temperatures) measurements to numerical weather prediction (NWP) models to provide historical datasets for wide range of geographical areas [9, 14–16]. These models are extremely computational heavy, but some datasets such as MERRA reanalysis dataset [8], are freely available. The reanalysis models, however, are not applicable for simulation methods such as Monte Carlo [17]. In addition, reanalysis models tend to underestimate extreme wind speeds and variability when time resolution is less than few hours [10], and thus, additional stochastic modelling of the short-term fluctuations is needed to accompany them as shown in [18].

Statistical models, on the other hand, are applicable for MC simulation studies, which is a notable advantage in the analysis of future scenarios. In addition, statistical models can be estimated also for higher time resolutions from, e.g., one hour to 15 or 5 minutes. Wind power generation modelling utilizing statistical models in long term MC simulations in measured wind speed or WPP locations have been considered extensively in [19–25].

The use of copula modelling (introduced in Section 3.1) enables the separation of the spatial and temporal dependency structures of WPPs from the modelling of marginal probability distributions (margins) of the individual WPP locations using data transformations [11, 19–21, 26]. When using the statistical modelling

approach with multiple generation locations, copula modelling, or data transformations motivated by copula modelling, are commonly used together with a statistical model. In addition, the copula approach can be also used in MC simulations as in [11, 19–21]. The copula approach allows the separation of the modelling in two parts, which are notably easier to model separately; the dependency structures (modelled with, e.g., time series models) and the margins (distributions modelled separately in each location). As this is an extremely useful strategy to simplify the modelling, it is also implemented in this work in Publications I–VII.

Time series models, such as AR or autoregressive-moving-average (ARMA) can be used to model the temporal dependencies in individual locations [21, 25, 27–29]. However, such univariate approaches are unable to analyse the spatial correlations between time series (or WPP locations) as such. However, there are mathematical methods to link the univariate processes to each other. In [30], the spatial correlations were modelled by combining univariate AR models with Cholesky decomposition of the spatial correlation matrix. This approach was also used and extended in this work and in Publications I, II and IV.

However, a spatial correlation matrix is not enough to capture the full complex spatial dependency structure between wind speeds or WPPs as presented later in Section 3.2.3 and illustrated in Section 4.1. According to [31], the dependency structure modelled with accurate cross-correlation functions, and consequently, proper spatial correlations between wind speed measurement locations are crucial. Thus, more detailed models, are required.

A popular approach to model both spatial and temporal dependency structures together is to use a vector autoregressive (VAR) model, which is a multivariate extension of the AR model, and also easily applicable for MC simulations [32, 33]. VAR model has been used to model dependency structures in existing WPP locations in [23, 29, 34, 35]. Other multivariate time series models such as multivariate ARMA models are used for wind generation forecasting error simulations in [36]. In addition to VAR and other time series models, also artificial neural networks [22, 23, 31] or Markov Chains [11, 37] have been used in the modelling of the dependency structures in wind power generation. In this work, the widely utilized approach based on time series models, such as various forms of AR and VAR based models, is also applied in Publications I–VII.

As mentioned earlier, data transformations are commonly used together with time series models such as the VAR model to separate the margins of individual WPPs from the dependency structures between and in the WPPs. The probability integral transformations do not fully preserve Pearson's correlations [38, 39], however rank order correlations such as Spearman's rho are preserved [40]. Though, the difference (the change in the Pearson's correlation values) over the transformations are minor [41]. Hence, both measures of correlation are used in the publications related to this thesis depending on the case.

The marginal probability distributions, i.e., the wind speed or wind power generation margins in individual WPP locations can be modelled with a wide

array of different approaches. The most common parametric distribution used to describe wind speed distribution in an arbitrary location is Weibull distribution [42, 43]. Non-parametric distributions or estimates, such as Gaussian kernels [44, 45] or empirical cumulative distribution functions (ECDFs) [20] are also widely used. These distributions can be combined with Generalized pareto distribution, which can be used to model tails of the distribution to enable estimates for values larger than the maximum value observed in the estimation data [32, 33]. In this work, all of the above mentioned approaches are used to describe wind speed distributions depending on the application.

The monthly diurnal structures, i.e., the hourly intra-day and monthly structures have been modelled using either dummy variables with the time series models, as shown in [46] or removing them from the data before estimation and adding back in simulation phase as done in [34]. Both approaches have been applied in this work.

The transformation from wind speeds to power can be done with various different power generation models. One of the most common ways is to use a third degree power curve as done in [42], although this approach is simplified and inaccurate for some wind turbines. More advanced state-of-the-art approaches include the utilization of logistic functions as power curves as done in [47, 48], however the accuracy of these approaches is heavily dependant on the correctness of the used parameters. Data for such approaches can be obtained from turbine data sheets and other important parameters, e.g., average unavailability rates from freely available reports such as [49]. As measured data is not required, the above mentioned approaches are suitable for the modelling of new WPP locations, which is one of the main focus areas in this thesis.

The wake effects inside WPPs can be modelled with a wide range of approaches from a single coefficient describing the whole wake effect in a WPP to detailed dynamic models [50, 51]. However, the utilization of the detailed wake effect models can be problematic with new WPPs as the real topology of the wind farm is not necessarily known. When modelling wake effect in a WPP in detail, the modelling of wind direction is also crucial, as it determines how the turbines of a wind farm align in respect to each other [52].

Wind direction has been modelled with time series models, e.g., ARMA model, in short-term forecasting in existing locations in [53]. Direction modelling using Markov Chains and dividing the directions into eight sectors have been considered in [54]. In this thesis and in Publication VI, the time series modelling approach is taken with the wind direction modelling, as it does not require simplifications such as using sectors as in [54].

To summarize, various approaches are available for the long term modelling of wing power generation time series. Data transformations combined with time series models has been a widely utilized approach and also used in this work. However, even though published research on the modelling of WPPs is plenty, the modelling of new non-measured locations is scarce. The main approach for the modelling of WPPs without actual data from the locations is to use the

meteorological reanalysis datasets. However, as mentioned, this approach does not allow MC simulations. In this work, this problem is addressed by developing MC simulation based modelling methodologies for long term wind generation analyses.

## 2.2 Long Term Solar and Combined Wind and Solar Power Modelling

In solar power generation modelling, whether it be the modelling of power or solar irradiance, similar approaches as with wind power can be used. Again, the two most viable options for long term simulation studies are the reanalysis datasets (which usually contain also irradiance data) created using historical data and physical NWP models [9,14–16], e.g., the previously mentioned MERRA dataset [8], which is used with solar modelling in [55], or the approaches based on plethora of different statistical models. However, as also mentioned earlier, only the statistical approaches enable the implementation of MC simulations, which are beneficial when conducting long term case studies or scenario analyses in large systems.

Copula modelling, or data transformations motivated by it, can be applied similarly for PVP and solar irradiance modelling as with wind speeds and WPPs in earlier section. The benefit of copula method is exactly the same, i.e., to separate the modelling of the spatial and temporal dependency structures between PVP locations from the modelling of the margins in individual PVP locations. Copulas are applied in PVP modelling, e.g., in [56].

As with the copula modelling, similar statistical modelling approaches, which are applicable for WPPs, are also viable for the PVP modelling. Statistical modelling of solar irradiance and PVPs have been considered extensively with various different models and approaches [57–59]. Time series modelling approaches are commonly used both in short term forecasting and in long term simulations. Short term forecasting using time series models such as state-space models, ARMA, simplified VAR and VAR models have been considered in [60–63]. As visible, similar modelling approaches are applied in PVP modelling as with WPPs previously. Long term statistical analysis and approaches to generate artificial (simulated) time series with temporal and spatial correlations are developed in [64,65], but only for measured locations.

As the penetration of PV generation grows rapidly, a notable amount of research has been focusing on the large scale generation. Recently, large scale PVP modelling with multiple locations, instead of considering only a single or few locations, in, e.g., a countrywide setting, is studied and assessed in [65–69].

As with WPPs, several different approaches for the long term simulations and modelling of PVPs are available, however, the main focus of the research is on the existing measured locations. Thus, this work aims to develop methods, which can be used to model new PVPs in long term simulation studies using

MC simulations. The chosen approach is similar to the one chosen for WPP modelling, i.e., data transformations combined with time series models. Next, previous research on the modelling of non-measured PVP locations is shortly described.

Clear-sky irradiance, i.e., the theoretical possible solar irradiance, in any geographical location can be estimated using a model developed in [70]. The clear-sky irradiance estimates the potential for PVP generation in clear-sky conditions, i.e., without any clouds or large amount of particles in the atmosphere blocking the irradiance. The clear-sky irradiance model proposed in [70] has a wide applicability as it doesn't require any measured data to produce estimates for the ideal conditions. Thus, it can be easily applied in the modelling of new PVP locations.

For the transformation from solar irradiance to generated power, several models with varying degree of detail are available. This work utilizes a commonly used power generation model for PV polycrystalline silicon panels proposed in [71]. The model developed in [71] requires the separation of solar global irradiance (which is usually the quantity simulated with the models) into its components for more detailed modelling. The separation from global to diffuse and beam irradiance can be done using a relatively simple, but also more inaccurate, models as in [72], or using detailed models, which also require accurately estimated parameters to function well, as proposed in [73, 74]. For the decomposition, this work applies the state-of-the art logistic model proposed in [73].

In addition to the separate modelling of WPPs and PVPs, also joint modelling of both is a topic of growing interest as penetrations of both types of VRE are growing in energy systems. Thus, the combined variability and output of these stochastic energy sources has to be studied. Combined VRE analysis has been considered in detail in [75–78]. These studies include also the analysis of the correlations between WPPs and PVPs. A negative correlation between wind and solar has been observed in [76, 78], which is a topic studied also in this work.

To summarize the chapter, long term simulations of multiple WPPs, PVPs or a joint VRE modelling have been a topic of interest in research during the previous years. For the long term modelling, two viable options are available; the use of reanalysis models (if sufficient computational capacity is available) and freely available reanalysis datasets or MC simulation based approaches using statistical modelling. Statistical models have been a popular approach, though, utilized mainly on the modelling of existing generation locations with measured data available. Consequently, statistical approaches with the ability to simulate new locations are scarce. Hence, the modelling of new non-measured WPP and PVP locations using MC simulations is the application this work focuses on.



## 3. Modelling Methods

This chapter presents the methods developed for the modelling of wind and solar power generation in Publications I–VII, with the main focus on the modelling of WPPs and PVPs in new (non-measured) generation locations. The chapter is divided into two parts with different focuses. The first part introduces shortly the theory related to copula modelling, which is applied in the modelling approaches presented in this thesis. Then, the time series models applied in the modelling of temporal and spatial dependency structures are presented in detail.

In the second part, the focus is on the developed modelling methodologies. It presents the methodologies for the stochastic modelling of new WPPs, PVPs and the combined modelling of both generation types. These methodologies are the main focus of the thesis. This is followed by the modelling approach developed for wind direction modelling. Finally, a methodology for wind generation modelling in a regional scale without the modelling of individual WPPs is shortly presented.

### 3.1 Copula Theory

This section presents shortly the basic theory related to copula modelling. A copula specifies a multivariate probability distribution with uniformly distributed margins that describes the dependence structure between two or more random variables. In this thesis, an approach motivated by copula modelling is used in the modelling methodologies as it allows the separate modelling of the margins and the dependency structures of the data (the utilized approach is *de facto* copula modelling, but does not necessarily fulfil all theoretical definitions in all cases). This is crucial as the modelling of both simultaneously can be very difficult. The methods presented in this thesis utilize data transformations to Gaussian data (comparable to Gaussian copula), which are suitable for wind speed analyses [41], and hence, other copulas are not discussed.

For a  $k$ -dimensional random variable  $\mathbf{Y} = [Y_1, \dots, Y_k]'$  the cumulative distribution function (CDF) can be specified as

$$H(\mathbf{y}) = P[Y_1 \leq y_1, \dots, Y_k \leq y_k], \quad (3.1)$$

where the right hand side of the equation describes the probability that the random variable  $Y_1$  takes a value less than or equal to  $y_1$  et cetera.  $H$  describes the statistical distribution of  $\mathbf{Y}$ , however, it can be difficult to analyse  $H$  directly [41]. Copula theory provides an approach to separate the analysis of  $H$  into two parts, which is demonstrated next. Sklar's theorem states that  $H$  can be specified as

$$H(\mathbf{y}) = C(u_1, \dots, u_k) = C[F_1(y_1), \dots, F_k(y_k)], \quad (3.2)$$

where  $C$  is a copula and  $F_i$  is the CDF of the margin  $i$  [79]. The marginal CDFs  $F_i$  are specified as

$$F_i(y_i) = P[Y_i \leq y_i]. \quad (3.3)$$

In this thesis, margins can be considered as continuous and strictly increasing, and therefore, Sklar's theorem states that  $C$  is unique. Thus,  $H$  can be separated into two parts;  $C$  contains the information on the dependency structures between the components of  $\mathbf{Y}$ , and the CDFs  $F_i$  contain the information on the marginal distributions [79].

The Gaussian copula, considered in this thesis, defines the CDF of  $\mathbf{U}$ , which is the copula  $C$ , as follows

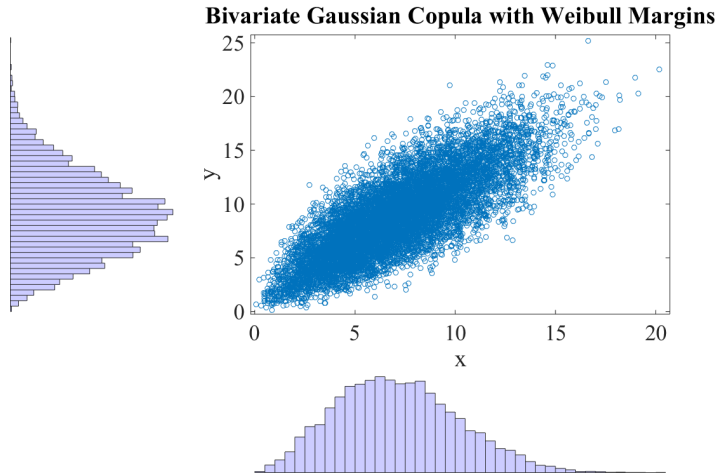
$$C_{\Sigma}(\mathbf{u}) = H_{\Sigma}[F_N^{-1}(u_1), \dots, F_N^{-1}(u_k)], \quad (3.4)$$

where  $F_N^{-1}$  is the inverse CDF of the standard normal distribution and  $H_{\Sigma}$  is the CDF of the  $k$ -dimensional multivariate normal distribution with zero mean vector and covariance matrix  $\Sigma$ , which is equal to the correlation matrix  $\mathbf{C}_{\text{Gauss}}$ , and  $F_N^{-1}$  is the inverse CDF of the standard normal distribution [41, 79]. Figure 3.1 illustrates an example of a dependency structure specified by a bivariate Gaussian copula, with correlation between the two random variables in the Gaussian domain specified as  $\rho = 0.8$ , with Weibull distributed margins.

## 3.2 Modelling of Temporal and Spatial Dependency Structures

This section presents the time series models and related parameter specification methods utilized in this thesis. The presented models can be applied in the modelling of the spatial, temporal and spatio-temporal dependency structures of both wind speeds and solar irradiances. The most notable difference between the modelling of wind speeds and solar irradiances is the data pre-processing phase and the different scale of the hourly and daily variations. The larger variations in solar irradiance compared with wind speeds, however, do not require completely different modelling approaches, but can be addressed with some additions to the models presented in this section as discussed later in Section 3.4.

First, the univariate AR modelling approach is presented, and is followed by the simplified and full VAR approaches. Last, this section presents the error



**Figure 3.1.** Dependency structure specified by a bivariate Gaussian copula ( $\rho = 0.8$ ) with Weibull distributed margins.

term modelling for the above mentioned models.

### 3.2.1 Univariate Autoregressive Approach

In this work, time series models, such as AR model, are used to model the temporal dependency structures in WPPs and PVPs, e.g., how wind speed or solar irradiance values vary in time and what are the relations between the values in consecutive time steps. For example, AR model describes how wind speed at time  $t$  in location  $i$  depends on the previous values of wind speeds in location  $i$  at, e.g., time  $t - 1$  and  $t - 2$ . Thus, a time series model such as AR model can be considered as a regression model where the explanatory variables are the lagged values of the dependent variable.

The most simple of the presented models is the univariate AR model, which is utilized in the modelling of temporal dependency structures in individual locations with both WPPs and PVPs in Publications I, II and IV. In all of the studied applications, there were no need to include the moving-average part of the general ARMA model as AR part of the model was sufficient. In addition, ARMA processes can be also approximated using long AR models [80], and hence, the MA part or ARMA models are not considered in the thesis.

An univariate  $p$ -order AR model for a process  $x_t$  is defined as

$$x_t = c + \sum_{h=1}^p a_h x_{t-h} + u_t, \quad (3.5)$$

where  $a_1, \dots, a_p$  are the model coefficients,  $c$  is the intercept term, and  $u_t$  is the error term of the model at time  $t$ . In the standard approach  $a_1, \dots, a_p$  and  $c$  are assumed to be constants, and thus, remain the same for all  $t$ . The error terms of

the model,  $u_t$ , are assumed to be independent and identically distributed (i.i.d.), however no assumptions concerning the distribution of the errors are made.

An univariate AR process models, by definition, only the temporal dependencies in a single time series  $\mathbf{x}_t$ . Therefore, to model spatial dependency structures between multiple time series, univariate AR models are combined using a correlation matrix  $\mathbf{C}$  as done in [30].

The mathematical approach for the implementation of the spatial correlations between univariate time series  $\mathbf{x}_t$  (with temporal correlations modelled with the AR models), is the utilization of the Cholesky decomposition of correlation matrix  $\mathbf{C}$ . The Cholesky decomposition of  $\mathbf{C}$  can be written as

$$\mathbf{C} = \mathbf{L}\mathbf{L}^\top, \quad (3.6)$$

where  $\mathbf{L}$  is a lower triangular matrix with real and positive diagonals and  $\mathbf{L}^\top$  is the transpose of  $\mathbf{L}$ . The spatial dependency structures are implemented by multiplying a matrix consisting of the spatially uncorrelated time series with  $\mathbf{L}$  [30].

The combination of these methods is referred as the ARC model in this thesis according the combination of an univariate AR models and Cholesky decomposition. For the ARC model, the parameters that need to be estimated from data are AR model coefficients  $a_1, \dots, a_p$ , intercept term  $c$  and the correlation matrix  $\mathbf{C}$ . The estimation of these parameters is straightforward from the data, and is presented in detail for WPPs in Publication II and for PVPs in Publication IV and in Section 3.4.

### 3.2.2 Simplified Vector Autoregressive Approach

In the ARC model, as presented in the previous section, the spatial correlations for lag  $h = 0$  are created using Cholesky decomposition of spatial correlation matrix  $\mathbf{C}$ . However, Cholesky decomposition does not fully preserve the shape of the autocorrelation function (ACF), i.e., the temporal dependencies, in individual locations. This is not an issue if all locations have relatively similar shape of the ACF, which is true when modelling only WPPs (wind speeds) or PVPs (solar irradiances). However, when modelling, e.g., both generation types with the same model to capture the correlations between WPP and PVP locations, it distorts the ACFs in the individual locations as the ACFs differ notably between wind speeds and solar irradiances.

Therefore, an approach which preserves ACFs in individual locations is required, and hence, the simplified VAR model, which models both temporal dependencies in individual locations and spatial dependencies between them, was proposed in Publication V. In addition, the simplified VAR model based approach was also a crucial intermediary step after the ARC model in the development of the full VAR model based approach, which is presented in the next section.

In Publications V and VI, the temporal and spatial dependency structures

are modelled with a simplified VAR model. In Publication V, an approach to specify the simplified VAR model parameters for new WPP and PVP locations was proposed and is presented shortly in this section. In Publication VI, a simplified version of the presented approach, applicable only for model order  $p = 1$ , is utilized.

A VAR model is a multivariate generalization of an univariate AR model which models temporal and spatial dependencies in multiple time series. Therefore, Cholesky decomposition of a spatial correlation matrix or any similar approach is not required to accompany the VAR model.

A  $k$ -dimensional  $p$ -order VAR $_k(p)$  model is defined for a process  $\mathbf{x}_t = [x_{1,t}, x_{2,t}, \dots, x_{k,t}]'$  for time steps  $t = 1, \dots, T$  as

$$\mathbf{x}_t = \mathbf{c} + \sum_{h=1}^p \mathbf{A}_h \mathbf{x}_{t-h} + \mathbf{u}_t, \quad (3.7)$$

where  $\mathbf{A}_1, \dots, \mathbf{A}_p$  are the  $k \times k$ -dimensional VAR coefficient matrices,  $\mathbf{c}$  is a  $k \times 1$ -vector of intercept terms and  $\mathbf{u}_t = [u_{1,t}, u_{2,t}, \dots, u_{k,t}]'$  is the error term of the model [46]. Index  $k$  is the number of individual locations (e.g., WPP or PVP locations) and  $i$  represents one of those locations. The model order  $p$  describes how many previous values (time steps) of  $\mathbf{x}_t$  are used to predict the next value of  $\mathbf{x}_t$ .

The model is called a simplified VAR because the off-diagonal components of the VAR coefficient matrices  $\mathbf{A}_1, \dots, \mathbf{A}_p$  are assumed to be zero, yielding

$$\mathbf{A}_h = \begin{bmatrix} a_{h,1} & 0 & \cdots & 0 \\ 0 & a_{h,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{h,k} \end{bmatrix}, \quad (3.8)$$

where  $a_{h,1}, \dots, a_{h,k}$  are model coefficients, which specify the temporal dependency structures separately for each location  $i$ . In general, the diagonal components of the VAR coefficient matrices specify the temporal dependencies in each individual location, i.e., how the value  $x_{i,t}$  depends on the past values of the same location  $i$  with lags  $1, \dots, h$ . The off-diagonal components specify the spatial dependencies between locations, i.e., how  $x_{i,t}$  depends on the past values of all other  $k - 1$  locations at lags  $1, \dots, h$ .

The estimation of the off-diagonal components of  $\mathbf{A}_1, \dots, \mathbf{A}_p$  is difficult, as they cannot be linked to straightforwardly to, e.g., the temporal correlations of wind speeds or solar irradiances in a single location or to the distances between two locations, which is discussed in more detail in Publications I and III. Hence, the off-diagonal components are assigned to zeros to enable the easier estimation of the model parameters for the modelling of new WPPs and PVPs.

As the VAR coefficient matrices are as specified in (3.8), the spatial correlations between the components of  $\mathbf{x}_t$  have to come exclusively from the error terms  $\mathbf{u}_t$ . They can be specified using the covariance matrix of the error terms  $\boldsymbol{\Sigma}_u = \text{cov}(u_t)$ .

$\Sigma_u$  can be specified using VAR coefficient matrices  $\mathbf{A}_1, \dots, \mathbf{A}_p$  and autocovariance matrices  $\Gamma_x(h)$ . These parameters, on the other hand, can be estimated from data. This section shows how  $\Sigma_u$  is specified and the estimation of the parameters from data is presented in Sections 3.5.2 and 3.5.3. In addition, more information can be found in Publication V.

To be able to determine  $\Sigma_u$ , the VAR model has to be assessed in the state-space presentation form, i.e., as a  $\text{VAR}_{kp}(1)$  model [46]. The state-space representation allows a  $\text{VAR}_k(p)$  model with any order  $p$  to be written as a  $kp$  –dimensional  $\text{VAR}_{kp}(1)$  model, i.e., with order  $p = 1$ , which is necessary to determine  $\Sigma_u$  with the utilized approach. First, the  $kp$  –dimensional state-space autocovariance matrix  $\Gamma_X(0)$  of the process  $\mathbf{x}_t$  is formed from the autocovariance matrices  $\Gamma_x(h)$  for lags  $h = 0, \dots, p - 1$ . Matrix  $\Gamma_X(0)$  specifies the temporal and spatial autocovariances of  $\mathbf{x}_t$  for lags  $h = 0, \dots, p - 1$ , and is specified as follows

$$\Gamma_X(0) = \begin{bmatrix} \Gamma_x(0) & \Gamma_x(1) & \cdots & \Gamma_x(p-1) \\ \Gamma_x(-1) & \Gamma_x(0) & \cdots & \Gamma_x(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_x(-p+1) & \Gamma_x(-p+2) & \cdots & \Gamma_x(0) \end{bmatrix}. \quad (3.9)$$

In addition to  $\Gamma_X(0)$ , a  $kp \times kp$  –dimensional state-space representation of the VAR coefficient matrices  $\mathbf{A}_1, \dots, \mathbf{A}_p$  is also required, and can be formed in the following manner

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_k & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{I}_k & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{I}_k & 0 \end{bmatrix}, \quad (3.10)$$

where  $\mathbf{I}_k$  is a  $k \times k$  –dimensional identity matrix [46]. According to [46], using (3.9) and (3.10) the  $kp \times kp$  –dimensional covariance matrix  $\Sigma_U$  can be obtained as follows

$$\Sigma_U = \Gamma_X(0) - \mathbf{A}\Gamma_X(0)\mathbf{A}^\top. \quad (3.11)$$

The  $k \times k$  –dimensional covariance matrix of the error terms  $\Sigma_u$ , which specifies the spatial dependency structures is obtained by selecting first  $k$  rows and first  $k$  columns from  $\Sigma_U$  as shown in Publication V.

### 3.2.3 Full Vector Autoregressive Approach

The previous section presented the simplified VAR model for the modelling of spatial and temporal dependencies in new WPPs and PVPs. Simplified VAR model considers the spatial dependency structures only with lag  $h = 0$ , as does the ARC model with the spatial correlation matrix  $\mathbf{C}$ . This is enough to capture

the spatial dependency structures between solar irradiance locations accurately. However, modelling the spatial dependencies only with lag  $h = 0$  is not enough for the accurate modelling of wind speeds in multiple WPP locations.

The spatial dependency structures in wind speed data are such that lags  $h > 0$  have to be taken into account also in the spatial correlations to capture the correct dependency structure between locations. Hence, a full VAR model based methodology, which is able to model the spatial dependencies with multiple lags is proposed for the modelling of WPPs in multiple locations.

In Publication III, the simplified VAR model was developed further into a full VAR model to capture the full spatio-temporal dependency structure between WPP locations and an approach to specify the model parameters for new WPP locations was presented. In addition, Publication VII also utilized a VAR model, but only with existing locations. This section presents the full VAR approach proposed in Publication III and how the model parameters are specified when modelling new locations.

Equation (3.7) presented the  $k$ -dimensional  $p$ -order  $\text{VAR}_k(p)$  model for a process  $\mathbf{x}_t = [x_{1,t}, x_{2,t}, \dots, x_{k,t}]'$ . With the full VAR model approach, the VAR coefficient matrices  $\mathbf{A}_1, \dots, \mathbf{A}_p$  associated with the model are specified without any components forced to zeros, as

$$\mathbf{A}_h = \begin{bmatrix} a_{h;1,1} & a_{h;1,2} & \cdots & a_{h;1,k} \\ a_{h;2,1} & a_{h;2,2} & \cdots & a_{h;2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{h;k,1} & a_{h;k,2} & \cdots & a_{h;k,k} \end{bmatrix}, \quad (3.12)$$

where, e.g.,  $a_{h;1,2}$  denotes the coefficient between  $x_{1,t}$  and  $x_{2,t}$  at lag  $h$ .

In contrast to the specification of the simplified VAR model parameters, with the full VAR model the VAR coefficient matrices  $\mathbf{A}_1, \dots, \mathbf{A}_p$  have to be specified using autocovariance matrices  $\Gamma_x(h)$ . This section shows how  $\mathbf{A}_1, \dots, \mathbf{A}_p$  are specified for new locations. The estimation of the required autocovariance matrices  $\Gamma_x(h)$  from the data is presented in Section 3.3.3.

The specification of the parameters starts with Yule-Walker equations, which can be written, according to [46], for lags  $h > 0$  as

$$\Gamma_x(h) = \mathbf{A}_1 \Gamma_x(h-1) + \dots + \mathbf{A}_p \Gamma_x(h-p). \quad (3.13)$$

Equation (3.13) specifies the relation between matrices  $\mathbf{A}_1, \dots, \mathbf{A}_p$  and  $\Gamma_x(h)$  as a function of lag  $h$ , and, can be formulated also as

$$\Gamma_x(h) = [\mathbf{A}_1, \dots, \mathbf{A}_p] \begin{bmatrix} \Gamma_x(h-1) \\ \vdots \\ \Gamma_x(h-p) \end{bmatrix}. \quad (3.14)$$

According to [46], using the state-space representation, i.e., the  $\text{VAR}_{k,p}(1)$  form of the VAR model, and thus, using matrices (3.9) and (3.10), (3.14) can be written

as

$$[\Gamma_x(1), \dots, \Gamma_x(p)] = \mathbf{A}\Gamma_X(0). \quad (3.15)$$

As the VAR model coefficients need to be specified for the modelling of new locations, elements of  $\mathbf{A}$ , where the matrices  $\mathbf{A}_1, \dots, \mathbf{A}_p$  are specified as in (3.12), can be solved from (3.15) as follows

$$\mathbf{A} = [\Gamma_x(1), \dots, \Gamma_x(p)]\Gamma_X(0)^{-1}, \quad (3.16)$$

where  $\Gamma_X(0)$  is non-singular. Using (3.16), the VAR coefficient matrices can be specified for the modelling of new WPP locations using autocovariance matrices  $\Gamma_x(h)$ , which can be estimated from data, as shown in Publication III.

The covariance matrix of the error terms  $\Sigma_u$  is the only parameter left to specify for the full VAR model.  $\Sigma_u$  can be calculated from the already specified matrices  $\mathbf{A}$  and  $\Gamma_X(0)$  similarly as with the simplified VAR model in the previous section using (3.11) and then selecting the first  $k$  rows and first  $k$  columns from  $\Sigma_U$ .

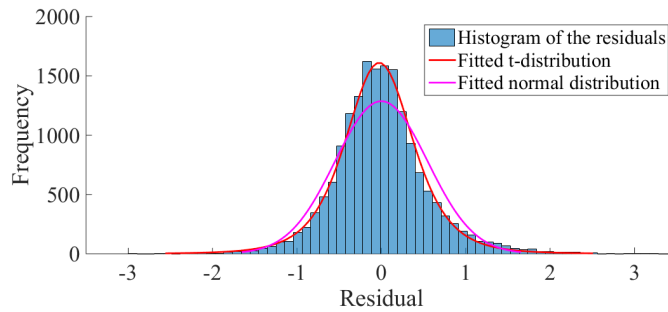
### 3.2.4 Error Term Modelling

In Publication I, the error terms of the AR model were assumed to be normally distributed when modelling transformed wind speeds. However, in Publication II it was found out, when analysing the error terms in detail, that the errors were  $t$ -distributed instead of following a normal distribution. The error terms  $\mathbf{u}_t = [u_{1,t}, u_{2,t}, \dots, u_{k,t}]'$  were found to have high kurtosis in all cases and the distributions had notably fatter tails and more peakedness compared with the normal distribution. Similar observation was made in Publication IV, the error terms of the AR models, when applied in the modelling of solar irradiance, had also fat tails, and hence, were  $t$ -distributed.  $t$ -distributions were also used with the simplified and full VAR models when modelling new WPPs and PVPs in Publications III and V. In addition, in Publication VII  $t$ -distributed errors were used with the modelling of wind generation in regional areas.

The mean value of the degrees of freedom from the  $t$ -distributions fitted for the error terms of solar irradiance locations were 3.68 in Publication IV and for wind speed locations 5.81 in Publication III. Hence, it was clear that  $t$ -distribution gave a much better fit as  $t$ -distribution approaches normal distribution when the degrees of freedom increase (usually with  $> 30$  degrees of freedom  $t$ -distribution is close to normal distribution). The fitted normal and  $t$ -distribution are also illustrated in Figure 3.2 for the errors of a time-varying AR model fitted for a measured solar irradiance location in Publication IV.

The error terms of the model were assumed to follow  $t$ -distribution with the modelling of both WPPs and PVPs and with all used models (except the modelling of wind directions where errors followed multivariate normal distribution). The model parameters were estimated with ordinary least squares (OLS) method in

all cases. According to Gauss-Markov Theorem, OLS is the best linear unbiased estimator if model errors  $\mathbf{u}_t$  have zero mean, have constant variance and are uncorrelated [81]. Although, with VAR and simplified VAR models, the error terms are allowed to be correlated at lag  $h = 0$ , but have to be uncorrelated between time steps [46]. In addition to these conditions, OLS does not require normality of the errors and it allows their separate modelling [82]. Thus, the usage of  $t$ -distributed errors were justified.



**Figure 3.2.** Histogram of the error terms of a time-varying AR model fitted for solar irradiance measurements and the fitted normal and  $t$ -distributions from Publication IV.

### 3.3 Modelling of New Wind Power Plant Locations

This section presents the modelling methodology for new WPP locations. New WPPs have been modelled in Publications I, II and III and during the course of the publications, the modelling methodology has been further developed and improved. The section goes through the main modelling methodologies step by step with an emphasis on the most developed approach utilizing the full VAR model developed in Publication III.

#### 3.3.1 Estimation Data and Marginal Distributions

The estimation data for the models can be any measured wind speed data as long as the data is measured at least partly at the same time period, the measurement period is over one year (to allow the simulation of full years) and has at least hourly time resolution. The models presented here have been estimated using a dataset which consists both low and high altitude measurement data.

The low altitude data consisted of hourly wind speed data from 19 low altitude measurement locations in Finland (the measurement height is approximately 15 meters above the surrounding ground level). The measurement period was three years from July 2008 to July 2011 for all locations. The data was provided by the Finnish Meteorological Institute (FMI). The high altitude wind speed data are from 12 locations in Finland and the time resolution is one hour. The

measurement height varies from location to location 74 to 150 meters above the surrounding ground level. The length and dates of the measured data varies from location to location from a minimum of one year to several years.

In addition, measured aggregated wind power generation data from Finland in 2012 and 2015 are also utilized in the estimation or verification of the models. The resolution of the aggregated data is hourly and it was provided by the Finnish Energy Industries.

Marginal distributions, i.e., margins are the probability distributions describing the wind speed conditions in a location. Several different distributions can be used to model the distribution of wind speeds, e.g. Weibull distributions, empirical cumulative distribution functions (ECDFs) [20] or ECDFs combined with generalized Pareto distribution [32, 33] for the upper tail (highest 10% of the data) in order not to limit the distribution at the highest measured value. In the model estimation ECDFs combined with Pareto tails (in this section just ECDFs) are used to model the margins of the measured locations. However, these distributions cannot be used to specify new locations as they require measurement data from the modelled location. Therefore, the margins of new WPP locations are specified with Weibull distributions, defined by two parameters, and commonly used to describe wind speed distributions [42, 43].

The measured wind speeds from  $k$  locations, used in the estimation, can be denoted as  $\mathbf{y}_t = [y_{1,t}, y_{2,t}, \dots, y_{k,t}]'$ , where  $k$  denotes the number of the locations. The AR or VAR models usually assume the estimation data to be normally distributed [34, 46], and thus,  $\mathbf{y}_t$  is transformed to data with normally distributed margins  $\mathbf{z}_t^{\text{wd}}$ , where *wd* denotes *with day structures*. This means that the monthly diurnal variations are still present in the data. The probability integral transformation, similar to the one used in the Gaussian copula, to Gaussian data can be written as

$$z_{i,t}^{\text{wd}} = F_N^{-1} [\hat{F}_i(y_{i,t})], \quad (3.17)$$

where  $F_N^{-1}$  is the inverse CDF of the standard normal distribution and  $\hat{F}_i$  is the estimated ECDF margin (as specified earlier in this section), which describes the wind speed distribution (i.e. wind conditions) in location  $i$  [79].

The monthly diurnal structures, i.e., the fluctuations of the average values in the data are still included in the transformed data  $\mathbf{z}_t^{\text{wd}}$ . These monthly day structures are caused by average monthly and daily changes in the weather and wind conditions. These changes are caused by the change of seasons, and thus, the average temperature, but also the diurnal cycle of day and night and the related temperature patterns.

The daily and monthly average values are calculated for each hour of the day and for each month, yielding  $24 \times 12 = 288$  averages describing the mean values for the 24 hours for each of the 12 months. These averages are subtracted from  $\mathbf{z}_t^{\text{wd}}$ , and thus, transformed data suitable for the model estimation  $\mathbf{z}_t = [z_{1,t}, z_{2,t}, \dots, z_{k,t}]'$  is obtained similarly as in [34]. In addition, as the VAR model assumes stationarity [46], the stationarity of  $\mathbf{z}_t$  is ensured with Dickey-Fuller

test [83].

### 3.3.2 Model Selection and Parameter Estimation

The time series model used for the modelling of the temporal and spatial dependency structure in the new WPP locations can be estimated for  $\mathbf{z}_t$ . The model can be the ARC model as presented in Section 3.2.1 or the more advanced full VAR model presented in Section 3.2.3. As the full VAR model is the most advanced of the developed methods to model new WPP locations, this section focuses mainly on it. The models selection and parameter estimation for the ARC model is presented in Publication II.

The model identification can be done with several different approaches, both graphical and numeric. One of these is the analysis of the autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs) of the estimation data  $\mathbf{z}_t$ . For the full VAR model, the suitable order  $p$  was found to be 4 in Publication III.

The adequacy of the model can be assessed also with various approaches. The adequacy of the  $\text{VAR}_k(4)$  model was verified with the assessment of the ACFs, PACFs and cross-correlation functions (XCFs) of the model residuals  $\mathbf{u}_t$ , and with the Ljung-Box Q-test [84]. The residuals passed the Q-test and with visual assessment, it was noted that there were no statistically significant components, except XCF with lag  $h = 0$ , which is allowed for the residuals.

As already discussed in Section 3.2.4, the margins of the errors  $\mathbf{u}_t$  were modelled with  $t$ -distributions. For the new WPPs, the average degrees of freedom estimated from the estimation locations were used. The monthly diurnal variations for new WPP locations are estimated from transformed data  $\mathbf{z}_t^{\text{wd}}$  using the averages of the estimation locations. The Weibull parameters for the wind speed margins describing the wind speed conditions in the new locations are obtained from the Wind Atlas database [85] according to the heights of the turbines and the coordinates of the location.

### 3.3.3 Estimation of the Temporal and Spatial Correlations

Section 3.2.3 presented the approach to specify the VAR coefficient matrices for the new WPP locations. The covariance matrix  $\Sigma_u$  of the errors, on the other hand, is specified as shown in Section 3.2.2. Both of these specifications require autocovariance matrices  $\Gamma_z(h)$  for lags  $h = 0, \dots, 4$ , which are estimated as follows.

First, autocorrelation matrices  $\mathbf{R}_z(h)$  are estimated for  $h = 0, \dots, 4$  from  $\mathbf{z}_t$ . The temporal correlations, i.e., the diagonal components of  $\mathbf{R}_z(h)$ , are the average values of the ACFs estimated from the components of  $\mathbf{z}_t$ . The spatial and spatiotemporal (spatial with lags  $h \neq 0$ ) correlations, i.e., the off-diagonal components of  $\mathbf{R}_z(h)$  are estimated using the distances between the locations, which are linked to the underlying correlations in  $\mathbf{z}_t$ , as shown in Figure 3.3 from

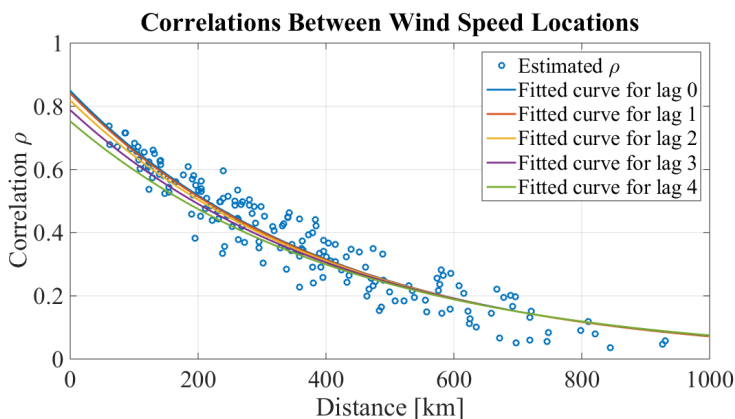
Publication III. Figure 3.3 illustrates the relationship between the correlations of  $\mathbf{z}_t$  at  $h = 0$  and distances between the measurement locations, and the fitted curves for lags  $h = 0, \dots, 4$ . These fitted curves are used to estimate the spatial and spatiotemporal correlations for the new WPP locations depending on the distances between the locations. Similar approach has been taken for the estimation of lag  $h = 0$  correlations as a function of distances between WPPs in [34, 86].

When the  $\mathbf{R}_z(h)$  are fully specified for  $h = 0, \dots, 4$ , they can be transformed to the autocovariance matrices  $\mathbf{\Gamma}_z(h)$  with the following equation

$$\mathbf{\Gamma}_z(h) = \mathbf{D}\mathbf{R}_z(h)\mathbf{D}, \quad (3.18)$$

where  $\mathbf{D}$  is a  $k \times k$  diagonal matrix with the averages of the standard deviations  $\sigma_i$  of the new locations in the diagonal [46]. The averages of  $\sigma_i$  estimated from the components of  $\mathbf{z}_t$  are used for the new locations. When  $\mathbf{\Gamma}_z(h)$  for lags  $h = 0, \dots, 4$  are fully specified, VAR coefficient matrices for new WPPs can be estimated using (3.16).

After the VAR coefficient matrices are estimated,  $\mathbf{\Sigma}_u$  can be estimated using (3.11) and selecting the  $k$  first rows and  $k$  first columns from  $\mathbf{\Sigma}_U$ .  $\mathbf{\Sigma}_u$  is modelled as a Gaussian copula [79], and its margins are  $t$ -distributed. Correlation matrix for the errors  $\mathbf{C}_u$  can be estimated from  $\mathbf{\Sigma}_u$  and is used to approximate the correlation matrix  $\mathbf{C}_{\text{Gauss}}$ , which specifies the Gaussian copula. This approach is justified in Publication III.



**Figure 3.3.** Spatial correlation  $\rho$  estimated from  $\mathbf{z}_t$  between wind speed locations for lag  $h = 0$ , plotted against the distances between the locations, and fitted curves for spatial correlations for lags  $h = 0, \dots, 4$  from Publication III.

### 3.3.4 Wind Turbine Power Generation Models

As the used time series model, e.g., the full VAR model, is estimated for wind speed measurement data, the output data from the simulations is also wind

speed time series. The transformation from the simulated wind speed data to generated power data can be done with various approaches. As the aim of the presented methodology is to model new WPPs, the power generation model used for the transformation cannot be based on approaches requiring power measurement data. Thus, the approach has to rely purely on the data sheet parameters of the turbines, which limits the number of viable approaches, some of which are presented next.

In Publication I, a simple linear piecewise power curve was used for the transformation. With this approach, the power generation was zero below the cut-in speed of the turbine and constant nominal power above the nominal speed. Between the cut-in and nominal speeds the generation was linearly increasing. In Publication II, a 3rd degree curve was fitted for the piecewise power curve between the cut-in and nominal speeds to increase the accuracy. In addition, the emergency shut down was implemented by cutting the generation to zero above the cut-out speed. After an emergency shut down, the wind speed has to decrease below a turbine specific re-cut-in speed before the turbine restarts.

In Publication III, a state-of-the-art power generation model, based on logistic functions was used to model the transformation between the cut-in and cut-out speeds [48]. The utilized three parameter logistic power curve can be defined for location  $i$  as

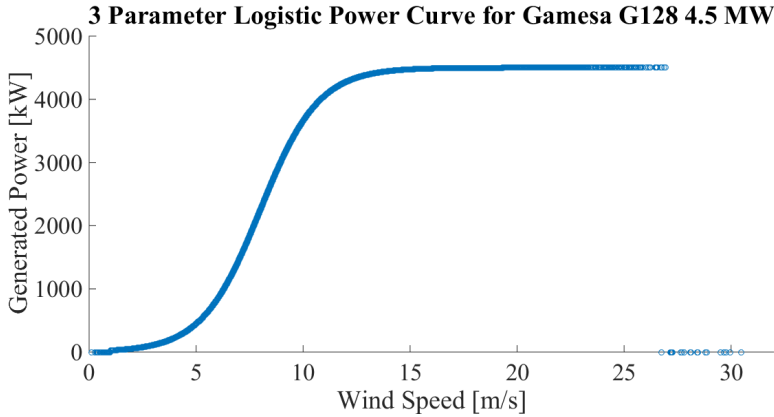
$$P(\tilde{y}_i) = \frac{P_r}{1 + e^{4s(y_{ip} - \tilde{y}_i)/P_r}}, \quad (3.19)$$

where  $P_r$  is the rated power,  $\tilde{y}_i$  is the simulated wind speed,  $y_{ip}$  is the wind speed in the inflection point, i.e., the point where the gradient of the power curve reaches its maximum and  $s$  is the value of the gradient (the slope) in the inflection point, all in location  $i$  [47]. All of these parameters can be estimated from the data sheet parameters. An example power curve for Gamesa 128 4.5 MW wind turbine [87] modelled with (3.19) is illustrated in Figure 3.4.

So far, only the output of a single wind turbine has been considered. However, the wake effect inside a wind farm has to be taken into account when modelling new WPPs with multiple turbines. One approach to model the wake effect is to use a single wake coefficient, which is estimated using the number of the turbines in the WPP and the average distance between the turbines [50], as done in Publications II, III and V. However, the wake effect can also be modelled in more detail using a dynamic approach, which requires wind direction information for each time step  $t$  [50]. Dynamic modelling of the wake effect detailed was studied in Publication VI and is shortly presented in Section 3.6.

### 3.3.5 Simulation Procedure

The modelling of new WPPs is based on long term MC simulations (however, the presented methods can also be used for short term simulations or forecasting, as shown in [41]). The general structure of the simulation process for the full VAR model approach is illustrated in the flowchart in Figure 3.5. As the full



**Figure 3.4.** Power curve of Gamesa G128 4.5 MW wind turbine [87] modelled with the three parameter logistic function. The cut-out-speed where the turbine shuts down is 27 m/s, and thus, the generated power above that wind speed is zero.

VAR approach is the most advanced of the WPP modelling approaches, only it is presented in detail. The simulation procedure for new WPPs utilizing the ARC model is presented in Publication II.

The process starts with the drawing of a random sample from the multivariate Gaussian copula for the  $k$  new WPPs for the length of the simulation runs  $t = 1, \dots, T$ . Then, the multivariate normal margins of the sample are transformed to  $t$ -distributions according to the estimated degrees of freedom of the errors  $\mathbf{u}_t$  for new WPP locations.

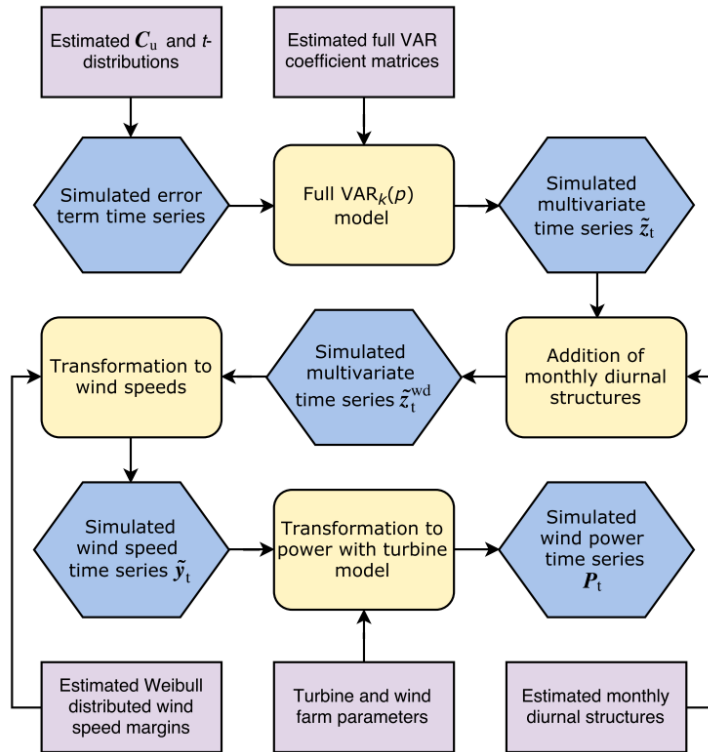
The next step is to simulate multivariate time series  $\tilde{\mathbf{z}}_t$  with the full  $\text{VAR}_k(4)$  model, using the correlated and  $t$ -distributed random sample as innovations. At this point, the simulated  $\tilde{\mathbf{z}}_t$  already contains the specified temporal and spatial dependency structures.

The addition of the estimated monthly diurnal structures follows as done in [34], and thus,  $\tilde{\mathbf{z}}_t^{\text{wd}}$  is obtained. Then,  $\tilde{\mathbf{z}}_t^{\text{wd}}$  is transformed to simulated wind speeds with the inverse transformation of (3.17) specified as

$$\hat{y}_{i,t} = \hat{F}_i^{-1} \left[ F_N(\tilde{z}_{i,t}^{\text{wd}}) \right], \quad (3.20)$$

where  $\hat{y}_{i,t}$  is the obtained simulated wind speed in new WPP location  $i$  at time  $t$ ,  $\hat{F}_i^{-1}$  is the estimated inverse CDF of the margin for location  $i$  (Weibull distributions describing the wind speed conditions in the location) and  $F_N$  is the CDF of the standard normal distribution.

Final step in the simulation procedure is to transform the simulated wind speeds  $\tilde{\mathbf{y}}_t$  to generated power with the three parameter logistic power curve approach as in 3.3.4. When the generated power for each new WPP is obtained with the power generation model, the resulting multivariate power time series can be utilized in the further analyses separately or combined into aggregated power time series representing the whole modelled system.



**Figure 3.5.** Flowchart of the MC simulation process for new WPP locations with the full VAR model based approach. Yellow boxes are operations, blue are simulated data and purple are estimated parameters.

### 3.4 Modelling of New Solar Power Plant Locations

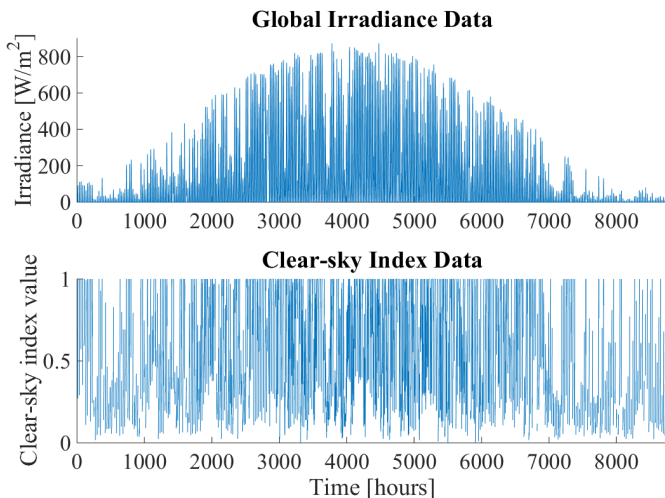
This section presents the modelling methodology for new PVP locations proposed in Publication IV. The methodology is based on the ARC modelling methodology for new WPPs proposed in Publications I and II and presented in Section 3.2.1. However, the ARC modelling is developed further and modified to be suitable for PVP modelling in Publication IV. To be applicable in the modelling of solar irradiance, the ARC model has to be able to handle the large daily and monthly variations of solar irradiance. Next, the PVP modelling methodology is presented in detail.

#### 3.4.1 Estimation Data and Marginal Distributions

The proposed ARC model is estimated using hourly solar (global) irradiance data, called as irradiance from this on, which is obtained from the Finnish Meteorological Institute (FMI). The estimation dataset consists of irradiance and temperature measurements from eight locations in Finland. The length of

the dataset varies from location to location from one to five years from January 2011 to January 2014, with at least one year of overlapping measurements with the other locations.

The irradiance data  $\mathbf{E}_t = [E_{1,t}, E_{2,t}, \dots, E_{k,t}]'$  from  $k$  locations consists of stochastic (caused by cloud movements and atmospheric conditions) and deterministic components. The deterministic component is caused by the rotations of the earth and the sun in relation to each other, and creates non-stationary fluctuations in monthly and daily time scales in the data. Therefore, it is not applicable to be used with autoregressive models, which require stationary estimation data. The solution is to use clear-sky index (CSI) data  $\mathbf{K}_t = [K_{1,t}, K_{2,t}, \dots, K_{k,t}]'$  instead the measured irradiance. Figure 3.6 illustrates  $\mathbf{E}_t$  and  $\mathbf{K}_t$  time series. To be able to obtain CSI data  $\mathbf{K}_t$ , clear-sky irradiance  $\mathbf{E}_t^{\text{CS}}$  data is required. Next, an approach to estimate clear-sky irradiance is presented.



**Figure 3.6.** One year of global irradiance and clear-sky index time series data from one of the measurement locations.

The clear-sky irradiance  $\mathbf{E}_t^{\text{CS}}$ , as the name suggests, provides the theoretical maximum irradiance for each hour of the year and can be obtained for any geographical location using a deterministic model proposed in [70]. The clear-sky irradiance model estimates both direct and diffuse components of the (global) clear-sky irradiance on hourly time resolution. The reflected radiation is not considered, as its contribution is minor and it is difficult to estimate accurately for new locations as the site specific conditions have a large impact on its values [88].

The utilized modelling approach was chosen as it requires only the coordinates and the Linke turbidity factors, which describe the amount of particles, and consequently the scattering and absorption of solar irradiance, in the atmosphere [70], in those coordinates as input parameters, and thus, is applicable for new

PVP locations. The Linke turbidity factors, which affect the transmittance of the solar irradiation, can be obtained from the worldwide Linke turbidity factor database [89] from the Solar Radiation Data service by Armines/MINES ParisTech, Centre Enerétique et Procédés. The database provides monthly averages of the Linke turbidity factors for specified coordinates.

The CSI data is obtained from the ratio between the measured irradiance  $\mathbf{E}_t$  and the theoretical clear-sky irradiance  $\mathbf{E}_t^{\text{CS}}$ , and thus, CSI can have only values between one and zero. The CSI data  $\mathbf{K}_t$  is calculated as

$$K_{i,t} = \frac{E_{i,t}}{E_{i,t}^{\text{CS}}}, E_{i,t}^{\text{CS}} \neq 0. \quad (3.21)$$

As visible from the transformation (3.21), only the hours when  $\mathbf{E}_t^{\text{CS}}$  is larger than zero, i.e. the hours when the sun is above the horizon, are included in the CSI data, which is used in the model estimation, otherwise  $\mathbf{K}_t$  is not specified. This is understandable for the model estimation perspective, as the stochastic component in the irradiance, that is actually modelled, is the cloudiness, and when measuring irradiance during night hours, there is no information about the cloudiness.

Next, ECDFs are fitted for  $\mathbf{K}_t$  to obtain the CSI margins for the estimation locations. The CSI data is stationary according to visual assessment and, e.g., augmented Dickey-Fuller test [83,90], but not normally distributed. As noted earlier, the AR models generally require normally distributed margins [46], and hence, the CSI data is transformed similarly as wind speed data in (3.17) to Gaussian data. This is done again utilizing the probability integral transformation, used also in the Gaussian copula, as follows

$$z_{i,t}^{\text{wd}} = F_N^{-1} [\hat{F}_i(K_{i,t})], \quad (3.22)$$

where  $F_N^{-1}$  is the inverse CDF of the standard normal distribution and  $\hat{F}_i$  is the estimated ECDF of the CSI margin for location  $i$  [79].

In Publication IV, some monthly diurnal structures were still found in the data despite the transformations to the CSI data, which removes the deterministic structures caused by the rotation of the earth and the sun, and then further to  $\mathbf{z}_t^{\text{wd}}$ . The remaining variations are caused by factors like local weather patterns. These monthly diurnal structures, i.e. the average values for each month and each hour of the day ( $12 \times 24 = 288$  average values in total), are calculated similarly as for wind speeds in Section 3.3.1. The calculated expected values are then subtracted from  $\mathbf{z}_t^{\text{wd}}$  and  $\mathbf{z}_t$  is obtained.

### 3.4.2 Model Selection and Parameter Estimation

Individual AR( $p$ ) models are used to model the temporal dependencies in the individual PVP locations and the locations are linked to each other (the spatial dependencies) using the Cholesky decomposition of the correlation matrix  $\mathbf{C}$ , as done in [30], and described in Section 3.2.1. In addition, with the PVP modelling,

time-varying model coefficients have to be used to capture the temporal dependency structures present in irradiance, and thus, the approach is called as the time-varying ARC model.

The time-varying AR( $p$ ) models are estimated from  $\mathbf{z}_t$  using OLS. The model identification was done by assessing the ACFs and PACFs of  $\mathbf{z}_t$  and by manually testing different models. In Publication IV, multiple different approaches were tested in the estimation of the time-varying coefficients (e.g., how many different groups should be used). The most suitable approach was to estimate the time-varying coefficients separately for six groups each consisting of time period of two months. In general, the summer months had smaller coefficients compared with the winter months, which indicates faster changes in the irradiance (cloudiness) during summer. The time-varying AR coefficients combined with the order of the model  $p = 3$  were found to be sufficient to remove all statistically significant ACF and PACF components from the residuals  $\mathbf{u}_t$ , and thus, verified the adequacy of the model. According to the these specification, the AR( $p$ ) models are specified as

$$\mathbf{z}_t = c + \sum_{h=1}^3 a_{h,t-h} x_{t-h} + \mathbf{u}_t, \quad (3.23)$$

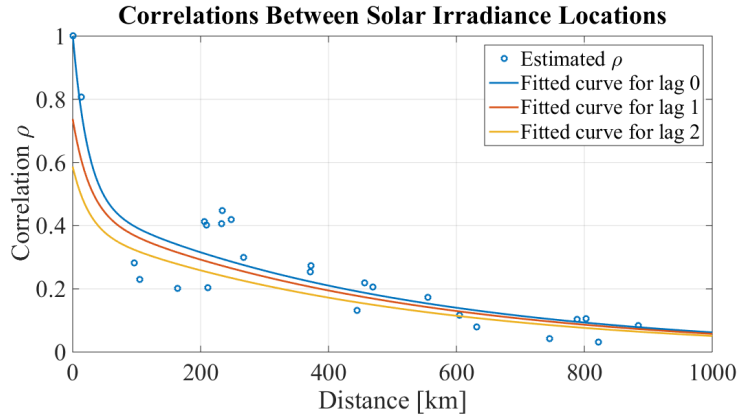
where model coefficients  $a_{1,t-1}, \dots, a_{p,t-p}$  change in time.

The residuals  $\mathbf{u}_t$  of the models were observed to be  $t$ -distributed in all estimation locations, as discussed in Section 3.2.4. Thus,  $t$ -distributions were fitted for the errors and the average of the estimated degrees of freedom are used for the new PVP locations. The average values of the estimated AR coefficients are also used as the AR coefficients specifying the temporal correlations in new PVP locations.

The ECDF from the closest measured location is used for the clear-sky index margin for new locations. Also, the monthly changing diurnal structures for new locations are from the closest estimation location. The clear-sky irradiance time series  $E_t^{\text{CS}}$  for the new locations are obtained with the theoretical clear-sky irradiance model [70] similarly as for the irradiance estimation locations in Section 3.4.1.

### 3.4.3 Estimation of Spatial Correlations

The spatial correlations between the new PVP locations for the correlation matrix  $\mathbf{C}$  are estimated from  $\mathbf{z}_t$ . As with the WPP locations in Section 3.3.3, the underlying correlations in  $\mathbf{z}_t$  are linked to the distances between the locations. Figure 3.7 illustrates the correlations of  $\mathbf{z}_t$  and distances between the measurement locations (at lag  $h = 0$ ), and the fitted curves for lags  $h = 0, 1, 2$ . The curve fitted for  $h = 0$  is used to estimate the correlations between the new PVP locations depending on the distances between the locations. Curves for lags 1 and 2 are used in the combined modelling of WPPs and PVPs with a more advanced simplified VAR model, as presented later in Section 3.5.



**Figure 3.7.** Spatial correlation  $\rho$  estimated from  $\mathbf{z}_t$  between solar irradiance locations for lag  $h = 0$ , plotted against the distances between the locations, and fitted curves for spatial correlations for lags  $h = 0, 1, 2$ .

### 3.4.4 Photovoltaic System Power Generation Model

Publication IV combines the ARC model based methodology with a power generation model to allow the simulation of the power output of a PV system. The modelling methodology produces global irradiance time series for the new PVP locations. However, detailed power generation models, which transform the irradiance to generated power, require the separation of global irradiance into beam and diffuse components. Solar generation other than PV systems are not considered in Publication IV or in this thesis, however, as the modelling methodology is modular, it can be used with any kind of power generation model requiring global irradiance or its components as input data.

In Publication IV, a well-performing logistic Boland-Ridley-Laurent (BRL) model is utilized to decompose the global irradiance [73,74]. The parameters of the model can be estimated from the clear-sky irradiance model and from the simulated CSI data  $\tilde{K}_{i,t}$ . A detailed presentation of the model and its parameter estimation can be found in Publication IV.

The irradiance data  $\mathbf{E}_t$  used in the estimation is for a horizontal surface, and thus, so is the simulated  $\tilde{\mathbf{E}}_t$  and its beam and diffuse components. However, usually the PV panels are tilted with a fixed angle or using sun tracking. Therefore, the horizontal beam and diffuse components are transformed to the desired tilt angle using an approach presented in [91]. A commonly used power generation model for polycrystalline silicon PV panels is used for the transformation to power [71]. It takes into account the ambient temperature, and the separate transmittances for both beam and diffuse components of the irradiance. The detailed presentation of the model and related equations can be found in Publication IV.

### 3.4.5 Simulation Procedure

The modelling of new PVPs follows a relatively similar MC simulation procedure as the WPP modelling in Section 3.3. The flowchart for this procedure is illustrated in Figure 3.8. The process begins with drawing a  $t$ -distributed sample with the size of the length of the simulation period for each new location. These  $t$ -distributed random numbers are used as the innovations for the time-varying AR(3) models to simulate temporally correlated time series for each location separately. Next, the spatial correlations are added to the time series using the Cholesky decomposition (3.6) of the correlation matrix  $\mathbf{C}$  as in [30]. Thus, multivariate time series  $\tilde{\mathbf{z}}_t$  with the required temporal and spatial correlations are obtained.

As with the WPPs earlier, next the monthly diurnal structures are added back to  $\tilde{\mathbf{z}}_t$  to obtain  $\tilde{\mathbf{z}}_t^{\text{wd}}$ . This is followed by the transformation to CSI data  $\mathbf{K}_t$  (the inverse transformation of (3.17) using the following transformation

$$\tilde{K}_{i,t} = \hat{F}_i^{-1} \left[ F_N(\tilde{z}_{i,t}^{\text{wd}}) \right], \quad (3.24)$$

where  $\tilde{K}_{i,t}$  is the simulated CSI data in new PVP location  $i$  at time  $t$  and  $\hat{F}_i^{-1}$  is the inverse of the ECDF margin for location  $i$  (obtained from the closest measured location).

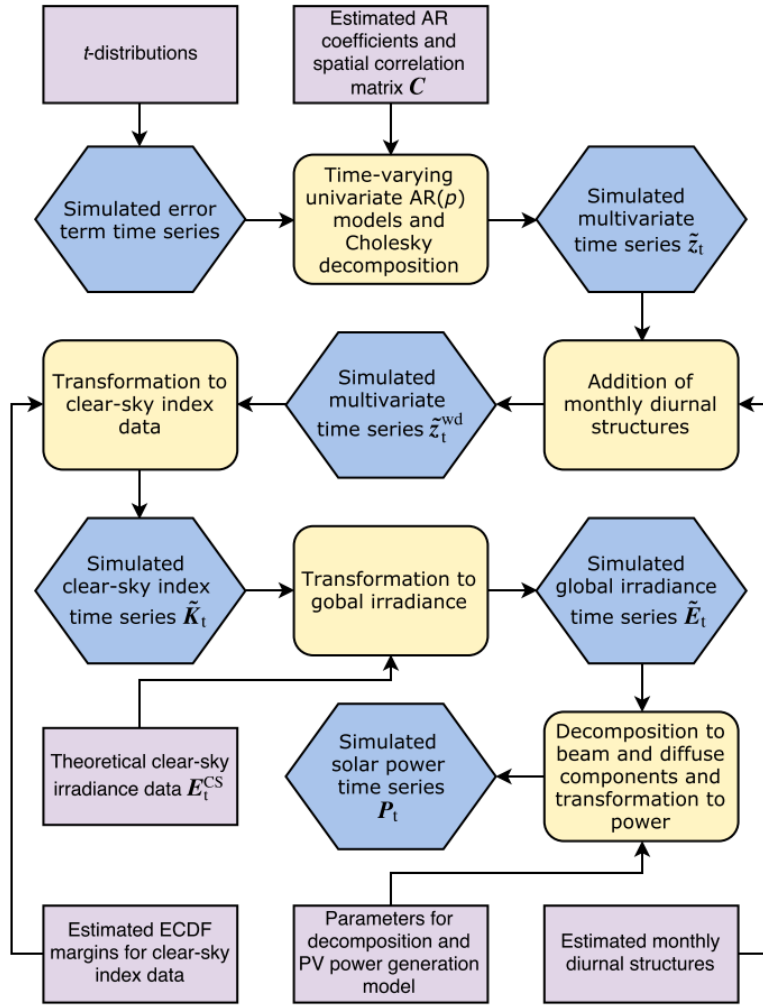
This is followed by the data transformation from  $\tilde{\mathbf{K}}_t$  to simulated irradiance data  $\tilde{\mathbf{E}}_t$  using (3.21) and the clear-sky irradiance  $\mathbf{E}_t^{\text{CS}}$  estimated for the new PVP locations. The last step is to transform the simulated irradiances  $\tilde{\mathbf{E}}_t$  to its beam and diffuse components, and further to the generated power as shown in Section 3.4.4. Then, the resulting generated power time series can be used in the further analyses separately or as an aggregate for system level analyses.

## 3.5 Combined Modelling of New Wind and Solar Power Plant Locations

This section presents the combined modelling of new WPPs and PVPs. Publication V proposes a methodology for the modelling of systems with both new WPPs and PVPs. The developed methodology combines the WPP and PVP modelling presented in Publications II and IV to a more advanced simplified time-varying VAR methodology to enable the combined modelling, which takes also into account the correlations between WPPs and PVPs.

### 3.5.1 Marginal Distributions

The combined modelling approach utilizes the same datasets as WPP modelling in Section 3.3 and PVP modelling in Section 3.4. The margins and the data pre-processing are handled separately for the WPPs and PVPs following the exact approaches already presented for wind speed data and margins in Sections 3.3.1



**Figure 3.8.** Flowchart of the MC simulation process for new PVP locations with the time-varying ARC model based approach. Yellow boxes are operations, blue are simulated data and purple are estimated parameters.

and for solar irradiance data and margins in Section 3.4.1. Global irradiance data  $\mathbf{E}_t$  is transformed to CSI data  $\mathbf{K}_t$  using the clear-sky irradiance model and then both  $\mathbf{K}_t$  and wind speed data  $\mathbf{y}_t$  are transformed to normally distributed  $\mathbf{z}_t^{\text{wd}}$  with the respective transformations (3.22) and (3.17). The monthly diurnal structures are then estimated and removed as with wind speeds in Section 3.3.1 and solar irradiances in 3.4.1 to obtain stationary  $\mathbf{z}_t$  to be used the estimation of the simplified VAR model.

### 3.5.2 Model Selection and Parameter Estimation

Publications II and IV utilized the ARC modelling approach, however, in Publication V with the combined modelling, a simplified time-varying VAR model with  $t$ -distributed error terms is used to model the temporal and spatial dependency structures between the generation locations.

The following specifications are made for the simplified VAR model. As the time series model follows the simplified VAR approach, the non-diagonal components of the VAR coefficient matrices are assigned to zeros as shown in (3.8). The intercept term of the VAR model,  $\mathbf{c}$ , are also assumed to be zeros, which is justified as  $\mathbf{z}_t$  follows the standard normal distribution. The model selection was conducted by analysing the ACFs and PACFs and using Ljung-Box Q-test [84], similarly as with the full VAR model in Section 3.3.2. The sufficient order for the model was found to be  $p = 3$ , as discussed in Publication V.

As the modelling of irradiance requires time-varying model coefficients, as discussed in Section 3.4.2, the simplified VAR coefficient matrices  $\mathbf{A}_{1,t}, \dots, \mathbf{A}_{p,t}$  become time-varying and thus, can be specified as in (3.8), but in the case of time-varying model, the coefficient  $a_{h,i}$  change in time and are thus specified as  $a_{h,t,i}$

Unique VAR coefficients are estimated separately for six two month groups, similarly as in the PVP modelling in Section 3.4.2. However, it should be noted that as the wind speed modelling does not require the time-varying VAR coefficients, and thus, the coefficients related to WPP locations remain the same with all  $t$ . Last, the errors of the model  $\mathbf{u}_t$  are  $t$ -distributed as discussed in Section 3.2.4. Consequently, the VAR model, as presented in (3.7) is now specified for the combined modelling as

$$\mathbf{z}_t = \sum_{h=1}^3 \mathbf{A}_{h,t-h} \mathbf{x}_{t-h} + \mathbf{u}_t. \quad (3.25)$$

The parameter estimation follows a similar approach as in Section 3.2.2. The diagonal coefficients for the time-varying VAR matrices  $\mathbf{A}_{1,t}, \dots, \mathbf{A}_{3,t}$  can be estimated with univariate time-varying AR(3) models using OLS, similarly as with solar irradiances in Section 3.4.2. This is possible, as each coefficient specifies the temporal dependency structure only in one individual location, as shown in Section 3.2.2. The average values of the time-varying VAR coefficients of the same type are used for new WPP and PVP locations when forming the state-space representation of the coefficient matrices as shown in (3.10). This is done to simplify the calculation of  $\Sigma_U$  with (3.11) (and as it did not have notable effect on the results in MC simulation studies).

The monthly changing diurnal structures and the ECDFs of the CSI margins for new PVP locations are from the closest measured location. For new WPPs and PVPs, the average degrees of freedom estimated from  $\mathbf{u}_t$  are used to specify the  $t$ -distributed margins of the innovations. The Weibull parameters for the wind speed margins of the new WPP locations are from the Wind Atlas database [85]

similarly as in Section 3.3. The clear-sky irradiances  $E_t^{\text{CS}}$  for new PVP locations are estimated with the clear-sky irradiance model [70] as in Section 3.4.

### 3.5.3 Estimation of Spatial Correlations

As the off-diagonals of the VAR coefficient matrices are zeros, all spatial correlations have to come from the covariance matrix of the error terms  $\Sigma_u$ , which is estimated following the process presented in Section 3.2.2. This process requires the specification of the VAR coefficient matrices  $A_{1,t}, \dots, A_{3,t}$  and the autocovariance matrices  $\Gamma_z(h)$  for lags  $h = 0, 1, 2$ .  $\Gamma_z(h)$  are obtained from the correlation matrices  $R_z(h)$  as in Section 3.3.3 with (3.18), where the required standard deviations  $\sigma_i$  are averages of the  $\sigma_i$  estimated separately from WPPs and PVPs from the components of  $z_t$ .

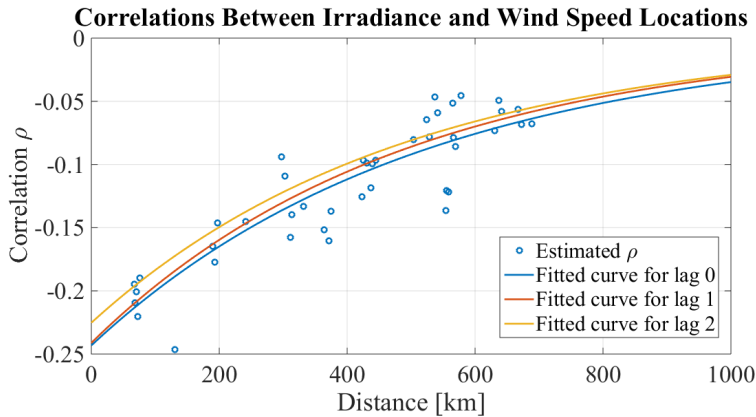
The temporal correlations, i.e., the diagonal components of  $R_z(h)$ , are obtained analytically from the averages of the estimated time-varying VAR coefficient matrices  $A_{1,t}, \dots, A_{3,t}$  using Yule-Walker equations [46]. This is discussed more in Publication V.

The spatial correlations, i.e., the off-diagonal components of  $R_z(h)$ , are estimated using the fitted curves for lags  $h = 1, 2, 3$  shown in Figures 3.3 and 3.7. These are used for the components specifying the correlation between two wind or two irradiance locations. The components specifying the dependence between wind and irradiance locations are obtained with the curves illustrated in Figure 3.9. Wind speeds and irradiances are negatively correlated as shown in [76, 78]. However, as visible from Figure 3.9, also the underlying correlations in  $z_t$  between wind speeds and irradiances are negatively correlated even when the deterministic components, such as the daily and monthly variations, are removed from the data.

With the VAR coefficient and the autocovariance matrices fully estimated,  $\Sigma_u$  can be estimated with (3.11) as for wind speeds in Section 3.3.3. Similarly,  $\Sigma_u$  is modelled here also as a Gaussian copula [79] with  $t$ -distributed margins, and approximated from the correlation matrix of the errors  $C_u$ .

### 3.5.4 Simulation Procedure

The MC simulation procedure for the simplified VAR model follows a similar approach as with the full VAR model in Section 3.3.5. The simulation process for the simplified time-varying VAR based approach is illustrated in the flowchart in Figure 3.10. When simulated  $\tilde{z}_t$  (contains the spatial and temporal dependencies) is obtained, the further processing is separated for the WPP and PVP locations. First, the monthly diurnal structures are added and then the data is transformed to wind speeds  $\tilde{y}_t$  with (3.20) for the WPP locations and to CSI data  $\tilde{K}_t$  with (3.24) for the PVP locations. CSI data is transformed further to  $\tilde{E}_t$  as shown in Section 3.4.5. The final step is to transform the both generation types into power with the respective power generation models.



**Figure 3.9.** Spatial correlation  $\rho$  estimated from  $\mathbf{z}_t$  between solar irradiance and wind speed locations for lag  $h = 0$ , plotted against the distances between the locations, and fitted curves for spatial correlations for lags  $h = 0, 1, 2$ .

### 3.6 Wind Direction Modelling

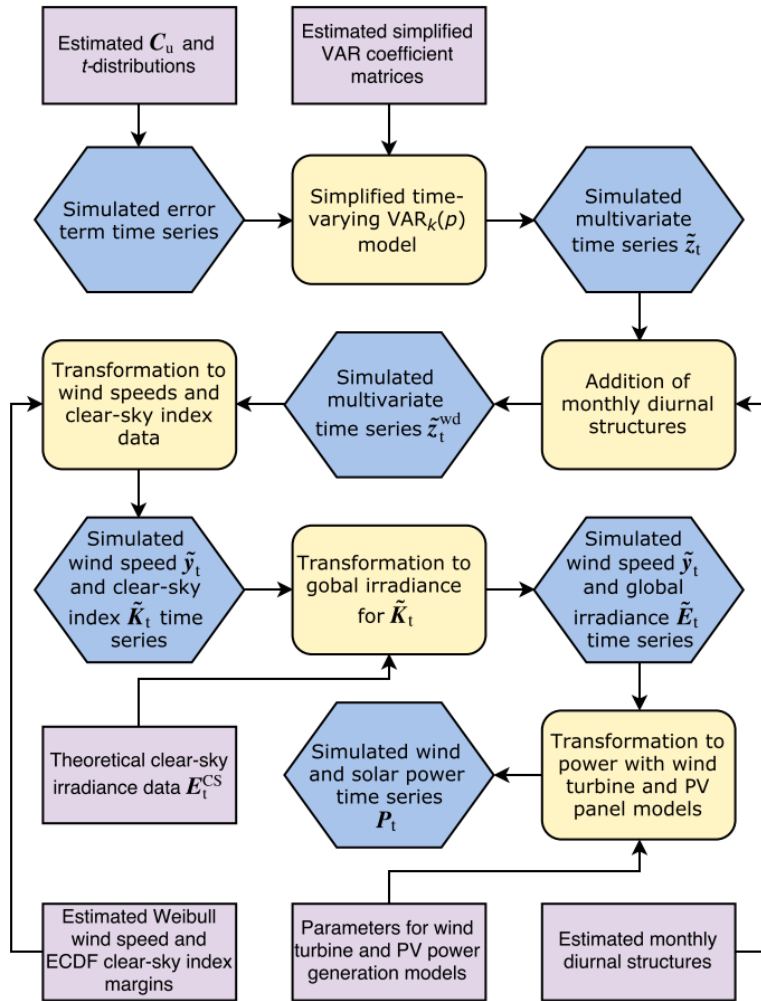
Publication VI proposes an approach for the modelling of wind direction in new WPP locations. The presented methodology is based on the wrapped Gaussian VAR process, where VAR model is the simplified VAR model (Section 3.2.2). In the proposed approach, the circular wind direction data is considered as an outcome of an underlying Gaussian process and a wrapping operation. This approach is taken, as time series models such as the VAR model, are not applicable to circular data.

An additional simplification is made to the simplified VAR model so that the order of the model is fixed to  $p = 1$ . This stipulation was made as it allowed a straightforward estimation of the model parameters in the case of  $\text{VAR}_k(1)$  model, and as the model was still sufficient for the modelling of the underlying Gaussian process in the wind directions modelling.

The wind direction modelling can be used jointly with the wind speed modelling methodology to provide wind direction simulations, in addition to wind speeds, for new WPPs. The wind direction data allows the use of dynamic wake effect models instead of static model as in Section 3.3.4. The wind direction data for the estimation of the model is from the same high altitude wind speed dataset from Finland as used in Section 3.3. The wind direction modelling methodology is presented here in short and can be found in detail from Publication VI.

#### 3.6.1 Estimation and Simulation Procedure

The simplified  $\text{VAR}_k(1)$  model assumes that the intercept term  $\mathbf{c} = \mathbf{0}$ , and thus, the parameters required for the estimation are the simplified VAR coefficient matrix  $\mathbf{A}_1$  (3.8) and the covariance matrix  $\Sigma_u$  of the errors  $\mathbf{u}_t$ . The error term  $\mathbf{u}_t$  is specified to follow multivariate normal distribution, and thus, making the



**Figure 3.10.** Flowchart of the MC simulation process for new WPP and PVP locations with the simplified time-varying VAR model based approach. Yellow boxes are operations, blue are simulated data and purple are estimated parameters.

process Gaussian as shown in Publication VI.

As the underlying Gaussian process, modelled with the simplified  $\text{VAR}_k(1)$  model, cannot be measured directly, the wind direction measurements  $\omega_t = [\omega_{1,t}, \omega_{2,t}, \dots, \omega_{k,t}]'$  are used to estimate the process. However, standard measures of correlation, such as Pearson's correlation, are not applicable for circular data. Therefore, circular correlation  $\rho_T$  [92] is used as a measure of correlation for the wind directions.

First, circular correlations  $\rho_T$  are estimated from the measured  $\omega_t$ , including both temporal (ACFs at lag  $h = 1$ ) and spatial correlations between the loca-

tions at  $h = 0$ . These circular correlations are then transformed to Pearson's correlations  $\rho$  following the approach presented in detail in Publication VI, and as specified in [92]. The obtained Pearson's correlations can be used in the estimation of the  $\text{VAR}_k(1)$  model parameters to specify the underlying Gaussian process. As the order of the model is 1, the diagonal components of  $\mathbf{A}_1$  are the estimated ACF values (temporal correlations) at lag  $h = 1$  [93].

Covariance matrix of the errors  $\Sigma_u$ , on the other hand, can be estimated with the simplified version of (3.11), where  $\Gamma_z(0) = \Sigma_z$  (because  $p = 1$ ). The required covariance matrix  $\Sigma_z$  of the underlying Gaussian process is calculated using the estimated spatial Pearson's correlations  $\rho_{i,j}$  and transformation (3.18), where the needed standard deviations  $\sigma_i$  are specified so that  $\sigma_i^2 = 3$ , as justified in Publication VI.

When simulating from the underlying Gaussian process, the resulting normally distributed  $\tilde{\mathbf{z}}_t$  is transformed to simulated circular data after first adding a mean vector  $\boldsymbol{\mu} = [\mu_1, \dots, \mu_k]'$  which specifies the prevailing wind direction, i.e., the mean angle, for each location (average of the means of the estimation locations is used for new WPPs). The transformation is done by using a wrapping operation as follows

$$\omega_{i,t} = \tilde{z}_{i,t} [\text{mod}(2\pi)], \quad (3.26)$$

where  $\omega_{i,t}$  is the simulated circular data, restricted between  $0 \leq \omega_{i,t} < 2\pi$ , in location  $i$  at time  $t$  and  $\text{mod}$  is the modulo operator [94, 95].

The addition of the mean wind direction with  $\boldsymbol{\mu}$  before the transformation is not enough to specify proper margins for the simulated wind direction data in new WPPs. Thus,  $\omega_t$  is transformed further with the following transformation

$$\tilde{\omega}_{i,t} = \hat{F}_i^{-1} [\hat{F}_{sim}(\omega_{i,t})], \quad (3.27)$$

where  $\hat{F}_i^{-1}$  is the inverse of the estimated ECDF margin for location  $i$  and  $\hat{F}_{sim}$  is the estimated margin of the simulated  $\omega_t$ . Average ECDFs of the estimation locations are used for  $\hat{F}_i$ , however, also ECDFs from the closest measured location could be used.

### 3.6.2 Dynamic Wake Effect Model

Simulated wind direction data  $\omega_t$  can be used for the more accurate modelling of the wake effect in new WPPs instead of a static wake coefficient as discussed in Section 3.3.4. Simulated wind direction data allows the use of dynamic wake effect models, which estimate a wake coefficient for each  $t$  depending on the current wind direction at  $t$ .

The wind directions are divided into eight  $45^\circ$  sectors and each sector yields a specific effective wind farm topology (depending on the wind direction and the topology of the WPP). For different effective topologies, the individual turbines align differently in relation to each other, assuming that the turbines follow the wind direction without a notable lag.

Different effective topologies result in different effective wind speeds experienced by individual turbines as each upstream turbine draws energy from the wind, and thus, decreases the downstream wind speed. The effective wind speed experienced by any individual turbine is estimated using the single wake model proposed in [50], although, the effective wind speeds can be estimated using various different approaches. The approach is presented in more detail in Publication VI.

The dynamic wake coefficient for a single turbine can be estimated from the ratio between the effective wind speed for that turbine at  $t$  and the ideal free wind speed provided by the wind speed simulation model at  $t$ . The wake coefficient for a WPP at  $t$  can be calculated as the average of the wake coefficients of the individual turbines in that WPP at  $t$ .

### 3.7 Regional Wind Generation Modelling

Publication VII took a somewhat different approach for the modelling of wind power generation compared with Publications I–VI. In Publication VII aggregated generation in geographical areas (Nord Pool Spot bidding areas [96]) were modelled as a whole instead of individual WPPs. This is useful when information of individual WPPs is not available or if more crude modelling is sufficient.

The modelling is done using a full VAR model as shown in (3.7), however, as the proposed approach deals with existing geographical areas, complex parameter estimation for new locations as presented in Section 3.2.3 is not required. The model estimation is done using measured hourly wind generation data from Denmark, Estonia, Finland and Sweden, presented in detail in Publication VII.

The installed capacity in larger geographical areas change in time as new capacity is being constantly installed. Consequently, the power generation time series are not stationary as the average generation and its volatility changes in time, and thus, not applicable for the estimation of a VAR model [46]. Therefore, the modelling approach utilizes the proportion of installed capacity (PIC) data for the model estimation to obtain stationary time series. The PIC data can be calculated as

$$p_{i,t}^{\text{pic}} = \frac{P_{i,t}}{\text{cap}_{i,t}}, \quad (3.28)$$

where  $P_{i,t}$  is the average hourly power generation and  $\text{cap}_{i,t}$  is the installed generation capacity in area  $i$  for time  $t$ . The PIC values are always between one and zero and were found to be stationary for all areas. For example, augmented Dickey-Fuller test [83, 90] can be used to ensure that the PIC time series are stationary.

Next, the PIC distributed data is transformed to normally distributed data  $z_t$  with probability integral transformations, similarly as with the WPP or PVP modelling in Sections 3.3.1 and 3.4.1. The estimated ECDFs combined with fitted

GP tails [32, 33] are used as the margins of the PIC data in the transformation.

The spatial and temporal dependency structures are modelled with a  $\text{VAR}_g(5)$  model (as there were eight geographical areas included).  $p = 5$  was found to be sufficient and the adequacy of the model was assessed as in Section 3.3.2. The monthly changing diurnal variations are modelled with time-varying intercept term  $\mathbf{c}_t$  with different values for every hour of every month, similarly as with the monthly diurnal structures with WPP and PVP modelling in Sections 3.3.1, 3.4.1 and 3.5.2. The error terms of the model are modelled with a Gaussian copula [79] with  $t$ -distributed margins as with the WPP and combined WPP and PVP modelling in Sections 3.3 and 3.5.

The simulation procedure follows a similar structure as in Sections 3.3.5 and 3.5.4. A random sample is drawn from the Gaussian copula, the margins are transformed to  $t$ -distributions and then the random sample is used as innovations for the  $\text{VAR}_g(5)$  model. The only exception is that the monthly changing diurnal structures are added through  $\mathbf{c}_t$  in the simulation of  $\tilde{\mathbf{z}}_t$  with the VAR model instead of adding them after the simulation.

Then, the simulated  $\tilde{\mathbf{z}}_t$  is transformed to PIC data using a similar transformation as in, e.g., (3.27), with specific PIC margins estimated for each area. As with the PIC margins, ECDFs combined with GP tails are used also to estimate the margins of the simulated  $\tilde{\mathbf{z}}_t$ . It should be noted that other margins instead of the the original PIC margin for the considered location can be used when simulating future scenarios as the increased capacity changes the distribution, as discussed in Publication VII.

Last, the simulated PIC data  $\hat{\mathbf{p}}_t^{\text{pic}}$  (with proper margins describing the simulated case) is transformed back to aggregated generated power for each area with the installed capacities used in the simulated scenario.

## 4. Simulation Results

This chapter presents a general overview of the results obtained in Publications I–VII. The first part presents the main results for the validation of the developed modelling methods for long term MC simulations. These results are divided into three sections by the main application of the methodology, i.e., new WPP locations, new PVP locations and the combined VRE modelling. The second part presents the results related to the effects of the geographical distribution of the WPP and PVP locations, which has been a prevailing topic of study in Publications I–VII.

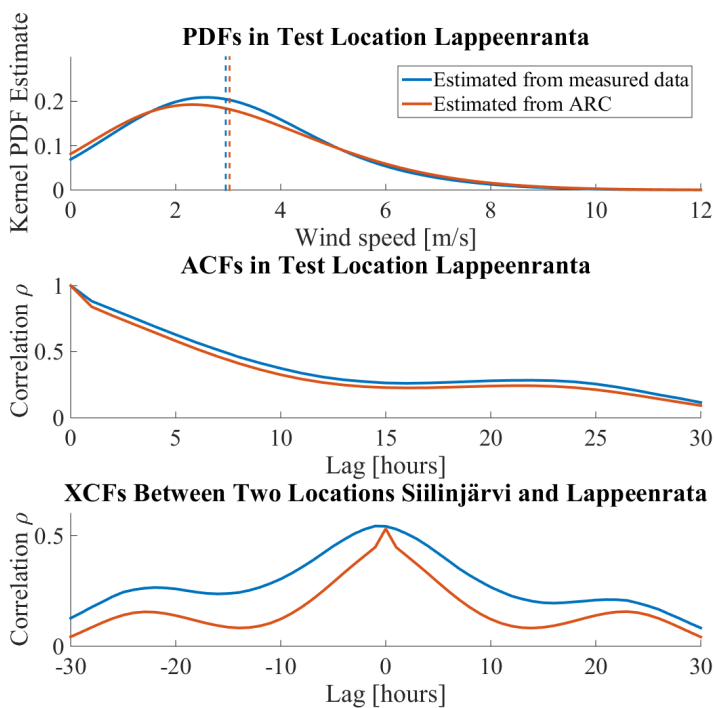
### 4.1 New Wind Power Plant Locations

This section focuses on the overview of the results obtained for WPPs in Publications I–III, where the focus was mainly on the modelling of WPP locations without measurement data available from the locations, i.e. modelling of new (non-measured) locations. The modelling methodology was developed further in each of the three consecutive publications. Next, the key results for the two main modelling approaches based on the ARC model and on the full VAR model are presented.

The aim of Publication I was to model individual new wind speed locations. The proposed modelling methodology was based on the ARC model, which was described in Section 3.2.1. This modelling approach is the first of the developed methods, and thus, also the least advanced with simplifications such as the use of normally distributed model errors. The model was estimated with the low altitude wind speed data from 19 locations in Finland, as presented in Section 3.3.1.

The key simulation results from 100 yearly MC runs (8760 hours  $\times$  100 runs =  $8.76 \times 10^5$  samples) for two out-of-sample wind speed locations (located in Finland and measured approximately 10–15 meters above the surrounding ground level) are illustrated in Figure 4.1. According to Figure 4.1 the estimated probability distribution function (PDF) and average wind speed in test location Lappeenranta were similar to the ones estimated from the data. Furthermore,

ARC based approach was able to produce ACF with a similar shape to the ACF estimated from the measurements. However, the estimated XCF between test locations Siilinjärvi and Lappeenranta had a peaked shape near lag  $h = 0$  instead of the round shape of the XCF calculated from the data. The value of the XCF at lag  $h = 0$  is modelled correctly, but the slope of the XCF around lag  $h = 0$  is too steep and the values with other lags are notably too low. This is due the lack of information regarding the spatial correlations with other lags (spatial information is provided by the correlation matrix, which contains only  $h = 0$  values). Despite the shape of the XCF, good accuracy is also obtained in the simulation results for the probabilities for events when wind speeds are high or low simultaneously in several simulated locations in Publication I.



**Figure 4.1.** The PDFs and ACFs estimated from simulated and measured low altitude wind speed data from test location Lappeenranta and the XCFs between test locations Siilinjärvi and Lappeenranta. The dashed lines in PDF sub-figure illustrate the corresponding hourly mean wind speeds. The simulation results are averages of 100 MC runs.

For the modelling of new individual wind speed locations, the wrong shape of the XCF is not an issue. However, for the modelling of the aggregated generation of multiple WPPs, the wrong shape of the XCF causes problems with the aggregated ACF and ramp rates. This is because the ACF of an aggregate of multiple time series depend on the XCFs between its components,

as demonstrated in Publication III.

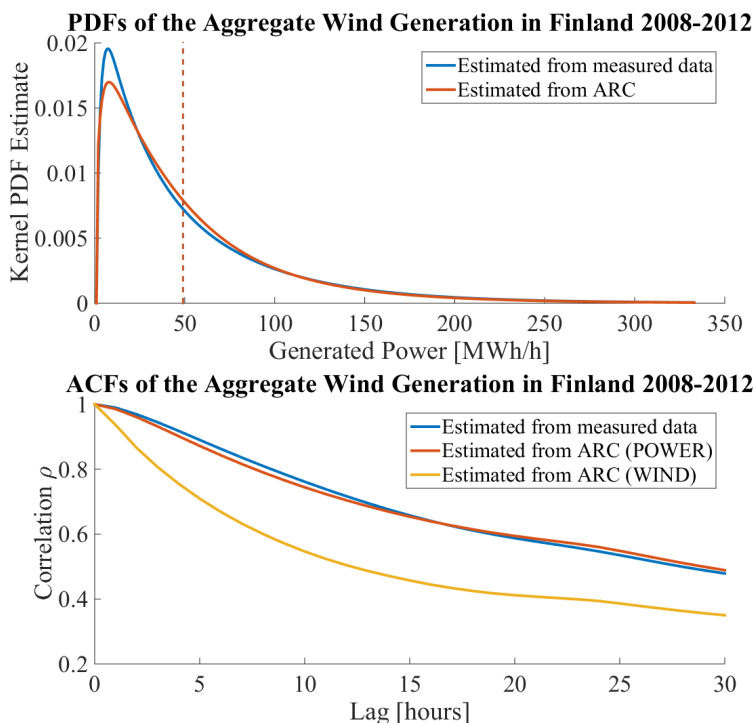
In Publication II, the aim was to apply the ARC based methodology to the modelling of aggregated wind power generation of multiple WPPs. The ARC model was combined with the general WPP modelling methodology presented in Section 3.3. The aggregated wind power generation structure of Finland from 2008 to 2012 was modelled in detail, including the changing number of installed wind turbines (212 in the end of 2012) as new WPP locations. The ARC model based approach was developed further by using, e.g.,  $t$ -distributed margins for the model errors and a more advanced power generation model including third degree power curve and taking into account the wake effect inside a wind farm, as discussed in Section 3.3.4.

As discussed earlier, the limitations of the ARC model in the modelling of the XCFs between individual locations result in an ACF with too low values for the aggregated power generation (when using AR coefficients estimated from (high altitude) wind speed measurement data). In Publication II, the issue was solved by using AR coefficients estimated from the measured aggregated wind power generation from Finland in 2012 instead of wind speed measurements.

Figure 4.2 presents the ACFs estimated from 100 MC simulation runs for the aggregated wind generation in Finland from 2008 to 2012. It is clear that the standard approach using coefficients estimated from wind speed data resulted in too low ACF values, but the coefficients estimated from power generation data produced similar ACF shape for the simulations and the measurement data. Both approaches, however, resulted in correct PDF shapes and mean hourly power generations as shown in Figure 4.2. When the coefficients were estimated from the power data, also the numerical results assessing the accuracy of the methodology were good in terms of various numerical statistics, such as, mean hourly generation, power ratios, yearly generated energies and standard deviations of the hourly generation as presented in detail in Publication II.

However, the proposed solution doesn't come without problems. If the AR coefficients are estimated from aggregated power generation instead of measured wind speeds, the individual locations have wrong ACFs (as they have now the temporal structure of the aggregated wind generation of a system). Consequently, this approach is only valid for the analysis of aggregated generation and cannot be used for the analysis of the individual WPPs. In addition, if the generation structure in the modelled scenario would differ notably from the structure of the system used for the AR coefficient estimation, the results could be inaccurate.

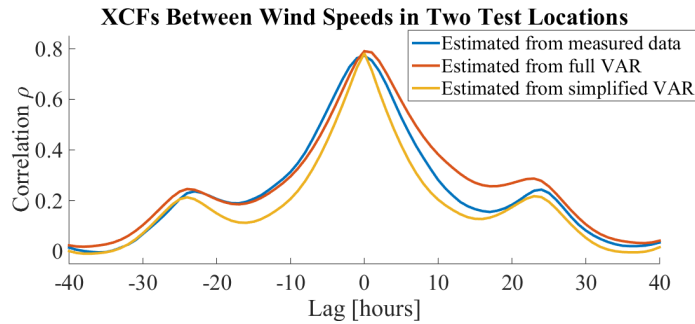
Publication III proposed the full VAR model based methodology and an approach to estimate the model parameters for scenarios with new WPP locations. This approach allows the detailed modelling of the XCFs between individual WPPs, while estimating the model only using wind speed data. Consequently, the approach also yields correct shapes for the aggregated ACF and distributions of the aggregated power ramp rates. Thus, the full VAR based approach solves the problems related to the ARC based modelling, as it performs as good as or better in all aspects compared with the ARC based approaches.



**Figure 4.2.** The PDFs and ACFs of the simulated and measured aggregated wind power generation data in Finland from 2008 to 2012. The dashed lines in PDF sub-figure illustrate the corresponding hourly mean power generation (the lines are overlapping). In the ACF sub-figure POWER indicated that the AR coefficients are estimated from the aggregated power generation and WIND that the coefficients are estimated from the high altitude wind speed measurements. The simulation results are averages of 100 MC runs.

The full VAR model was estimated with high altitude wind speed locations (except the spatial and spatiotemporal correlations were estimated with low altitude wind speed data) and the methodology followed the approach presented in Section 3.3. For the power generation, a state-of-the-art approach based on logistic functions was utilized.

Figure 4.3 presents the XCFs estimated from 100 MC runs for two out-of-sample high altitude wind speed locations in Finland compared with measurement data. In addition, Figure 4.3 illustrates also XCFs calculated from data simulated with a simplified VAR model. The simplified VAR model the model used with the combined VRE modelling in Publication V. It can be considered as an improvement for the ARC based approaches, but it is not able to model the shape of the XCF properly. As visible, the XCF modelled with the full VAR model is able to capture the correct round shape of the function near lag  $h = 0$ , which is not the case with the simplified VAR or ARC models.

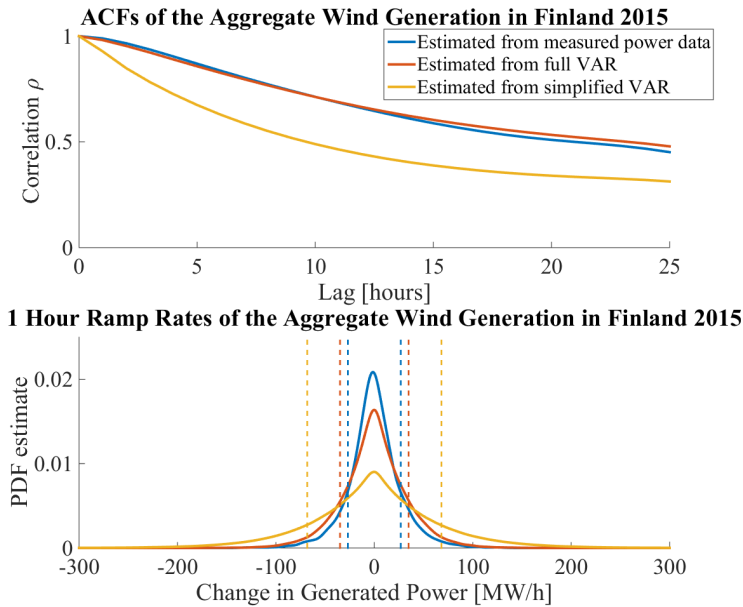


**Figure 4.3.** The XCFs estimated from data simulated with full and simplified VAR models and from the measured high altitude wind speed data for two test locations. The simulation results are averages of 100 MC runs.

Figure 4.4 illustrates the main MC simulation results from 100 runs for the aggregated wind power generation in Finland in 2015. The simulations are an out-of-sample test, i.e., aggregated measured power time series or any measurement data from the individual WPPs weren't used in the model estimation. Figure 4.4 illustrates the ACFs estimated from the 2015 power generation data and with the full and simplified VAR models. The full VAR model succeeds to capture the correct shape of the ACF with good accuracy. This is a significant improvement, as the full VAR approach obtains excellent results in a complete out-of-sample test. The model performance and accuracy were also good in the out-of-sample test when measured with various numerical statistics as displayed in detail in Publication III.

As the full VAR model succeeds in the modelling of XCFs between individual WPPs, it can be utilized in the modelling of aggregated up and down power ramps of the WPPs. Figure 4.4 presents the estimated one hour ramp rate PDFs and the standard deviations for the aggregated wind generation in Finland in 2015. The full VAR model is able to produce ramp rate PDFs and standard deviations close to the ones estimated from the data (it should be noted that some differences are expected as this is a complete out-of-sample test).

To conclude, the full VAR model performs well with all aspects of long term modelling of WPPs, and can be considered superior to ARC and simplified VAR based approaches. It captures the temporal and spatial dependency structures accurately for both individual WPP locations and for the aggregated generation, and thus, also allows the modelling and analysis of aggregated power ramps with new WPPs. Consequently, the methodology has a wide applicability in the assessment of future VRE scenarios and, e.g., the estimation of required amount of balancing power or need for new grid investments when new WPPs are added to the system, and thus, is one of the main contributions of this work to the research field.



**Figure 4.4.** The ACFs estimated from data simulated with full and simplified VAR models and from the measured aggregated wind power generation data from Finland in 2015. The one hour ramp rate PDFs and standard deviations estimated from data simulated with full and simplified VAR models and from the measured aggregated wind power generation data from Finland in 2015. The simulation results are averages of 100 MC runs.

## 4.2 New Solar Power Plant Locations

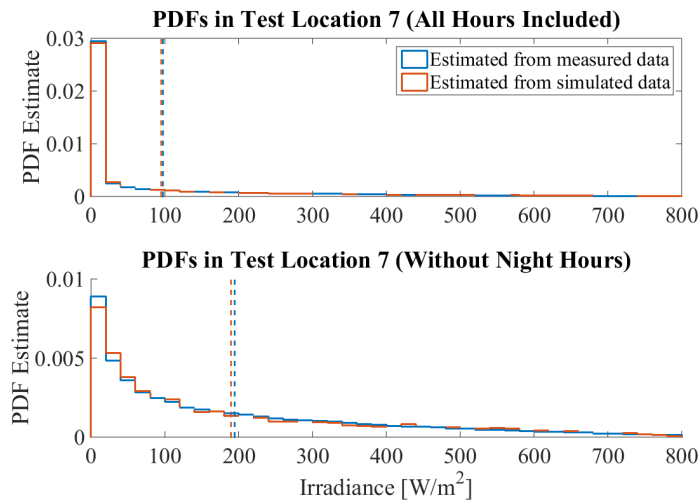
This section presents an overview of the results obtained in Publication IV. In Publication IV, a time-varying ARC model was utilized in the long term PVP modelling in new generation locations. The modelling methodology followed the approach presented in Section 3.4.

The long term MC simulation results for solar (global) irradiance are presented in Figures 4.5 and 4.6. The results are averages from 100 MC runs for two out-of-sample irradiance test locations. Figure 4.5 illustrates the PDFs and the mean values estimated from the measured and simulated data for one of the test locations in two cases. First, with all hours included, and second, for the case where only hours when the sun is above the horizon are considered, i.e., when irradiance can be larger than zero. As visible, the simulation methodology performs well in both cases.

Figure 4.6 shows the ACFs estimated for one of the test locations from the simulated and measured data. As visible from the figure, the methodology is able to capture the shape of the ACF accurately. In addition, Figure 4.6 illustrates the cyclical ACF structure of irradiance, which is caused by the daily diurnal cycle dependent on the geographical location.

Figure 4.6 presents also the XCFs between the test locations from simulated and measured data. Again, the cyclical structure is dominant in the shape of the XCF. According to Figure 4.6, the time-varying ARC based methodology is able to produce simulated data with similar XCF shape compared with the XCF estimated from the measurements. Hence, unlike with WPP locations, the ARC based methodology is able to capture all of the relevant dependency structures of solar irradiance in new PVP locations. This is also supported by the good accuracy of the simulation results in wide range of numerical statistics assessing the model performance in Publication IV.

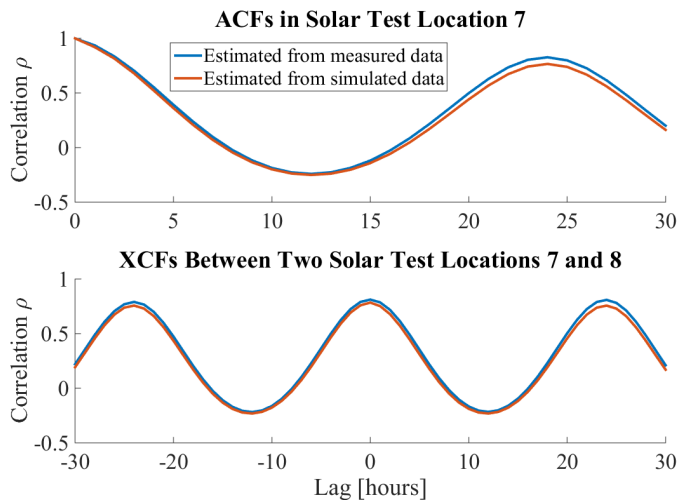
To summarize, the PVP modelling methodology yield accurate results in the out-of-sample simulation test, and thus, can be considered applicable in the modelling of solar irradiance in new PVP locations. The methodology has a wide applicability from the modelling of individual buildings to aggregated solar generation in a country and can be utilized, e.g., by the DSOs to analyse future scenarios with large installed capacities of consumer PV panels in their systems, and consequently, to plan the required grid investments for their networks. Hence, it is another of the main contributions of this thesis to the research field, as it allows detailed simulation of new PVPs in future scenarios.



**Figure 4.5.** The PDFs and the average global irradiance values estimated from simulated and measured global irradiance data from one of the test locations. The dashed lines illustrate the corresponding hourly mean global irradiances. The simulation results are averages of 100 MC runs.

### 4.3 Combined New Wind and Solar Power Plant Locations

This section presents an overview of the results obtained in Publication V for the combined modelling of WPPs and PVPs. In Publication V, simplified time-



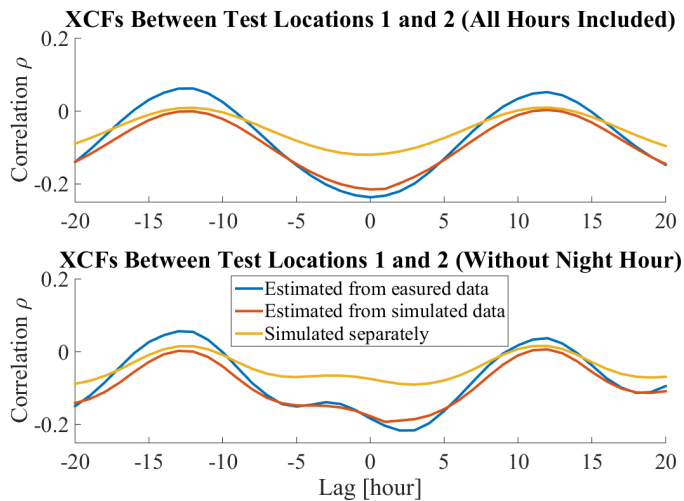
**Figure 4.6.** The ACFs estimated from simulated and measured solar irradiance data from test location 7 and the XCFs estimated from simulated and measured solar irradiance data for test location 7 and 8. The simulation results are averages of 100 MC runs.

varying VAR model based methodology, introduced in Section 3.5, was proposed for the joint modelling of WPPs and PVPs. The aim of Publication V was to model the effect of the spatio-temporal dependencies between WPPs and PVPs and implement it to the modelling of new generation locations. The approach combined the ARC model based methodologies proposed in Publications II and IV into a simplified time-varying VAR based methodology and added the modelling of the spatio-temporal correlations between the WPPs and PVPs. Consequently, the impact of the negative correlations between WPPs and PVPs, as discussed in Section 3.5.3, were implemented in the modelling methodology.

The combined modelling approach was verified by analysing 1000 MC runs for two out-of-sample test locations, one wind speed and one solar solar irradiance measurement location, both located in Finland. Figure 4.7 illustrates the XCFs between the test locations both with the night time hours (when sun is below the horizon and the solar irradiance is always zero) included in and excluded from the estimation of the XCFs. Figure 4.7 also presents the simulation results obtained with the ARC model approaches utilized in Publications II and IV used separately. In both cases, the combined modelling approach produces correlation structures, which are close to the structures found from the measurement data. Similar findings were observed in Publication V when the accuracy of the simulation results was assessed with numerical statistics as the combined modelling and the separate simulations both obtained good accuracy in all other measures except the spatial correlation between the test locations, in which only combined modelling was accurate. Thus, it is clear that the separate simulations are unable to produce proper correlation structures, as the models lack the information concerning the dependencies between wind speeds and

solar irradiances.

To conclude, the combined WPP and PVP modelling approach is crucial when analysing energy systems or future scenarios with both generation types included, as the negative correlations between the WPPs and PVPs have a significant impact on the shape of the XCF. Consequently, these dependency structures have also a notable effect on the aggregated VRE generation and its variation. The joint modelling can be applied, e.g., by power producers, who have both WPPs and PVPs in their portfolio, and TSOs to analyse the total variations of the VRE generation in their system or to assess the effects of new WPPs or PVPs on the aggregated generation. Thus, the combined modelling approach is the third main contribution of this thesis, as it allows detailed simulation of future VRE generation.



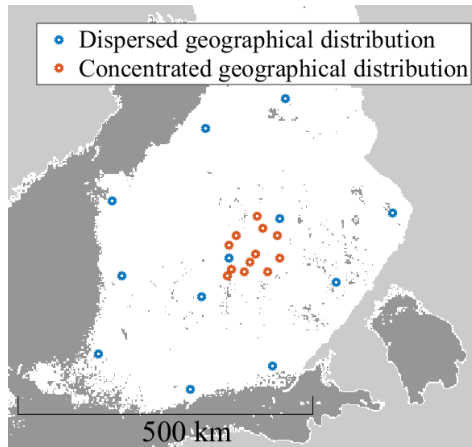
**Figure 4.7.** The XCFs estimated from data simulated with the combined WPP and PVP modelling using the simplified VAR model, from the separate simulations with the ARC based models and from the measured wind speed and solar irradiance data from one wind and one solar test locations. In the lower figure, the night time hours, when the sun is below the horizon, are omitted. The simulation results are averages of 1000 MC runs.

#### 4.4 Effect of Geographical Distribution of Generation Locations

This section presents an overview of the MC simulation results obtained in Publications I–VII considering the geographical distribution of the WPP and PVP locations. The effect of the geographical distribution of the WPPs, PVPs or the combination of both to the generated energy and its variability has been a prevailing topic of the case studies and scenario analyses in Publications I–VII. Quantifying the effect of the geographical distribution of the VRE generation through various studies and analyses is also the fourth main contribution of this

work to the research field.

In most of the publications, two cases, a dispersed and a concentrated distribution of the generation locations have been analysed. The length of the MC runs in each of the presented cases is one year. An example of these geographical distributions in Southern Finland are illustrated in Figure 4.8.



**Figure 4.8.** Dispersed and concentrated geographical distribution of VRE generation locations in Southern Finland.

#### 4.4.1 Wind Power Plants

The effect of geographically dispersed or concentrated distribution of WPPs on the probabilities for very high or low generation events was first studied in Publication I with scenarios consisting of ten find farms. In Publication II, similar case studies were conducted for several future wind power generation scenarios in Finland. In these case studies, in addition to the existing installed wind generation in 2012, the planned generation from 2012 to 2017, and the Wind Europe's (formerly EWEA) target installed capacity for 2020 [97], were distributed according to the actual planned WPP sites and with a concentrated and dispersed distribution of the WPPs.

Publication III also assessed the impact of the geographical distribution of WPPs with the full VAR approach combined with a more accurate power generation model. An example of the differences in the PDFs of the aggregated mean hourly generation with geographically dispersed and concentrated distribution is illustrated in Figure 4.9. These example cases consists of 12 WPPs with 20 Gamesa G128 4.5 MW turbines [87] in each, and thus, with the total installed generation capacity of 1080 MW in both cases. The geographical distributions in the example cases are the ones presented in Figure 4.8.

The analyses of the results from Publications I–III concluded that the dispersed distribution resulted in a smaller number of extreme events, where the

aggregated generation is very high or low, and also in smaller variance of the hourly average generation compared with more concentrated distribution of the WPPs. All of the related results obtained in Publications I–III were in line with each other. Similarly, reduced variability of the generation was also observed with reanalysis datasets with and actual measurements from multiple existing WPPs in [98,99].

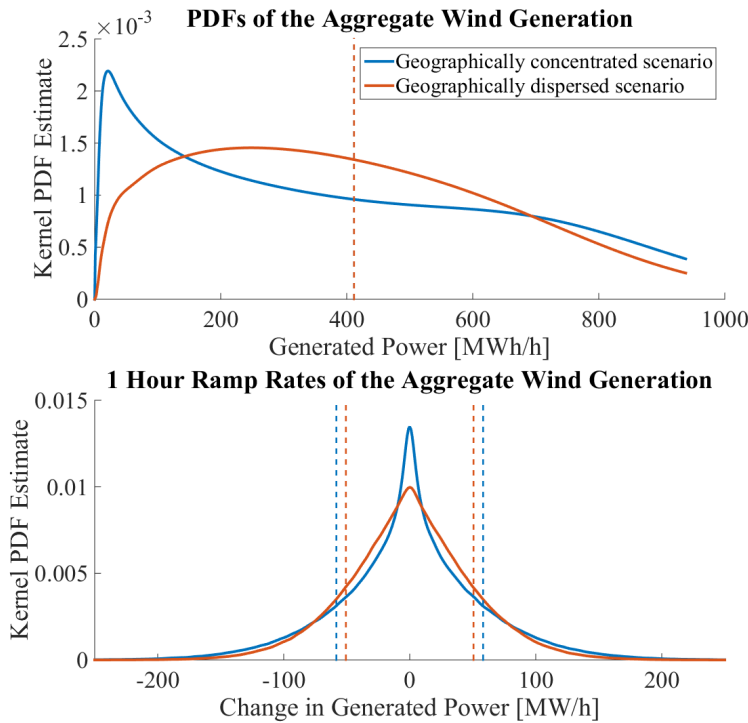
In Publication III, the proposed methodology also allowed the assessment of the effect of geographical distribution of the WPPs on the aggregated wind power ramp rates. The case study showed that the volatility of the ramp rates was reduced with dispersed placement of the WPPs with both the up and down ramp rates despite the length of the ramps. Similar results were observed with measured existing WPPs in [99]. In addition, the differences were found to increase longer the ramps were. Figure 4.9 presents an example of one hour ramp rate PDFs of aggregated wind generation with both geographical distributions in the example case described earlier. These kind of analyses can be applied, e.g., by TSOs to assess the volatility and ramp rates of the aggregated VRE generation to estimate the need for, e.g., balancing power and demand response in the system to compensate the variations and power ramps.

Publication VI studied the impact of the wind direction modelling in individual WPPs on the aggregated power generation in geographically concentrated and dispersed scenarios, average distance between WPPs 67.9 km and 499 km respectively. Both scenarios consisted of 12 WPPs with 36 Vestas V112 3.3 MW turbines [100] in each (more details in Publication VI) and the wind speed simulations for the WPPs were done with the ARC based approach as in Publication II.

Figure 4.10 illustrates the impact of the wake effect modelling, and hence, the modelling of wind direction in individual WPPs, on the PDFs of the aggregated power generation with the two geographical distributions. As shown in Figure 4.10, the dynamic modelling of wake effect (i.e. the wake effect at time  $t$  depends on the simulated wind speed at  $t$ ) has a notable effect on the PDF of the aggregated generation near the maximum with the concentrated, but not with the dispersed distribution of the WPPs. This is because with the concentrated scenario, the aggregated generation reaches the maximum more often, and with the dynamic (i.e., more detailed) wake effect modelling the maximum output of a WPP depends on the current wind direction at time  $t$  as the wind turbines align differently in relation to each other depending on the wind direction.

Consequently, the dynamic wake effect modelling assesses the maximum generation events, which are more common with the concentrated scenario, with greater detail compared with the static wake effect, where the average wake effect specifies the same possible maximum output for a WPP for all  $t$ . Furthermore, it can be said that the wake effect has greater effect to the aggregated generation with concentrated geographical distribution of the WPPs.

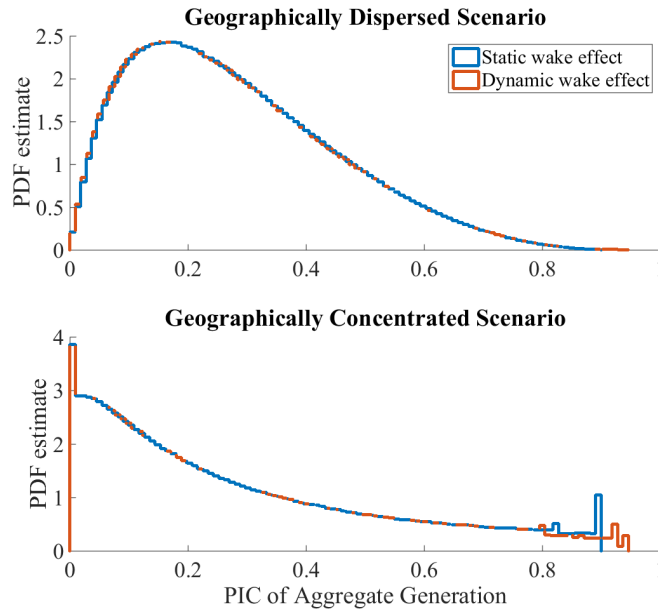
In addition, the impact of geographical distribution of wind generation was studied also in regional level for the Nord Pool Spot bidding areas in Denmark,



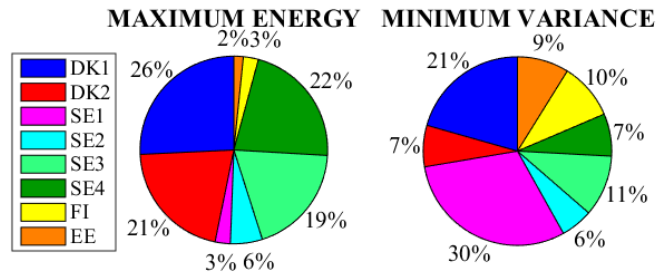
**Figure 4.9.** The PDFs for the aggregated power generation and one hour ramp rates estimated from simulated data with concentrated and dispersed geographical distribution using the full VAR approach. In the upper sub-figure, the dotted lines illustrate the mean hourly power generation and are overlapping. In the lower sub-figure, the dotted lines show the standard deviations of the one hour ramp rates. The simulation results are averages of 100 MC runs.

Sweden, Finland and Estonia in Publication VII. The conducted case study showed that depending on whether to optimize to maximize the yearly energy generation or to minimize the volatility of the hourly generation, the optimal placement of new capacity, and thus, the geographical distribution of the total installed generation capacity was notably different between the bidding areas.

When maximizing the energy output, the new WPPs were placed to areas with the best generation conditions for the WPPs, i.e., to Denmark and southern Sweden. However, when minimizing the volatility of the aggregated generation, new WPPs were assigned to areas with low initial installed capacity to balance the volatility of the whole system. The results are shown for the bidding areas in Figure 4.11, and are also in line with the ones obtained with the new WPP modelling approaches. When it comes to applications, these kind of evaluations and assessments can be beneficial for policy makers when designing, e.g., subsidies for renewables or when making policy decisions at, e.g., the European Union level.



**Figure 4.10.** The PDFs of the aggregated hourly wind generation for the geographically dispersed and concentrated scenarios using the static (same wake effect for all  $t$ ) and dynamic (the wake effect at time  $t$  depends on the simulated wind direction at  $t$ ) wake effect modelling with the WPPs. The aggregated generation is displayed as PIC values (proportion of the installed capacity). The simulation results are averages of 1000 MC runs.



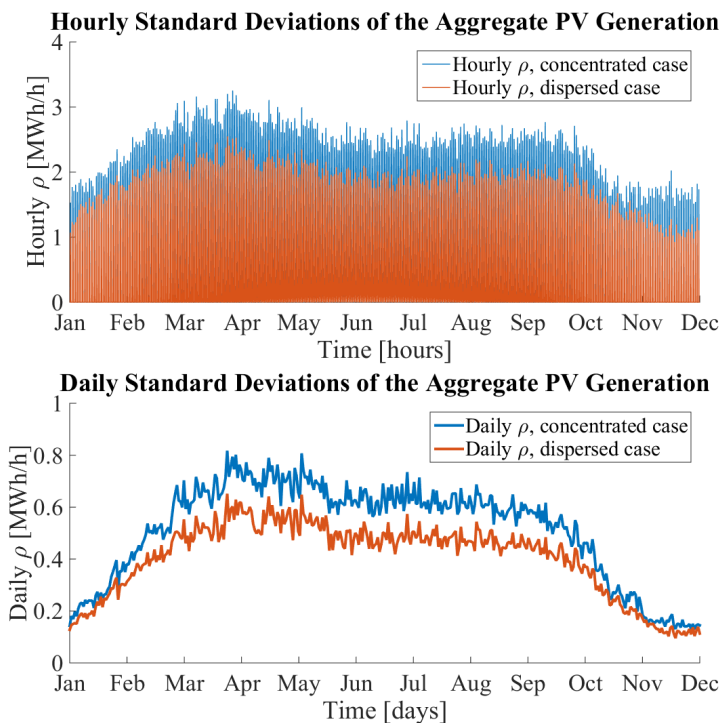
**Figure 4.11.** The shares of total installed generation capacity between the Nord Pool Spot bidding areas in two scenarios where the total generated energy is maximized or the variance is minimized. DK indicates bidding areas in Denmark, SE in Sweden, FI in Finland and EE in Estonia.

#### 4.4.2 Photovoltaic Power Plants

Publication IV presented a case study analysing the effects of the geographical distribution of PVPs on the variability of the aggregated PV generation. Figure 4.12 presents hourly and daily standard deviations for the aggregated PV gener-

ation in geographically dispersed and concentrated cases, as in Figure 4.8, both consisting of 12 PVPs with the capacity of 1 MW in each, depicting aggregates of consumer PV systems in the surrounding areas (more details can be found in Publication IV).

As visible from Figure 4.12, also with the PVPs the dispersed geographical distribution yielded lower variability of the aggregated generation. This is caused by the smoothing effect, (the changes in cloudiness smooths over large geographical areas) which reduces the volatility of the aggregated generation, as also discussed in [101]. However, it should be noted that with PVPs the impact of the geographical distribution to the volatility of the generation was found to be notably smaller, at least in the case studies of Finland with hourly time resolution, compared with the impact observed with WPPs.



**Figure 4.12.** The hourly and daily standard deviations of the aggregated PVP generation with geographically dispersed and concentrated distribution. The simulation results are averages of 100 MC runs.

#### 4.4.3 Wind and Photovoltaic Power Plants

In Publication V, the effect of the geographical distribution of VRE generation locations was studied for a system with both WPPs and PVPs in it. The optimal

ratio of the installed capacity between 12 WPPs and 12 PVPs was analysed with a case study consisting of 12 scenarios, presented in Table 4.1., using the geographical distributions presented in Figure 4.8. (Additional details about the setup of the case study can be found from Publication V.)

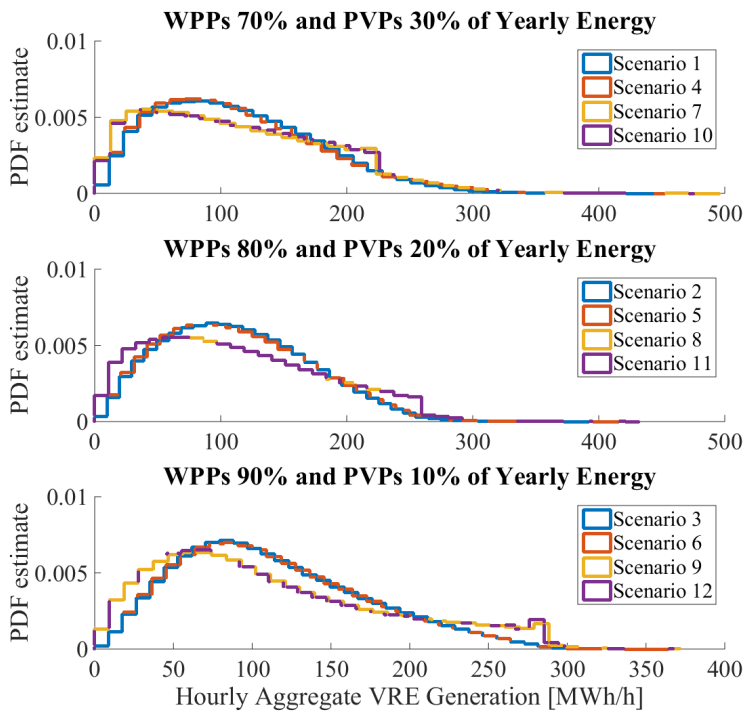
The hourly variability was minimized in cases where the WPPs and PVPs had both dispersed geographical distribution (Scenarios 4, 5 and 6). However, the difference with the cases where PVPs were concentrated was minor. This is in line with the previous observations that the distribution of WPPs has greater effect on hourly variations compared with the distribution of PVPs.

The ratio of the installed capacities to minimize to volatility was found in Scenario 5, where WPPs produced 80% and PVPs 20% of the annual generated energy. In terms of proportional installed capacities, the optimal ratio was 60% of the total installed VRE capacity in WPPs and 40% in PVPs as the power ratio was much higher with WPPs compared with PVPs. This result indicates that the negative correlations between WPPs and PVPs are so notable that they reduce the hourly volatility of the aggregated generation more than the higher volatility of PVPs compared with WPPs increases it, and thus, the volatility is reduced when both generation types are included in the generation mix. These kind of assessments and case studies can be valuable for the TSOs and policy makers to optimize the subsidies and other incentives so that large scale integration of VRE generation to future energy systems is possible. Figure 4.13 presents the PDFs of the aggregated VRE generation in the 12 scenarios.

To summarize, it was consistently observed in Publications I–VI that the dispersed geographical distribution of the generation locations is an effective way to decrease the hourly variability of the aggregated VRE generation.

**Table 4.1.** The specifications of the 12 VRE power generation scenarios with different geographical distributions and yearly energy proportions.

Scenario	WPP Geogr. Distribution	PVP Geogr. Distribution	WPP Yearly Energy (%)	PVP Yearly Energy (%)
Scenario 1	Dispersed	Concentr.	70	30
Scenario 2	Dispersed	Concentr.	80	20
Scenario 3	Dispersed	Concentr.	90	10
Scenario 4	Dispersed	Dispersed	70	30
Scenario 5	Dispersed	Dispersed	80	20
Scenario 6	Dispersed	Dispersed	90	10
Scenario 7	Concentr.	Concentr.	70	30
Scenario 8	Concentr.	Concentr.	80	20
Scenario 9	Concentr.	Concentr.	90	10
Scenario 10	Concentr.	Dispersed	70	30
Scenario 11	Concentr.	Dispersed	80	20
Scenario 12	Concentr.	Dispersed	90	10



**Figure 4.13.** The PDFs of the aggregated hourly generation from the 12 VRE scenarios. The plots with the same color have the same geographical distributions for the WPPs and PVPs. The simulation results are averages of 1000 MC runs.

# 5. Discussion and Conclusions

This chapter summarizes and discusses the key findings of this thesis including both the developed VRE modelling methods and the obtained long term simulation results, and answers to the formulated research questions. In addition, the practical applicability of the developed methods is considered and future work recommendations are suggested.

## 5.1 Discussion of the Modelling Methods and Simulation Results

### 5.1.1 Summary of the Findings

The main research question this thesis focused on answering was how VRE generation can be modelled in new (non-measured) generation locations to aid the large scale VRE integration to the smart energy systems of the future. The main research question was divided into two more detailed questions, the first focusing on how power generation and its volatility can be modelled in detail in new WPPs, PVPs and jointly with both generation types. The second more detailed research question concentrated on how the geographical distribution of the future large scale VRE generation affects the aggregated power generation and its variability, and consequently the need of balancing power. The main findings answering these research questions, and consequently, the main contributions to the research field, can be summarized as:

- Statistical methods for the modelling of new WPPs, including also the modelling of wind direction and changing generation capacity in large areas, were developed and verified with measurement data (Publications I, II, III, VI and VII).
- A statistical method for the modelling of new PVPs was developed and verified with measurement data (Publication IV).

- A statistical method for the joint modelling of WPPs and PVPs was developed and verified with measurement data (Publication V).
- The effect of the geographical distribution of the VRE generation locations to the aggregated generation was analysed and its impact quantified (Publications I-VII).

The first contribution in the above mentioned list of findings developed modelling methods for new wind power generation. This task was divided into three separate areas of the modelling. The first, and the most emphasized, was the detailed modelling of new WPPs. In Publications I–III statistical modelling methods for the analysis of new WPPs from a single turbine to a whole WPP generation in a country were developed and further improved and extended as presented in Section 3.3. A transformed ARC model was first developed in Publication I and then extended to a modular wind power generation modelling methodology and improved with more detailed error term modelling using  $t$ -distributions in Publication II. The methodology was further improved in Publication III by developing a transformed full VAR model based approach, which also enabled the analysis of aggregated wind power ramps in addition the assessment of the energy generation. A procedure for the full VAR model parameter estimation in the case of new WPPs was also derived and presented in Section 3.2.3. Using long term MC simulations, the methods were verified against out-of-sample data in Section 4.1 and applied in scenario analyses and case studies in Section 4.4.1.

The second focus area was the modelling of wind direction inside new WPPs to allow a more detailed analysis of the wake effect in WPPs. A statistical modelling methodology for wind direction modelling in new locations was developed in Publication VI to complement the WPP modelling methods. A transformed simplified VAR model based methodology was developed to model the circular wind direction data as presented in Section 3.6. An approach to estimate the model parameters from the circular data and to transform the simulated data back to wind direction data was also proposed. The methodology was verified using MC simulations against out-of-sample measurements and then applied in the dynamic wake effect modelling in new WPPs as shown in Section 4.4.1.

The last focus area related to wind power was the modelling of new wind capacity in large geographical areas instead of detailed modelling of new WPPs. A transformed VAR model with  $t$ -distributed errors was developed for this purpose in Publication VII as presented in Section 3.7. The methodology was used to model the wind power generation in large areas in the Nordic and Baltic power system without any knowledge of individual WPPs. It was shown that the model was able to simulate future scenarios with changing installed wind generation capacity in the studied areas. In addition, an approach to minimize the variance of the aggregated generation in the system composed of the modelled areas was proposed and demonstrated in Section 4.4.1.

The modelling of new PVPs was also one of the main goal of this work, as stated in the research questions. Hence, the methods developed for the modelling of new WPPs were further applied in the modelling of new PVPs in Publication IV. A methodology based on a transformed time-varying ARC model with  $t$ -distributed errors was developed to model new PVP locations from a PV system in a single building to an aggregated PVP generation in a country as presented in Section 3.4. MC simulations were used to verify the proposed methodology against out-of-sample data in Section 4.2 and apply it in case studies in Section 4.4.2.

The third main contribution of the this thesis to the research field was the model developed for the combined modelling of WPPs and PVPs. The separate methods developed for WPPs and PVPs were combined into joint modelling of VRE generation in new generation locations in Publication V. An approach utilizing a transformed simplified time-varying VAR model with  $t$ -distributed error terms was developed to capture the spatial dependency structures between wind and solar generation as presented in Section 3.5. The modelling methodology was verified against out-of-sample measurement locations using MC simulations in Section 4.3 and applied in case studies in Section 4.4.3. Furthermore, a notable negative correlation between WPP and PVP locations was observed, analysed and quantified (also in the underlying data with daily and monthly structures removed) as shown in Sections 3.5.3, 4.3 and 4.4.3, which is a notable contribution to the field.

Last, the second smaller research question considered the effect of the geographical distribution of the VRE generation to the aggregated power and its variability. Analysis of these effects has been a prevailing topic of the case studies and scenario analyses in Publications I–VII. The developed models were utilized in the study of these effects using long term MC simulations in various scenarios. The obtained results, presented in Section 4.4, were in-line quantifying how much dispersed geographical distribution of VRE generation locations decreased variations and volatility in the aggregated generation when compared with geographically concentrated distribution. The impact of the geographical distribution was observed to be notably larger with WPPs compared with PVPs in hourly time scale in Publication IV. In Publication III, it was analysed how much smaller the wind power ramps were with dispersed distribution of generation compared with the concentrated, and how much the difference increased with longer the ramp lengths. Furthermore, the analysis of combined VRE generation in Publication V and Section 4.4.3 the ratio of 80% of the annual energy generated with WPPs and 20% with PVPs, while both generation types having a dispersed geographical distribution, minimized the volatility of the aggregated VRE generation in terms of hourly generated energy. All these results can help quantifying the required amount of balancing power in the smart energy systems of the future with large VRE penetration.

In addition, it should be noted that the developed simulation models and modelling methodologies have certain limitations, which are related to the data used

in the model estimation. The models were estimated with measurement data obtained from Finland, and thus, can't be as such generalized for any geographical location or climate zone. However, provided the sufficient estimation data, e.g., wind speed and solar irradiance measurements from the geographical area of interest, the models can be estimated and should be applicable without problems. Furthermore, it should be noted that the same limitations of the generalizability also concern the simulation results on the effects of geographical distribution of VRE generation locations.

### **5.1.2 Practical Applications of the Work**

This section discusses the practical applications where the developed VRE simulation models and obtained results can be used to enable the transition to energy systems with high penetration of VRE sources. In addition, five beneficiary groups in both public and private sector, who can utilize the proposed methodologies and derived results in their operations or activities, were identified. Next, the practical implementations for each beneficiary group are discussed.

The TSOs can apply the models as an additional source of information to focus the grid investments more efficiently in long term grid planning, and as a way to estimate the effects of, and thus to enable, the increasing penetration of VRE generation in their systems. The models provide an accurate way to analyse future VRE generation structures and scenarios, and the variability of the generated power and steepness of the VRE power ramps without measured data. This knowledge can be used in the estimation of the required amount of balancing power, need for wind curtailment, potential of demand response, feasibility of energy storages or need for new grid components such as additional transmission lines.

The DSOs can also use the proposed methodologies to assess the effects of the increasing amount of consumer PV modules and, and thus, the increasing variations in the generated power and energy consumption in distribution systems. This kind of information can be vital for the effective planning of future smart grids and grid investments, so that the system is able to handle large number of prosumers with bidirectional power flows, local PV generation and household appliances with demand response.

The developed models can also provide vital information for the energy industry. Energy companies with WPPs, PVPs or both in their portfolio can use them as an additional source of information when, e.g., conducting cost-benefit analyses for, or evaluating feasibility of, possible VRE acquisitions or new planned VRE generation. Also, the analysis of the effects of the geographical distribution of generation locations to the volatility of their aggregated generation can be beneficial for the optimization of their generation portfolio. The information and insights provided by the proposed models can thus help the companies to allocate investments and resources related to future smart energy systems and VRE generation more efficiently, and therefore, to increase their profits.

In addition, the policy makers can benefit from the simulation of different future VRE scenarios and results obtained from the scenario analyses when designing ways to reduce CO<sub>2</sub> emissions or making other policy decisions related to, e.g., renewable energy subsidies or carbon taxes. This is important application of the work as it can allow more effective allocation of resources, and thus, faster transition to future smart energy systems with low CO<sub>2</sub> emissions.

Last, the researches working with topics related to future smart energy systems requiring plausible VRE generation data from specific cases for their own modelling and analyses can use the developed models as a source of artificial data. The developed models have been applied in various research topics to provide tailor made simulated data from PV panels of a single building to aggregated future VRE generation of a nationwide system. The models have been applied already in, e.g., demand response research [102], assessment of energy storages [103], economical analyses of VRE generation [104], development of distribution system grid planning algorithms [105] and combined analyses of consumption and VRE generation [106].

## 5.2 Recommendations for Future Work

This section suggests some possible further improvements and extensions that could be implemented in the developed models. The potential topics for future improvements can be divided into two main categories; the further development of the proposed models and the possible improvements to, and new applications for, the models due new estimation data.

A crucial future work aspect is the implementation of the most advanced of the developed models, i.e., the full VAR based approach developed for the modelling of WPPs, also for the joint modelling of WPPs and PVPs to improve the performance of the simplified VAR based methodology. Also, in this thesis the case studies and scenario analyses have been focusing on Finland, but should be expanded to Nordic and European scale in the future.

Relevant improvements that should be implemented in the wind direction modelling methodology include expanding the simplified VAR model to the full VAR model as the approach has been already developed for joint WPP modelling. In addition, the accurate modelling of wind direction margins for new WPP locations should be researched. Finally, also the possible correlation between wind speed and wind direction should be studied and implemented in the models if appropriate.

To improve the accuracy of the modelling of new WPPs, the modelling of the wake effect inside a WPP could be improved further by implementing a state-of-the-art wake effect model to fully benefit from the simulated wind direction data. In addition, as the models are modular by nature, additional power generation models for different PV panel types could be implemented to the PVP modelling methodology.

Another important topic is related to the measured data used in the estimation of the models. If more measurement data from larger number of locations could be obtained, the model parameters for new WPPs and PVPs could be estimated separately for, e.g., off-shore, onshore and inland locations, and for different climate zones, which could improve the accuracy even further. Also, more representative ECDFs for the modelling of new locations could be obtained when the ECDFs from the closest measured location are used in the modelling of new PVPs.

In this work, the models developed for new WPPs and PVPs utilized hourly time resolution. However, the models can be applied in the modelling of VRE generation in sub-hourly time resolutions, if sufficient estimation data could be acquired. Particularly interesting topic would be the study of the effect of the geographical distribution of the generation locations in sub-hourly time scales.

Finally, the modelling methodology for new PVPs should be verified also in the power domain using measured power output of individual PVPs or aggregated PVP generation in a certain geographical area, in addition to the presented model verification with measured solar irradiance locations.

### 5.3 Conclusions

This thesis developed multiple statistical methodologies for the modelling of VRE generation with an emphasis on new non-measured WPPs and PVPs modelled both separately and jointly. In addition, wind direction in new WPPs and large scale wind power generation in larger geographical areas were modelled. The main focus in this thesis was on the long term MC simulations, however, the developed models are also applicable in short term simulations. The developed modelling methods were verified against out-of-sample measurement data using MC simulations to ensure the accuracy of the models. The accuracy of the simulation results compared with measurement data was found to be good in all tested applications. Hence, the developed models are applicable for the assessment of, e.g., the effects of new VRE generation and future energy systems related investments and policy decisions.

The effect of the geographical distribution of the VRE generation was assessed in various scenarios. The geographical distribution was found to have a large effect on the aggregated VRE power generation and its variability. It was quantified how much dispersed geographical distribution decreased the volatility and probabilities for very low or high aggregated generation events compared with concentrated geographical distribution. The effect of the distribution of generation locations was also found to be much larger with WPPs compared with PVPs. Furthermore, WPPs and PVPs were found to be negatively correlated and the effect of this negative correlation was quantified. Consequently, a combination of both WPPs and PVPs were found to reduce the overall volatility of the aggregated VRE generation in the system, and thus, is recommended.

## References

- [1] IPCC, *Summary for Policymakers*. United Kingdom and New York, NY, USA: Cambridge University Press, 2014, ch. Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.
- [2] C. Le Quéré, R. J. Andres, T. Boden, T. Conway, R. A. Houghton, J. I. House, G. Marland, G. P. Peters, G. R. van der Werf, A. Ahlström, R. M. Andrew, L. Bopp, J. G. Canadell, P. Ciais, S. C. Doney, C. Enright, P. Friedlingstein, C. Huntingford, A. K. Jain, C. Jourdain, E. Kato, R. F. Keeling, K. Klein Goldewijk, S. Levis, P. Levy, M. Lomas, B. Poulter, M. R. Raupach, J. Schwinger, S. Sitch, B. D. Stocker, N. Viovy, S. Zaehle, and N. Zeng, “The global carbon budget 1959–2011,” *Earth System Science Data*, vol. 5, no. 1, pp. 165–185, 2013.
- [3] Global Wind Statistics 2016. Global Wind Energy Council GWEC. Cited Oct. 2017. [Online]. Available: [http://www.gwec.net/wp-content/uploads/vip/GWEC\\_PRstats2016\\_EN\\_WEB.pdf](http://www.gwec.net/wp-content/uploads/vip/GWEC_PRstats2016_EN_WEB.pdf)
- [4] Wind in power - 2016 European statistics. WindEurope. Cited Oct. 2017. [Online]. Available: <https://windeurope.org/wp-content/uploads/files/about-wind/statistics/WindEurope-Annual-Statistics-2016.pdf>
- [5] Snapshot of global photovoltaic markets 2016. International Energy Agency - Photovoltaic Power Systems Programme. Cited Oct. 2017. [Online]. Available: [http://www.iea-pvps.org/fileadmin/dam/public/report/statistics/IEA-PVPS\\_-\\_A\\_Snapshot\\_of\\_Global\\_PV\\_-\\_1992-2016\\_\\_1\\_.pdf](http://www.iea-pvps.org/fileadmin/dam/public/report/statistics/IEA-PVPS_-_A_Snapshot_of_Global_PV_-_1992-2016__1_.pdf)
- [6] H. Holttinen, A. Robitaille, A. Orths, I. Pineda, B. Lange, E. M. Carlini, J. Dillon, J. O. Tande, A. Estanqueiro, E. Gomez-lazaro, L. Söder, M. Milligan, and C. Smith, “Summary of experiences and studies for wind integration – iea wind task 25,” 2013.
- [7] L. Bird, D. Lew, M. Milligan, E. M. Carlini, A. Estanqueiro, D. Flynn, E. Gomez-Lazaro, H. Holttinen, N. Menemenlis, A. Orths, P. B. Eriksen, J. C. Smith, L. Soder, P. Sørensen, A. Altiparmakis, Y. Yasuda, and J. Miller, “Wind and solar energy curtailment: A review of international experience,” *Renewable and Sustainable Energy Reviews*, vol. 65, pp. 577 – 586, 2016.
- [8] M. M. Rienecker, M. J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M. G. Bosilovich, S. D. Schubert, L. Takacs, G.-K. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva, W. Gu, J. Joiner, R. D. Koster, R. Lucchesi, A. Molod, T. Owens, S. Pawson, P. Pegion, C. R. Redder, R. Reichle, F. R. Robertson, A. G. Ruddick, M. Sienkiewicz, and J. Woollen, “Merra: Nasa’s modern-era retrospective analysis for research and applications,” *Journal of Climate*, vol. 24, no. 14, pp. 3624–3648, 2011.

- [9] D. Cannon, D. Brayshaw, J. Methven, P. Coker, and D. Lenaghan, "Using reanalysis data to quantify extreme wind power generation statistics: A 33 year case study in great britain," *Renewable Energy*, vol. 75, no. Supplement C, pp. 767 – 778, 2015.
- [10] X. G. Larsén, S. Ott, J. Badger, A. N. Hahmann, and J. Mann, "Recipes for correcting the impact of effective mesoscale resolution on the estimation of extreme winds," *Journal of Applied Meteorology and Climatology*, vol. 51, no. 3, pp. 521–533, 2012.
- [11] R. Coić, J. Krstulović, and D. Jakus, "Simulation of aggregate wind farm short-term production variations," *Renewable Energy*, vol. 35, no. 11, pp. 2602–2609, May 2010.
- [12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*, ser. Springer Series in Statistics. Springer New York, 2009.
- [13] S. Ho, M. Xie, and T. Goh, "A comparative study of neural network and box-jenkins arima modeling in time series prediction," *Computers & Industrial Engineering*, vol. 42, no. 2, pp. 371 – 375, 2002.
- [14] J. Olauson and M. Bergkvist, "Modelling the swedish wind power production using merra reanalysis data," *Renewable Energy*, vol. 76, no. Supplement C, pp. 717 – 725, 2015.
- [15] I. Staffell and S. Pfenninger, "Using bias-corrected reanalysis to simulate current and future wind power output," *Energy*, vol. 114, no. Supplement C, pp. 1224 – 1239, 2016.
- [16] S. Rose and J. Apt, "What can reanalysis data tell us about wind power?" *Renewable Energy*, vol. 83, no. Supplement C, pp. 963 – 969, 2015.
- [17] C. Z. Mooney, *Monte Carlo Simulation*. Sage, 1997, pp. 1–12.
- [18] P. Sørensen, A. D. Hansen, and P. A. C. Rosas, "Wind models for simulation of power fluctuations from wind farms," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 90, no. 12, pp. 1381 – 1402, 2002, fifth Asia-Pacific Conference on Wind Engineering.
- [19] G. Papaefthymiou and D. Kurowicka, "Using copulas for modeling stochastic dependence in power system uncertainty analysis," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 40–49, February 2009.
- [20] K. Xie, Y. Li, and W. Li, "Modelling wind speed dependence in system reliability assessment using copulas," *IET Renewable Power Generation*, vol. 6, no. 6, pp. 392–399, November 2012.
- [21] Y. Li, K. Xie, and B. Hu, "Copula-arma model for multivariate wind speed and its applications in reliability assessment of generating systems," *Journal of Electrical Engineering and Technology*, vol. 8, pp. 421 – 427, 05 2013.
- [22] H. B. Azad, S. Mekhilef, and V. G. Ganapathy, "Long-term wind speed forecasting and general pattern recognition using neural networks," *IEEE Trans. Sustain. Energy*, vol. 5, no. 2, pp. 546–553, April 2014.
- [23] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *North American Power Symposium (NAPS), 2010*, Sept 2010, pp. 1–8.
- [24] A. Sturt and G. Strbac, "Time series modelling of power output for large-scale wind fleets," *Wind Energy*, vol. 14, no. 8, pp. 953–966, 2011.

- [25] R. Billinton, H. Chen, and R. Ghajar, "Time-series models for reliability evaluation of power systems including wind energy," *Microelectronics Reliability*, vol. 36, no. 9, pp. 1253 – 1261, 1996.
- [26] H. Louie, "Evaluation of bivariate archimedean and elliptical copulas to model wind power dependency structures," *Wind Energy*, vol. 17, no. 2, pp. 225–240, 2014.
- [27] B. G. Brown, R. W. Katz, and A. H. Murphy, "Time series models to simulate and forecast wind speed and wind power," *Journal of Climate and Applied Meteorology*, vol. 23, May 1984.
- [28] J. Torres, A. García, M. D. Blas, and A. D. Francisco, "Forecast of hourly average wind speed with arma models in navarre (spain)," *Solar Energy*, vol. 79, no. 1, pp. 65 – 77, 2005.
- [29] B. Klockl, "Multivariate time series models applied to the assessment of energy storage in power systems," in *Proceedings of the 10th International Conference on Probabilistic Methods Applied to Power Systems*, May 2008, pp. 1–8.
- [30] D. Villanueva, A. Feijóo, and J. L. Pazos, "Simulation of correlated wind speed data for economic dispatch evaluation," *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, January 2012.
- [31] D. A. Bechrakis and P. D. Sparis, "Correlation of wind speed between neighboring measuring stations," *IEEE Transactions on Energy Conversion*, vol. 19, no. 2, June 2004.
- [32] K. Nystrom and J. Skoglund, "Univariate extreme value theory, garch and measures of risk," *Preprint, Swedbank*, September 2002.
- [33] A. J. McNeil and R. Frey, "Estimation of tail related risk measure for heteroscedastic financial time series: an extreme value approach," *Journal of Empirical Finance*, vol. 7, pp. 271–300, 2000.
- [34] D. C. Hill, D. McMillan, K. R. W. Bell, and D. Infield, "Application of autoregressive models to U.K. wind speed data for power system impact studies," *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, pp. 134–141, January 2012.
- [35] B. Klöckl and G. Papaefthymiou, "Multivariate time series models for studies on stochastic generators in power systems," *Renewable Energy*, vol. 80, no. 3, pp. 265–276, March 2010.
- [36] L. Soder, "Simulation of wind speed forecast errors for operation planning of multiarea power systems," in *2004 International Conference on Probabilistic Methods Applied to Power Systems*, Sept 2004, pp. 723–728.
- [37] G. Papaefthymiou and B. Klockl, "Mcmc for wind power simulation," *IEEE Transactions on Energy Conversion*, vol. 23, no. 1, pp. 234–240, March 2008.
- [38] B. Deler and B. L. Nelson, "Modeling and generating multivariate time series with arbitrary marginals using a vector autoregressive technique," in *Proceedings of the 2001 Winter Simulation Conference*, 2001.
- [39] Z. Qin, W. Li, and X. Xiong, "Generation system reliability evaluation incorporating correlations of wind speeds with different distributions," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 551–558, Feb 2013.
- [40] J. L. Myers and A. D. Well, *Research Design and Statistical Analysis*, 2nd ed. New Jersey, Mahwah: Lawrence Erlbaum, 2003, pp. 504–511.

- [41] M. Koivisto, J. Seppänen, I. Mellin, J. Ekström, J. Millar, I. Mammarella, M. Komppula, and M. Lehtonen, "Wind speed modeling using a vector autoregressive process with a time-dependent intercept term," *International Journal of Electrical Power Energy Systems*, vol. 77, pp. 91 – 99, 2016.
- [42] Z. Yu and A. Tuzuner, "Fractional weibull wind speed modeling for wind power production estimation," in *2009 IEEE Power Energy Society General Meeting*, July 2009, pp. 1–7.
- [43] A. Chauhan and R. P. Saini, "Statistical analysis of wind speed data using weibull distribution parameters," in *2014 1st International Conference on Non Conventional Energy (ICONCE 2014)*, Jan 2014, pp. 160–163.
- [44] Z. Qin, W. Li, and X. Xiong, "Estimating wind speed probability distribution using kernel density method," *Electric Power Systems Research*, vol. 81, no. 12, pp. 2139–2146, December 2011.
- [45] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1986.
- [46] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*, 2nd ed. Berlin, Heidelberg: Springer-Verlag, 2007.
- [47] D. Villanueva and A. E. Feijóo, "Reformulation of parameters of the logistic function applied to power curves of wind turbines," *Electric Power Systems Research*, vol. 137, pp. 51 – 58, 2016.
- [48] M. Lydia, S. S. Kumar, A. I. Selvakumar, and G. E. Prem, "A comprehensive review on wind turbine power curve modeling techniques," *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 452 – 460, 2014.
- [49] V. Turkia and H. Holttinen, "Wind energy statistics of finland. yearly report 2011," *VTT Technical Research Centre of Finland*, 2013.
- [50] F. González-Longatt, P. Wall, and V. Terzija, "Wake effect in wind farm performance: Steady-state and dynamic behavior," *Renewable Energy*, vol. 39, no. 1, pp. 329 – 338, 2012.
- [51] A. Duckworth and R. Barthelmie, "Investigation and validation of wind turbine wake models," *Wind Engineering*, vol. 32, no. 5, pp. 459–475, 2008.
- [52] R. J. Barthelmie, K. Hansen, S. T. Frandsen, O. Rathmann, J. G. Schepers, W. Schlez, J. Phillips, K. Rados, A. Zervos, E. S. Politis, and P. K. Chaviaropoulos, "Modelling and measuring flow and wind turbine wakes in large wind farms offshore," *Wind Energy*, vol. 12, no. 5, pp. 431–444, 2009.
- [53] E. Erdem and J. Shi, "Arma based approaches for forecasting the tuple of wind speed and direction," *Applied Energy*, vol. 88, no. 4, pp. 1405 – 1414, 2011.
- [54] N. Masseran, "Markov chain model for the stochastic behaviors of wind-direction data," *Energy Conversion and Management*, vol. 92, 03 2015.
- [55] S. Pfenninger and I. Staffell, "Long-term patterns of european pv output using 30 years of validated hourly reanalysis and satellite data," *Energy*, vol. 114, no. Supplement C, pp. 1251 – 1265, 2016.
- [56] W. Wu, K. Wang, B. Han, G. Li, X. Jiang, and M. L. Crow, "A versatile probability model of photovoltaic generation using pair copula construction," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1337–1345, Oct 2015.
- [57] H. Cheng, W. s. Cao, and P. j. Ge, "Forecasting research of long-term solar irradiance and output power for photovoltaic generation system," in *2012 Fourth International Conference on Computational and Information Sciences*, Aug 2012, pp. 1224–1227.

- [58] W. Ji and K. C. Chee, "Prediction of hourly solar radiation using a novel hybrid model of arma and tdnn," *Solar Energy*, vol. 85, no. 5, pp. 808 – 817, 2011.
- [59] L. Mora-López and M. S. de Cardona, "Multiplicative arma models to generate hourly series of global irradiation," *Solar Energy*, vol. 63, no. 5, pp. 283 – 291, 1998.
- [60] D. Yang, Z. Dong, T. Reindl, P. Jirutitijaroen, and W. M. Walsh, "Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage," *Solar Energy*, vol. 103, no. Supplement C, pp. 550 – 562, 2014.
- [61] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," *Solar Energy*, vol. 83, no. 3, pp. 342 – 349, 2009.
- [62] Z. Dong, D. Yang, T. Reindl, and W. M. Walsh, "Short-term solar irradiance forecasting using exponential smoothing state space model," *Energy*, vol. 55, no. Supplement C, pp. 1104 – 1113, 2013.
- [63] E. Nuño, M. Koivisto, N. Cutululis, and P. Sørensen, "Simulation of regional day-ahead pv power forecast scenarios," in *2017 IEEE Manchester PowerTech*, June 2017, pp. 1–6.
- [64] B. Ngoko, H. Sugihara, and T. Funaki, "Synthetic generation of high temporal resolution solar radiation data using markov models," *Solar Energy*, vol. 103, pp. 160 – 170, 2014.
- [65] C. M. Fernández-Peruchena, M. Gastón, M. Sánchez, J. García-Barberena, M. Blanco, and A. Bernardos, "Mus: A multiscale stochastic model for generating plausible meteorological years designed for multiyear solar energy yield simulations," *Solar Energy*, vol. 120, pp. 244 – 256, 2015.
- [66] D. Lingfors and J. Widén, "Development and validation of a wide-area model of hourly aggregate solar power generation," *Energy*, vol. 102, pp. 559 – 566, 2016.
- [67] W. Wu, K. Wang, B. Han, G. Li, X. Jiang, and M. L. Crow, "A versatile probability model of photovoltaic generation using pair copula construction," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1337–1345, Oct 2015.
- [68] M. Larrañeta, S. Moreno-Tejera, M. Silva-Pérez, and I. Lillo-Bravo, "An improved model for the synthetic generation of high temporal resolution direct normal irradiation time series," *Solar Energy*, vol. 122, pp. 517 – 528, 2015.
- [69] J. Bright, C. Smith, P. Taylor, and R. Crook, "Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data," *Solar Energy*, vol. 115, pp. 229 – 242, 2015.
- [70] C. Rigollier, O. Bauer, and L. Wald, "On the clear sky model of the {ESRA} — european solar radiation atlas — with respect to the heliosat method," *Solar Energy*, vol. 68, no. 1, pp. 33 – 48, 2000.
- [71] M. Mattei, G. Notton, C. Cristofari, M. Muselli, and P. Poggi, "Calculation of the polycrystalline (PV) module temperature using a simple method of energy balance," *Renewable Energy*, vol. 31, no. 4, pp. 553 – 567, 2006.
- [72] R. Perez, P. Ineichen, E. Maxwell, R. Seals, and A. Zelenka, "Dynamic global-to-direct irradiance conversion models," *ASHRAE Transactions*, vol. 98, pp. 354–369, 1992.
- [73] J. Boland, J. Huang, and B. Ridley, "Decomposing global solar radiation into its direct and diffuse components," *Renewable and Sustainable Energy Reviews*, vol. 28, no. Supplement C, pp. 749 – 756, 2013.

- [74] B. Ridley, J. Boland, and P. Lauret, "Modelling of diffuse solar fraction with multiple predictors," *Renewable Energy*, vol. 35, no. 2, pp. 478 – 483, 2010.
- [75] M. Marinelli, P. Maule, A. N. Hahmann, O. Gehrke, P. B. Nørgård, and N. A. Cutululis, "Wind and photovoltaic large-scale regional models for hourly production evaluation," *IEEE Trans. Sustain. Energy*, vol. 6, no. 3, pp. 916–923, July 2015.
- [76] J. Widen, "Correlations between large-scale solar and wind power in a future scenario for Sweden," *IEEE Trans. Sustain. Energy*, vol. 2, no. 2, pp. 177–184, April 2011.
- [77] J. Vasilj, P. Sarajcev, and D. Jakus, "Estimating future balancing power requirements in wind–pv power system," *Renewable Energy*, vol. 99, pp. 369 – 378, 2016.
- [78] P. E. Bett and H. E. Thornton, "The climatological relationships between wind and solar energy supply in Britain," *Renewable Energy*, vol. 87, no. Part 1, pp. 96 – 110, 2016.
- [79] J. Rank, *Copulas: From theory to application in finance*. Haymarket House, London: Risk Books, a Division of Incisive Financial Publishing Ltd, 2007, pp. 3–34.
- [80] P. M. T. Broersen, "Autoregressive model orders for durbin's ma and arma estimators," *IEEE Transactions on Signal Processing*, vol. 48, no. 8, pp. 2454–2457, Aug 2000.
- [81] C. Clapham and J. Nicholson, "Gauss–markov theorem," in *The Concise Oxford Dictionary of Mathematics*, ser. Oxford Quick Reference. OUP Oxford, 2009. [Online]. Available: <http://www.oxfordreference.com/view/10.1093/acref/9780199235940.001.0001/acref-9780199235940-e-1222>
- [82] J. Hamilton, *Time Series Analysis*. Princeton University Press, 1994, pp. 200–220, 657–672.
- [83] D. A. Dickey and W. A. Fuller, "Likelihood ratio statistics for autoregressive time series with a unit root," *Econometrica*, vol. 49, no. 4, pp. 1057–1072, 1981.
- [84] G. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Econometrica*, vol. 66, pp. 67–72, 1978.
- [85] B. Tammelin, T. Vihma, E. Atlaskin, J. Badger, C. Fortelius, H. Gregow, M. Horttanainen, R. Hyvönen, J. Kilpinen, J. Latikka, K. Ljungberg, N. G. Mortensen, S. Niemelä, K. Ruosteenoja, K. Salonen, I. Suomi, and A. Venäläinen, "Production of the Finnish wind atlas," *Wind Energy*, vol. 16, no. 1, pp. 19–35, 2013.
- [86] L. L. Freris and D. Infield, *Renewable Energy in Power Systems*. Hoboken, N.J.: John Wiley and Sons Inc., 2008.
- [87] Gamesa G128–4.5 MW wind turbine data sheet. Siemens Gamesa Renewable Energy. Cited Oct. 2017. [Online]. Available: <http://www.gamesacorp.com/recursos/doc/productos-servicios/aerogeneradores/catalogo-g10x-45mw-eng.pdf>
- [88] A. Luque and S. Hegedus, *Handbook of Photovoltaic Science and Engineering*. John Wiley and Sons Inc., 2003.
- [89] Solar radiation data (soda) service. Cited Oct. 2017. [Online]. Available: <http://www.soda-pro.com/web-services/atmosphere/linke-turbidity-factor-ozone-water-vapor-and-angstroembeta>

- [90] R. Harris, “Testing for unit roots using the augmented dickey-fuller test: Some issues relating to the size, power and the lag structure of the test,” *Economics Letters*, vol. 38, no. 4, pp. 381 – 386, 1992.
- [91] J. Duffie and W. Beckman, *Solar Engineering of Thermal Processes*. Hoboken, N.J.: John Wiley and Sons Inc., 2006.
- [92] N. I. Fisher and A. J. Lee, “A correlation coefficient for circular data,” *Biometrika*, vol. 70, no. 2, pp. 327–332, 1983.
- [93] R. S. Pindyck and D. L. Rubinfeld, *Econometric Models and Economic Forecasts*, 3rd ed. Singapore: The McGraw-Hill, 1991.
- [94] N. I. Fisher and A. J. Lee, “Time series analysis of circular data,” *Journal of the Royal Statistical Society*, vol. 56, no. 2, pp. 327–339, 1994.
- [95] G. Hughes, *Multivariate and time series models for circular data with applications to protein conformational angles. Doctoral Dissertation*. Leeds, UK: The University of Leeds, Department of Statistics, 2007.
- [96] Nord Pool Spot bidding areas. Nord Pool. Cited Oct. 2017. [Online]. Available: <http://www.nordpoolspot.com/the-power-market/Bidding-areas/>
- [97] Wind energy scenarios for 2020. Cited Oct. 2017. [Online]. Available: <http://www.ewea.org/fileadmin/files/library/publications/reports/EWEA-Wind-energy-scenarios-2020.pdf>
- [98] S. M. Fisher, J. Schoof, C. L. Lant, and M. Therrell, “The effects of geographical distribution on the reliability of wind energy,” *Applied Geography*, vol. 40, pp. 83–89, 06 2013.
- [99] J. Kiviluoma, H. Holttinen, D. Weir, R. Scharff, L. Söder, N. Menemenlis, N. A. Cutululis, I. Danti Lopez, E. Lannoye, A. Estanqueiro, E. Gomez-Lazaro, Q. Zhang, J. Bai, Y.-H. Wan, and M. Milligan, “Variability in large-scale wind power generation,” *Wind Energy*, vol. 19, no. 9, pp. 1649–1665, 2016, wE-14-0277.R2.
- [100] Vestas V105–3.3 MW wind turbine data sheet. Vestas Wind Systems A/S. Cited Oct. 2017. [Online]. Available: <https://en.wind-turbine-models.com/turbines/693-vestas-v112-3.3>
- [101] R. Perez and T. E. Hoff, *Solar Energy Forecasting and Resource Assessment: Solar Resource Variability*. Academic Press, 07 2013, pp. 133–148.
- [102] A. Alahäivälä, J. Ekström, J. Jokisalo, and M. Lehtonen, “A framework for the assessment of electric heating load flexibility contribution to mitigate severe wind power ramp effects,” *Electric Power Systems Research*, vol. 142, no. Supplement C, pp. 268 – 278, 2017.
- [103] M. Ali, J. Ekström, and M. Lehtonen, “Assessing the potential benefits and limits of electric storage heaters for wind curtailment mitigation: A Finnish case study,” *Sustainability*, vol. 9, no. 5, 2017.
- [104] V. Olkkonen, M. Koivisto, J. Ekström, and S. Syri, “Technical and economic implications of wind power integration in the Finnish energy system: A stochastic/probabilistic scenario analysis for 2030,” in *International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems (ECOS)*, Pau, France, June 2015.
- [105] R. J. Millar, M. Koivisto, J. Ekström, E. Saarijärvi, and M. Lehtonen, “Probabilistic prosumer node modeling for estimating planning parameters in distribution networks with renewable energy sources,” in *58th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON)*, Riga, Latvia, October 2017.

## References

- [106] M. Koivisto, M. Degefa, M. Ali, J. Ekström, J. Millar, and M. Lehtonen, “Statistical modeling of aggregated electricity consumption and distributed wind generation in distribution systems using amr data,” *Electric Power Systems Research*, vol. 129, no. Supplement C, pp. 217 – 226, 2015.

Reducing carbon dioxide emissions is imperative when trying to mitigate the effects of global warming. In energy generation, one approach to achieve this goal is to replace polluting generation with variable renewable energy (VRE) sources, such as wind and solar energy. However, these are intermittent by nature, and thus, increasing the VRE capacity also increases the variations in power generation. This, in turn, can cause problems for the operation of the power system. Hence, it is important to be able to model the VRE generation and its variations, so the effects can be understood and compensated. The dissertation tackles this problem by proposing statistical modelling methods for the simulation of VRE generation in new generation locations. It also quantifies the effect of the geographical distribution of VRE generation on the variability of the aggregated power generation. The developed methods can be used to assess the impact of new VRE generation, and thus, to focus the investments efficiently and to enable the large scale integration of VRE generation in energy systems.



ISBN 978-952-60-7881-6 (printed)

ISBN 978-952-60-7882-3 (pdf)

ISSN-L 1799-4934

ISSN 1799-4934 (printed)

ISSN 1799-4942 (pdf)

**Aalto University**  
**School of Electrical Engineering**  
**Department of Electrical Engineering and Automation**  
[www.aalto.fi](http://www.aalto.fi)

**BUSINESS +  
ECONOMY**

**ART +  
DESIGN +  
ARCHITECTURE**

**SCIENCE +  
TECHNOLOGY**

**CROSSOVER**

**DOCTORAL  
DISSERTATIONS**