

Integration of Stochastic Generation in Power Systems

Georgios Papaefthymiou

Front cover: simulated daily aggregate wind power in the Netherlands for the year 2003, based on the 5GW-integration scenario presented in Chapter 4. Measured data obtained by the Royal Netherlands Meteorological Institute (KNMI) [78].

Back cover: daily system load in 2003 in the Netherlands (Chapter 2, Fig. 2.3).

Integration of Stochastic Generation in Power Systems

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.dr.ir. J.T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op woensdag 13 juni 2006 om 12:30 uur

door

Georgios PAPAETHYMIU
Diploma Electrical and Computer Engineer, University of Patras, Greece

geboren te Mesolongi, Griekenland.

Dit proefschrift is goedgekeurd door de promotor:
Prof .ir. L. van der Sluis

Samenstelling promotiecomissie:

Rector Magnificus,	Voorzitter, Technische Universiteit Delft
Prof. ir. L. van der Sluis	Technische Universiteit Delft, promotor
Prof. ir. W.L. Kling,	Technische Universiteit Delft/Eindhoven, TenneT
Prof. dr. R.M. Cooke,	Technische Universiteit Delft, Resources for the Future
Prof. dr. N. Hatziargyriou,	National Technical University of Athens, Greece
Prof. ir. M. Antal,	Technische Universiteit Eindhoven, Emeritus
Dr. ir. D. Kurowicka,	Technische Universiteit Delft
Dr. ir. P.H. Schavemaker,	Technische Universiteit Delft
Prof. dr. ir. R.L. Lagendijk, reservelid,	Technische Universiteit Delft

This research has been performed within the framework of the research program 'Intelligent Power Systems' that is supported financially by SenterNovem, an agency of the Dutch Ministry of Economic Affairs.

Integration of Stochastic Generation in Power Systems.

Dissertation at Delft University of Technology.

ISBN: 978-90-8570-186-6

Copyright © 2007 by G. Papaefthymiou.
All rights reserved.

Cover design by Michela Pelusio and Marnix Dekker (<http://maakali.org/>).

Printed by: Wöhrmann Print Service, Zutphen, the Netherlands.

Contents

Contents	i
List of Figures	v
List of Tables	ix
1 Introduction	1
1.1 'Vertical' power system	1
1.2 Problems with CG	3
1.2.1 Depletion of fossil fuels	3
1.2.2 Greenhouse gases emissions	4
1.2.3 Nuclear energy	4
1.2.4 Hydro power	4
1.3 Renewable energy sources (RES)	4
1.4 Deregulation & liberalization of the energy markets	5
1.5 non-Conventional Generation (non-CG)	6
1.5.1 Types of non-CG	6
1.5.2 Characteristics of non-CG	11
1.5.3 Distributed Generation (DG)	11
1.5.4 Stochastic Generation (SG)	12
1.6 Horizontally-Operated Power System	13
1.7 Objectives and limitations	13
1.8 Research Framework: 'Intelligent Power Systems'	16
1.9 Outline of the Thesis	17
2 Power System Steady-State Uncertainty Analysis	19
2.1 Power system steady-state: a 'snapshot' approach	19
2.2 Deterministic System Model (DSM)	20
2.2.1 Steady-State System Model	20
2.2.2 Deterministic analysis and vertical power system	21
2.3 Stochastic System Model (SSM)	21
2.3.1 Uncertainty analysis methodologies: general	21
2.3.2 Uncertainty in power system analysis	21
2.4 Literature review	22
2.4.1 Fuzzy arithmetic analysis	22

2.4.2	Possibilistic vs Probabilistic Analysis	22
2.5	Probabilistic steady-state uncertainty analysis	23
2.5.1	Analytical methods	23
2.5.2	Stochastic Simulations (MCS)	26
2.5.3	Snapshot Approach and MCS: Memory-less system	26
2.5.4	Literature review: conclusions	27
2.6	Load Uncertainty vs SG Uncertainty	27
2.6.1	Load Uncertainty: time-dependent stochasticity	27
2.6.2	SG Uncertainty	28
2.7	Planning uncertainty vs operational uncertainty	31
2.8	Problem Formulation	31
2.8.1	Types of r.v.	31
2.8.2	Problem Decomposition	33
2.9	Marginal distributions	34
2.9.1	Sampling of a r.v. in MCS	34
2.9.2	Load	35
2.9.3	Stochastic Generation	39
2.10	Stochastic Dependence Structure	40
2.11	Conclusions	41
3	Models of Stochastic Dependence	43
3.1	Mean value of the sum of r.v.	43
3.2	Independence	43
3.2.1	Independent normal loads	44
3.2.2	Independent stochastic generators: Central Limit Theorem	44
3.3	Measure of Dependence: Product Moment Correlation	47
3.4	Correlated normal loads: Joint Normal Distribution	48
3.4.1	Bivariate case	48
3.4.2	Multivariate case	49
3.4.3	Positive-semidefiniteness	51
3.5	Correlated stochastic generators: where product moment correlation fails	51
3.5.1	Bivariate case	51
3.6	Rank correlation: transformation of marginals into ranks	54
3.6.1	Rank correlation	54
3.6.2	Decoupling of dependence structure and marginals: towards copulas	55
3.7	Copulas	57
3.7.1	Sklar's Theorem	57
3.7.2	Fréchet-Hoeffding bounds	58
3.7.3	Families of Copulas	58
3.7.4	Copulas and rank correlation	61
3.7.5	Sampling of two correlated r.v. using copula	63
3.8	Conclusions	65

4	Multidimensional Dependence Modeling	67
4.1	Joint Normal Transform methodology	68
4.2	Wind power integration in The Netherlands	68
4.2.1	Marginal distributions	71
4.2.2	Dependence structure: correlation matrices	74
4.2.3	5GW-integration scenario: positive semi-definite matrix	75
4.2.4	6GW-integration scenario: non-positive semidefinite matrix	77
4.3	Positive semi-definiteness: Problems and repairing methods	77
4.3.1	Problems with positive semi-definiteness	77
4.3.2	Repairing methods	80
4.4	Problems with JNT methodology	80
4.5	Dependence Trees with Copulas	81
4.5.1	General	81
4.5.2	Definitions	82
4.6	Dependence Vines with Copulas	84
4.6.1	General	84
4.6.2	Definitions	85
4.7	Conclusions	86
5	Stochastic Plants	89
5.1	The co-fluctuation of stochastic inputs	90
5.2	Stochastic Bounds Methodology (SBM)	90
5.2.1	Upper bound: comonotonicity	92
5.2.2	Lower bounds: countermonotonicity - independence	92
5.3	Stochastic plants	95
5.3.1	Stochastic model reduction	95
5.3.2	Stochastic plants comprising SG units and loads	97
5.4	Wind stochastic plant: example	98
5.4.1	Part I: Impact of the dependence structure	98
5.4.2	Part II: Impact of the marginal distributions	100
5.5	Conclusions	103
6	Applications	105
6.1	Integration of SG in Distribution Systems	105
6.1.1	Simulation data	105
6.1.2	Marginal distributions	107
6.1.3	Appropriate dependence model	108
6.1.4	Active distribution network	111
6.1.5	System losses	112
6.2	Voltage control by SG units	113
6.2.1	Voltage control: theory	113
6.2.2	Voltage control: results for the study case	118
6.3	Integration of SG in a Bulk Power System	120
6.3.1	Simulation data	120
6.3.2	System loads	123
6.3.3	System wind power	125

6.3.4	System power injections	127
6.3.5	System operation: CG units	127
6.3.6	System slack bus	128
6.3.7	System power flows	130
6.3.8	System losses	130
6.4	Conclusions	134
7	Conclusions and recommendations	137
7.1	Conclusions	137
7.2	Recommendations for future work	140
A	Steady-State System Model	143
B	Repairing violations of positive semi-definiteness	147
B.1	Hypersphere Decomposition	147
B.2	Spectral Decomposition	148
C	Simulation data	149
C.1	Correlation matrices for the example in section 4.2	149
C.2	Weibull parameters for wind speed distributions in section 5.4	150
C.3	Distribution test system data for the study case in section 6.1	151
C.4	Bulk power system data for the study case in section 6.3	152
	Bibliography	155
	List of Abbreviations	163
	Summary in English	165
	Integration of Stochastic Generation in Power Systems	165
	Samenvatting in het Nederlands	169
	Integratie van Stochastische Opwekking in Elektricititeitssystemen	169
	List of Publications	173
	Journal Papers	173
	Conference Papers	173
	Acknowledgements	177
	Biography	179

List of Figures

1.1	A Vertically-Operated power system [17].	2
1.2	A conventional (passive) distribution network (upper graph - a) and an active distribution network (lower graph - b) [76].	14
1.3	'Vertical-to-Horizontal' transformation of a power system [76].	15
1.4	The four parts of intelligent power systems research project.	16
2.1	Schematic diagram of the system analysis.	20
2.2	Daily load for a distribution system for one month.	28
2.3	Daily system load in 2003 in The Netherlands.	29
2.4	Daily system load in September 2003 in The Netherlands.	29
2.5	Seasonal daily system load in 2003 in The Netherlands.	30
2.6	Windpower uncertainty.	32
2.7	Sampling of a r.v. in MCS.	35
2.8	Sampling of a r.v. in MCS based on measured data.	36
2.9	Normalized load distribution as mixture of TF-distributions (10000-sample MCS).	37
2.10	pdf for the system load in 2003 in The Netherlands.	38
2.11	WTG wind speed/power characteristic and distributions.	39
3.1	Independent normal load distributions.	44
3.2	Independent Weibull wind speed distributions.	45
3.3	Independent wind turbine power output distributions.	46
3.4	Scatter diagrams for correlated normal loads.	49
3.5	Aggregate distribution of correlated normal loads.	50
3.6	Time-series and scatter diagrams for perfect correlation between loads and wind speeds.	52
3.7	Scatter diagrams between X and powers of X.	53
3.8	Scatter diagrams for independence and perfect dependence between normal and Weibull distributions and their respective ranks.	56
3.9	Scatter diagrams for the Independent and Fréchet copulas.	58
3.10	Scatter diagrams for four types of copulas under the same rank correlation.	60
3.11	Scatter diagrams for the diagonal band copula for different rank correlations.	62
3.12	Wind scatter diagrams and power output distributions for two WTGs under the same rank correlation but with different dependence structures.	63

3.13	Sampling procedure for the diagonal band copula.	64
4.1	Wind park locations in The Netherlands	70
4.2	Wind speed cdfs for the wind park locations.	72
4.3	Measured and simulated wind speed distributions at site 1 (onshore) and 14 (offshore).	73
4.4	System load distributions for the year 2020 obtained by extrapolation of the 15-min recordings of year 2003.	74
4.5	Onshore and offshore wind power distributions for the 5GW-integration scenario.	76
4.6	Total system wind power and system net load distributions for the 5GW-integration scenario.	76
4.7	Deviation between the non-positive semi-definite rank correlation matrix $\mathbf{R}_r^{\text{WLe}}$ and the one obtained after the application of the repairing method, $\mathbf{R}_r^{\text{WLe-psd}}$	78
4.8	Onshore and offshore wind power distributions for the 6GW-integration scenario.	78
4.9	Total system wind power and system net load distributions for the 6GW-integration scenario.	79
4.10	Sampling procedure for a tree of 3 r.v.	83
4.11	A tree with 6 nodes (left) and an undirected graph with a cycle (right). The notation r_{ij} refers to the rank correlation ρ_{rij}	84
4.12	A D-vine on 4 variables.	85
5.1	Time-series and scatter diagram of 5-minute measurements of active power output of a windpark.	91
5.2	Independent/comonotonic/countermonotonic sampling and generated sequences.	94
5.3	Model reduction by the application of the stochastic plants concept. A single random generator U is used for the modeling of each stochastic plant.	96
5.4	Clustering of stochastic plants comprising stochastic generators and loads.	97
5.5	Wind stochastic plant aggregate power output distributions.	99
5.6	Mean value and standard deviation distributions for the WSP aggregate power output distribution for the 5000 cases of different marginals.	102
6.1	Single-line diagram for the 34-bus/38-line radial test distribution system.	106
6.2	Voltage distributions at node 42 under different extreme dependence scenarios.	109
6.3	Power flow distributions in branch 47-43 under different extreme dependence scenarios.	110
6.4	Power flow distributions in the branch 47-43, in two system states under different wind power penetration levels.	111
6.5	System power loss distributions for two system states and three wind power penetration levels.	112
6.6	Voltage control characteristics of the <i>maximum reactive power threshold control</i> and the <i>droop control</i>	115
6.7	Voltage control characteristics for different points in time.	115

6.8	Voltage control iterative algorithms.	117
6.9	Voltage distributions at node 42 for the different voltage control strategies (5000-sample MCS).	119
6.10	Box-plots of the nodal voltage distributions for the highest wind penetration level in the future system state (5000-sample MCS).	121
6.11	Single-line diagram of the 39-bus New England test system [65].	122
6.12	4-TF load modeling for the New England test system (10000-sample MCS).	123
6.13	Scatter diagrams for the load modeling based on specified intra- and inter-TF rank correlation (10000-sample MCS).	124
6.14	WSP power output in the New England test system (10000-sample MCS).	126
6.15	Wind speed and wind power scatter diagrams (10000-sample MCS).	127
6.16	Power injection at bus 4 for the 4 wind power penetration levels (1000-sample MCS).	128
6.17	Slack bus power injection distributions and boxplots (10000-sample MCS).	129
6.18	Box-plot for the power flows in the system lines in case of no wind power penetration (10000-sample MCS).	131
6.19	Box-plot for the power flows in the system lines in case of 25% wind power penetration (10000-sample MCS).	131
6.20	Box-plot for the power flows in the system lines in case of 50% wind power penetration (10000-sample MCS).	132
6.21	Box-plot for the power flows in the system lines in case of 75% wind power penetration (10000-sample MCS).	132
6.22	Some specific power flow distributions (10000-sample MCS).	133
6.23	Distributions of system losses (10000-sample MCS).	134

List of Tables

1.1	Worldwide installed capacity (GW) on January 1,2000 [74].	3
1.2	Basic characteristics of CG and non-CG.	11
1.3	Classification of non-CG based on dispatchability.	13
2.1	Time-frames settings	36
4.1	Capacity factor for the wind parks in The Netherlands.	76
5.1	Wind stochastic plant aggregate power output mean values, standard deviations and 5%-95% quantiles.	100
5.2	Mean values and standard deviations of the distributions in Fig. 5.6a. . .	101
5.3	Mean values and standard deviations of the distributions in Fig. 5.6b. . .	101
6.1	Mean values and standard deviations for the voltage and power flow distributions in the different extreme dependence concepts.	110
6.2	Mean values/standard deviations of the power flow distributions in line 47-43.	111
6.3	Mean and annual system losses.	112
6.4	Mean system reactive power production for the SG units [MVAR].	118
6.5	Characteristics of the New England test system.	120
6.6	TF settings for a 4-TF load modeling of the New England test system. . .	123
6.7	Capacity factor for the WSPs in the New England test system.	125
6.8	Mean value and standard deviation for the power injections at bus 4 for the 4 wind power penetration levels.	127
6.9	Mean value and standard deviation for the slack bus power injection distributions.	130
6.10	Mean value and standard deviation for the distributions of the system losses.134	
C.1	System data for the 34-bus/38-line radial test distribution system.	151
C.2	Bus Data of the New England 39 Bus Test System [65]	152
C.3	Line Data of the New England 39 Bus Test System [65]	153

Chapter 1

Introduction

The overall purpose of an electric power system is to supply electric energy to the loads, i.e. the customers, in a safe, economic and reliable way. The generation, transmission and distribution of electrical power are the three main tasks, or primary functions of a power system.

Power systems operate under the restriction that derives from the non-storability of electrical energy: the electrical energy produced and consumed throughout the system should be equal at each moment in time. Consequently, the system operation relies on the control of the system power generation and its adaptation to the fluctuations of the uncertain system load, according to the restrictions set by the power transport networks.

1.1 'Vertical' power system

Power systems traditionally evolved based on a 'vertical' structure, in which power is generated by a relatively small number of large power plants. These large power plants are mainly constructed at remote sites, close to the energy resources or supply routes and relatively far from the load centers. The electrical energy is transported from these sources to the dispersed end-users by a hierarchical structure of high-voltage (HV) transmission networks and medium-voltage (MV) and low-voltage (LV) distribution networks, as shown in Fig. 1.1. To ensure both high security and availability, the transmission networks are meshed, to provide alternative routing in case of faults. The distribution networks are *passive systems** with a radial structure that enables the operation of selective protection. In this system structure, the electrical power flows from the higher to the lower voltage levels.

The power generation in a 'vertical' power system structure is mainly based on *controllable primary energy sources* such as fossil fuels in the case of large *thermo-electric power plants* or dammed water in the case of *hydro-electric power plants*. There are two basic types of thermo-electric power plants: the *fossil-fuel-fired power plants*, where fossil fuels such as oil, gas and coal are burnt and the resulting thermal energy is converted into electrical power by means of a steam cycle, and the *nuclear power plants*, where nuclear

*Passive systems are parts of the power system where only loads and no generation is connected as passive.

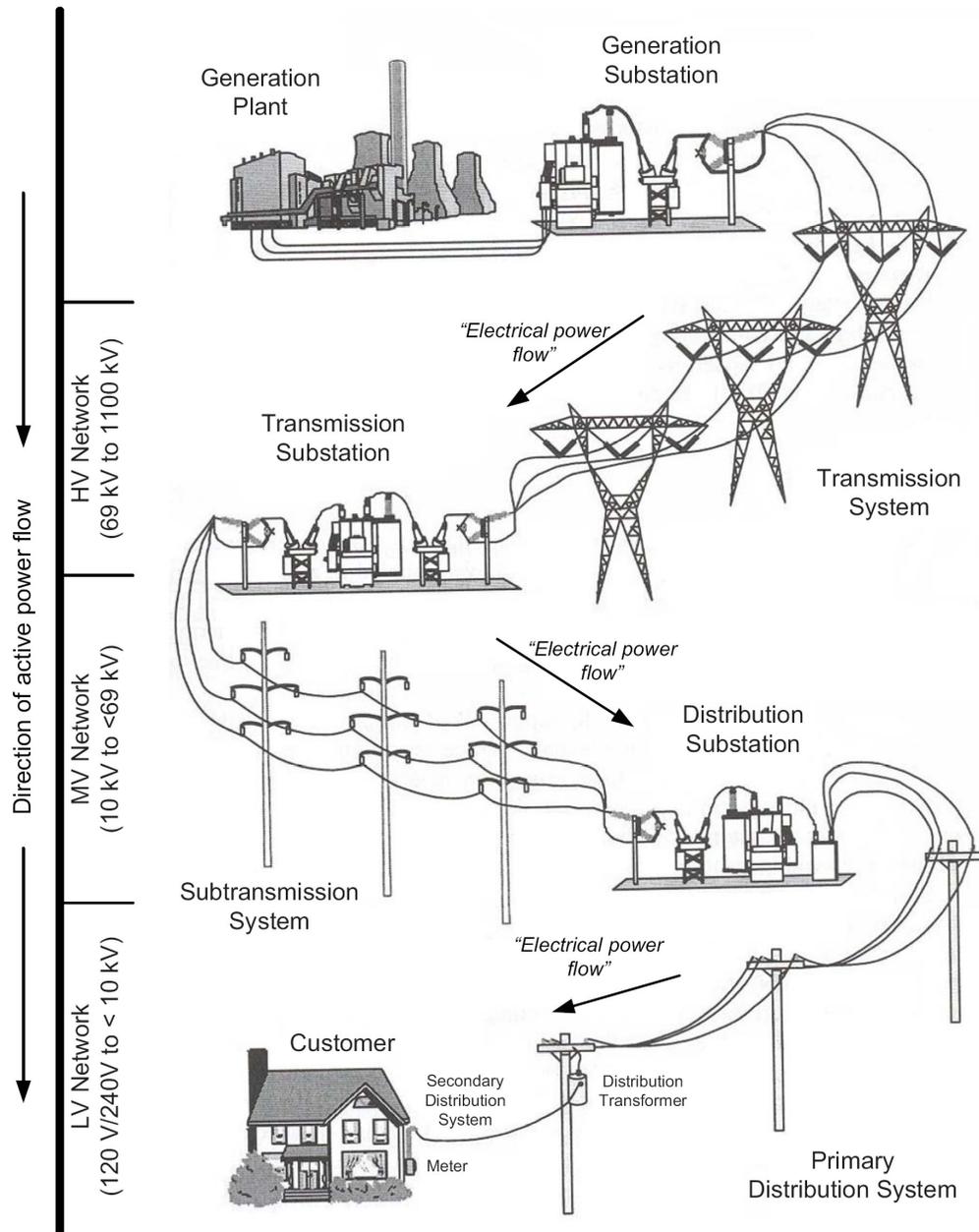


Figure 1.1: A Vertically-Operated power system [17].

Table 1.1: Worldwide installed capacity (GW) on January 1,2000 [74].

Region	Thermal	Hydro	Nuclear	Other/Renew	Total
North America	642	176	109	18	945
Central/South America	64	112	2	3	181
Western Europe	353	142	128	10	633
Eastern Europe/Former USSR	298	80	48	0	426
Middle East	94	4	0	0	98
Africa	73	20	2	0	95
Asia/Oceania	651	160	69	4	884
Total	2175	694	358	35	3362
Percentage	66.6	21.3	11.0	1.1	100

fission is used to release the energy contained in atom nuclei. This energy is then used to create high pressure steam that drives a turbine and a generator.

The use of controllable energy sources permits the robust control of the system power generation and therefore the reliable system operation. In nearly all large power plants synchronous generators are used for the conversion of mechanical energy into electrical energy. These power plants form the so-called *conventional generation (CG)*, with three main characteristics: large-scale, controllability and unified generation technology. In Table 1.1, the worldwide installed capacity for the several types of power generation technologies are presented. We can see that, in the year 2000, CG provided 99% of the total power, of which two thirds originated from thermal (fossil-fuel-fired) power plants.

This vertical structure of the power system brings considerable advantages, such as economies of scale in power generation, reduction of the required generator reserve margins for the individual plants, minimization of the cost of electrical power by shifting generation between units with different prime movers, flattening of the load curve due to the aggregation of loads in the higher system levels that enables a more effective use of the generation equipment, higher energy efficiency in large generating units and operation with a relatively small crew, reduction of the risk of common cause failures due to the use of equipment of various manufacturers and ages, etc. [88].

In spite of these advantages, conventional power generation bears inherent drawbacks. These drawbacks, in relation to special socioeconomic and political reasons that prevailed during recent years, motivated the development of a new, non-conventional power generation scheme, based on the utilization of new and distributed energy resources.

1.2 Problems with CG

1.2.1 Depletion of fossil fuels

As presented in Table 1.1, thermo-electric power generation corresponded to 67% of the worldwide installed capacity in the year 2000, while nuclear power generation provided another 11%. The prime mover for these power plants, i.e. fossil fuels, is extracted from the available natural reserves. However, these reserves are not infinite and will be depleted in the long term. This dictates the gradual incorporation of *sustainable energy sources*

in the power system generation mix, i.e. energy sources which are not expected to be depleted in a time-frame relevant to the human race.

1.2.2 Greenhouse gases emissions

On the other hand, the adverse environmental impacts of the operation of fossil-fuel-fired power plants, such as the global climate change and the greenhouse effect caused by the increase of the CO_2 concentration in the earth's atmosphere, brought further initiatives for the decoupling of the economic activity from fossil primary energy consumption. The Kyoto Protocol to the United Nations Framework Convention on Climate Change was adopted to achieve the 'de-carbonization' of the energy systems of the future based on the use of new and *renewable energy sources (RES)* as primary movers of electrical generation [96].

1.2.3 Nuclear energy

Nuclear energy, although energetically efficient and carbon-free, meets a negative public opinion due to the problem of disposal of the nuclear waste and the fear of the adverse effects of a nuclear accident. Except for a few economically emerging regions of the world, it is safe to observe that nuclear power production, using existing technologies, will decrease in the coming decades as old plants are retired and are not being replaced. Several European countries, such as Germany and Sweden, have enacted laws to accelerate the decommissioning of existing nuclear power plants. However, emerging technologies such as the *pebble bed reactor* technology, which allow for a highly standardized manufacturing of the power plants with modular installed capacities, may revive the nuclear power industry as they will most probably be required within any carbon-free generation mix [74].

1.2.4 Hydro power

Large-scale hydro power plants comprise a valuable renewable energy source alternative. Although large-scale hydro power represents very significant opportunities in several developing regions in the world, in the developed countries the available hydro potential has already been utilized for a large part. Furthermore, the construction of dams and basins for hydro power generation causes the flooding of large areas and affects the flow of rivers, which brings significant environmental consequences and often societal upheaval that go far beyond the national boundaries of the home countries [74].

1.3 Renewable energy sources (RES)

The move towards a non-conventional generation mix in the power system is firstly related to the incorporation of sustainable energy sources. The most prominent subset of sustainable energy is renewable energy, that is commonly defined as 'the energy from an energy resource that is replaced by a natural process at a rate that is equal to or faster than the rate at which that resource is being consumed'. RES capture their energy from existing flows of energy, from on-going natural processes, such as sunshine, wind, wave power, flowing water (hydropower), biological processes such as biomass and anaerobic

digestion, and geothermal heat flow. Most renewable forms of energy, other than geothermal and tidal power, ultimately derive from solar energy. Energy from biomass derives from plant material produced by photosynthesis using the power of the sun. Wind energy derives from wind activity, which is generated by the sun's uneven heating of the atmosphere. Hydropower depends on rain, which again depends on the sunlight's power to evaporate water.

Renewable power generation possesses two characteristics that distinguish it from CG: the *distributed nature* and *uncontrollability*. Although renewable energy is found in huge quantities in nature, it is geographically distributed, presenting a low energy density on each generation site. In order to capture this energy and convert it to electricity, small-scale converters should be spread in many sites in the power system. Their actual power contribution to the system comes from the aggregation of the output of a large number of such generating units. Depending on the level of aggregation, these small-scale power plants may be connected in different voltage levels in the system, including the distribution networks. The use of RES also introduces non-regulated prime energy movers in the power system generation mix, such as wind-, solar-, wave-, and run-of-river-hydro-energy. The power output of such power plants is mainly defined by the prime mover activity and not by the system management. This introduces *power generation uncertainty* in the system.

1.4 Deregulation & liberalization of the energy markets

Historically, the growth in electricity consumption was anticipated by adding new large central power plants, building new transmission lines and extending the traditional distribution systems. However, solving the transmission system capacity problem by erecting new HV lines becomes increasingly difficult, due to the investment cost, the lack of available physical space for expansion and the rejection from the public [74]. This problem, combined with the world-wide trend towards deregulation of the electricity markets and the introduction of new, more effective forms of small-scale generation like combined heat-power plants, micro-turbines, hydrogen and fuel cells [15], created a boost towards the use of distributed energy resources in the system.

An alternative solution to the transmission system capacity problem may come from the connection of generation in the distribution systems. Also, a deregulated environment with open access to the distribution networks provides better opportunities for small-scale units, which require lower capital costs and shorter construction times [47]. Benefits of generating power close to the loads include the use of waste heat for heating or cooling (*combined heat and power (CHP)*, co-generation) and the availability of standby power for critical loads during periods when electricity from the utility is unavailable. Furthermore, a number of small-scale power generation technologies have reached a development stage which allows for large-scale implementation within the existing electric utility systems [74].

1.5 non-Conventional Generation (non-CG)

1.5.1 Types of non-CG

The content of this section is largely based on the Ph.D thesis of M. Reza [76], Department of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, which is one of the Ph.D students working in the same research project.

As discussed in the previous section, the development and implementation of non-CG units is encouraged by environmental, social and political forces. This has stimulated research, promotion, development and increased use of new, renewable, and environmentally-friendly forms of energy. Non-CG comprises alternative energy sources like wind, biomass, sunlight, tidal-, wave- and geothermal energy, and new generation technologies as the *fuel cell*, *combined heat and power (CHP)* cogeneration, the *microturbine generator*, and also small-scale fossil-fuel-fired CG. The rise of non-CG is supported by the advancements in supporting technologies like power-electronic converters and controllers. The basic types of non-CG are briefly presented in the following sections.

Small-scale fossil-fuel-fired power plants

This type of non-CG refers to fossil-fueled power plants within a range of kW up to 100 MW [15,21,47]. The reciprocating engines and combustion turbines are the most common technologies used in this category.

Reciprocating engines are characterized by low capital cost, possible thermal and electrical cogeneration, good modularity and flexibility, and a high reliability. However, such utilization of engine generators creates location-specific environmental issues associated with the equipment's operational characteristics [74]. The emissions can be reduced to some extent by using natural gas. Also, the large number of moving parts leads to noise pollution (which is hard to control) and increases the maintenance cost.

Combustion turbines (gas turbines) are commonly used in industry [21]. In oil industry for example, the associated gas from the oilfield is frequently put into use to generate electricity. The use of natural gas for the combustion turbines results in lower emissions when compared to reciprocating engines. The use of combustion turbines is mostly encouraged by the development of microturbines, that are highlighted in the next paragraph.

Microturbines

A 'micro' gas combustion turbine (microturbine) produces electric power in the range of 25-500 kW. An electrical generator is integrated within the microturbine, that operates at a high rotational speed (50,000 to 120,000 RPM). The electric power is produced with a frequency (in the order) of 10,000 Hz. Therefore, a power-electronic converter is used to interface the generator to the grid. Within the power-electronic interface, the high-frequency electrical power is converted to DC before it is inverted back to low-frequency AC.

Most microturbines make use of natural gas. As a consequence, microturbines are typically characterized by low emission levels. The use of RES such as ethanol is also

possible, and makes the microturbine even more environmentally friendly. In addition, the low capital cost, high efficiency, durability and low noise encourage the use of micro-turbines [30].

Combined heat and power (CHP) plants

Combined heat and power (CHP), also known as cogeneration, is the simultaneous production of electrical power and useful heat [47]. Reciprocating engines, combustion turbines, and microturbines can be used in CHP schemes. CHP generation on a large scale is usually based on fossil fuel. The 'waste heat' resulting from the production of electricity is then used for other applications. With this simultaneous process, the overall efficiency of a CHP plant can be around 85%. In small-scale CHP units, the heat production is dominating the process and electricity is the by-product [21,47].

Fuel cells

Just like a battery, a fuel cell produces an electric current directly from a chemical reaction. However it is designed for continuous replenishment of the reactants consumed; it produces electricity from an external supply of fuel to the anode (usually hydrogen) and oxygen to the cathode as opposed to the limited internal energy-storage capacity of a battery.

One fuel cell only produces a small amount of electricity, and larger amounts can be obtained from a stack of fuel cells [15]. Fuel cells are very efficient, up to 80%. Because the emission is only pure water (in case of hydrogen as the fuel), they are also environmentally friendly. They are modular, portable and produce low noise pollution, because there are no moving parts. In the future, electrical networks can be combined with gas and hydrogen infrastructures. Such a development will further increase the implementation of fuel cells [39].

Geothermal power plants

Geothermal power plants convert the energy contained in hot rock into electricity by using water or an other fluid to absorb the heat from the rock and transport it to the surface of the earth.

It is estimated that the capacity of geothermal energy is nine times its existing installed capacity of about 8 GW. Geothermal energy is often referred to as a form of renewable energy, but because the heat at any location can eventually be depleted, technically it is not strictly renewable. In general, depending on the extraction rate, a field may remain productive over a period of 10 to 100 years. A geothermal power plant usually requires minimal land, as no excavation, transportation and storage are needed. It also produces no pollutant because any unwanted product, if any, can be disposed underground [85].

Biomass power plants

The term '*biomass*' refers to living and recently living biological material that can be used as fuel or for industrial production. It includes all water- and land-based vegetation and trees, municipal solid waste, municipal biosolids (sewage), animal wastes (manures),

forestry and agricultural residues, and certain types of industrial wastes. Biomass is considered a substitute for fossil fuels. Practically, biomass is converted to thermal energy, liquid, solid or gaseous fuels and other chemical products through a variety of conversion processes. These latter forms will then be converted into electricity. In general, biomass is abundantly available and can be considered as a renewable energy source [89].

Small hydro-power plants

A hydro-power plant generates electricity from the movement of a mass of water from a high level to a lower end, where a power house is installed. This water movement can be obtained, for example, from a run-of river or a river with a small impoundment. A small hydro-power plant produces electric power up to 10 MW. Nowadays, hydro power plant technology has reached maturity. A small hydro-power plant has less impact on the environment and ecosystem, when compared to a large hydro-power plant, and is easy to build within a short construction schedule. Once built, its maintenance cost is minimal [83].

Because of the non-availability of large power impounding (dam), the power output of a hydro turbine is practically driven by a direct-captured water flow (the prime mover in the small hydro-power plant). Thus, a simple expression of the power output for a small hydro-plant is [47]:

$$P = QH\eta\rho g \quad (1.1)$$

with P the output power [W], Q the flow rate [m^3s^{-1}], H the effective head [m], η the overall efficiency, ρ the density of water [kgm^{-3}], and g the gravitational constant [ms^{-2}]. For small hydro-power plants, H , η , ρ and g are deterministic and constant. Without significant storage capacity, a small hydro-power plant may experience a very large variation in available water flow (Q) and therefore in the output power (P). Thus, a small hydro-power unit is non-dispatchable.

Wind turbines

A wind turbine generates electricity by extracting kinetic energy from the wind passing through its blades. Wind energy is one of the most promising energy sources to be used for renewable electricity generation. The increasing interest for implementing wind turbines is mostly driven by the infinite availability of wind energy, limited environmental impacts and competitive electricity generation costs [88].

The power generated by a wind turbine (provided that the upstream wind velocity, u , is between the minimal and the maximal values, e.g. $4 < u < 25$ [ms^{-1}]) can be expressed as [47], [88]:

$$P = \frac{1}{2}C_p\rho u^3 A \quad (1.2)$$

with

$$C_p = \frac{1}{2}\left(1 + \frac{u_0}{u}\right)\left[1 - \left(\frac{u_0}{u}\right)^2\right] \quad (1.3)$$

In 1.2 and 1.3, P denotes the output power [W], C_p the power coefficient, u_0 the downstream wind velocity at the exit of the rotor blades [ms^{-1}], ρ the air density [kgm^{-3}], and A the swept area of the rotor disc [m^2].

In practice, ρ , A , and to some extent C_p , are deterministic and constant values. Thus, the power produced by a wind turbine is mainly characterized by the wind velocity. The wind velocity itself has a stochastic behavior; any wind speed can occur at any time. Moreover, when the upstream wind velocity (u) is either below the minimal or above the maximal operating values of the wind plant, e.g. $u < 4$ or $u > 25$ [ms^{-1}], the output power equals zero. As a result, a stochastic output power will be generated.

Photovoltaics

Photovoltaic (PV) power generation systems convert sunlight directly into electricity [72]. A PV cell consists of two or more semiconductor layers of specific physical properties. These layers are arranged in such a way that when the PV cell is exposed to sunlight, the photons cause the electrons to move in one direction (crossing the junctions of the layers) and a direct current (DC) is generated.

Currently, PV energy cost is still high. However, the capital cost of PV modules per Watt of power capacity have declined in the past decades. PV implementation is encouraged by the infinite availability of solar energy, long life cycle and simple maintenance (since there are no moving parts), high modularity and mobility, and short design, installation and start-up time of a new plant.

The power generated by a PV module is given by [72]:

$$P = \eta \times (E_{ed} \times A_{PV_{total}} + E_{es} \times A_{PV_{withsun}}) \quad (1.4)$$

where

$$A_{PV_{withsun}} = (\vec{S} \times \vec{P}) \times A_{PV_{total}} \quad (1.5)$$

and

$$\vec{S} = [S_x S_y S_z], |\vec{S}| = 1 \quad (1.6)$$

$$\vec{P} = [P_x P_y P_z], |\vec{P}| = 1 \quad (1.7)$$

$$\vec{S}_x = \cos(\theta) \times \cos(\alpha_{sun}) \quad (1.8)$$

$$\vec{S}_y = \cos(\theta) \times \sin(\alpha_{sun}) \quad (1.9)$$

$$\vec{S}_z = \sin(\theta) \quad (1.10)$$

$$\vec{P}_x = \cos(\beta) \times \cos(\alpha_{panel}) \quad (1.11)$$

$$\vec{P}_y = \cos(\beta) \times \sin(\alpha_{panel}) \quad (1.12)$$

$$\vec{P}_z = \sin(\beta) \quad (1.13)$$

In the above equations P denotes the power extracted from the sunlight [W], η the efficiency of the solar panel, E_{ed} and E_{es} the diffuse- and the direct-horizontal irradiance [Wm^{-2}], \vec{S} and \vec{P} the solar- and panel-orientation, θ and α_{sun} the altitude- and azimuth-angle of the sun [rad], and β and α_{panel} the altitude- and the azimuth-angle of the panel [rad].

In practice, $A_{PV_{total}}$, η , β and α_{panel} are deterministic and constant. Therefore, the generated electricity is characterized by E_{ed} , E_{es} , θ and α_{sun} . The altitude- and the azimuth-angle of the sun (θ and α_{sun}) have daily and seasonal patterns, whereas the characteristics of E_{ed} and E_{es} are intermittent. Weather changes and cloud movement, for example, strongly influence the values of E_{ed} and E_{es} and consequently the generated electricity. The power generation of PV power plants is therefore stochastic.

Tidal power plants

Tidal energy is derived from the gravitational forces of attraction that operate between the earth and the moon and between the earth and the sun. Energy is extracted either directly by harnessing the kinetic energy of currents due to the tides or by using a basin to capture potential energy from the difference in height of a rising and falling mass of water. To generate electricity, tidal flow is extracted by means of propellers with large diameters. In the latter technique, a huge dam, called a 'barrage' is built across a river estuary. When the tide goes in and out, the water flows through tunnels in the dam. The ebb and flow of the tides can be used to turn a turbine. When the tides comes into the shore, they can be trapped in reservoirs behind dams. Later, when the tide drops, the water dam can be used like in a regular operation of a hydroelectric power plant [86].

Tidal power is a renewable energy source. Tidal power plants produce no pollutant. They also cause no fundamental change of the natural rhythm of the tidal cycle and no inundation of the adjacent area. These factors encourage the implementation of tidal power plants. However, building a tidal power plant has to be planned carefully, considering the potential ecological impacts.

The power output of a turbine operating in flowing water is given by [93]:

$$P = \frac{1}{2}\rho AC_p u^3 \quad (1.14)$$

In 1.14, P denotes the output power [W], ρ the density of the fluid [kgm^{-3}], A the area of the flow interrupted by the device [m^2], C_p the power coefficient of the device (the percentage of power that the turbine can extract from the water flowing through the turbine), and u the velocity of the water [ms^{-1}].

For a tidal power plant, ρ , A , and C_p are deterministic and constant. Therefore, the output power P depends on the velocity of the water u . Thus, the tide, which is not continuous, is the only factor that affects the generating activity of a tidal power plant. This makes the tidal power generation non-dispatchable.

Wave power plants

Waves are generated on the surface of oceans by wind effects which in turn result from the differential heating of the earth's surface. Wave energy is complementary to tidal power, it uses the essentially up-and-down motion of the sea surface (wave power), instead of using the energy of the sea rushing back-wards and forwards (tidal power). A wave power plant extracts wave energy and converts it into electricity.

The wave power plant is promoted as electricity generation that is available in abundance throughout the world; it is clean and non-polluting, renewable, and suited to elec-

Table 1.2: Basic characteristics of CG and non-CG.

	CG	non-CG
Size	large	small, medium
Dispatchability	dispatchable	non-dispatchable
Generation technology	synchronous generator	different types

trify remote communities. However, just like a tidal plant, the erection of a wave power plant should be planned carefully, so that the ecological impacts are minimized.

The power production of a wave power plant can be assessed using [64]:

$$P_{abs} = \alpha A_w H_s^{1.5} \quad (1.15)$$

In 1.15, P_{abs} denotes the average absorbed power [W], A_w the float water plain area [m^2], H_s the significant wave height [m], and α is a coefficient that equals $0.166 [kgm^{-1.5}s^{-3}]$ in an ideal condition. For a wave power plant, α and A_w are deterministic. The output power depends practically on the wave height (H_s), that is not constant nor dispatchable. Hence, a wave power plant is non-dispatchable.

1.5.2 Characteristics of non-CG

Based on the analysis of the different non-CG technologies, we may epitomize its basic characteristics, that distinguishes it from traditional CG, as follows:

1. *Size*: whereas CG comprises mainly large central power plants that are connected to the HV networks, non-CG units are small- to medium- scale generators that, due to their size, are connected to different voltage levels in the system.
2. *Dispatchability*: CG power plants have controllable prime energy movers and they are centrally dispatched by the system operator or generation company. The non-CG units are either *locally dispatched* or *non-dispatchable*, in the case they make use of non-controllable primary energy sources.
3. *Generation technology*: in nearly all CG power plants, synchronous generators are used for the conversion of mechanical energy into electrical energy. In the case of non-CG, new technologies are introduced for the generation of electrical power, including other types of electrical generators and grid coupling via power-electronic converters.

These characteristics are summarized in Table 1.2. Each type of power generation that holds at least one of the characteristics of non-CG may be defined as non-CG.

1.5.3 Distributed Generation (DG)

In the literature, it is a usual practice to use the first characteristic of Table 1.2 (size), in order to identify non-CG units. In this case, the term *Distributed Generation (DG)* is used. This characteristic determines subsequently the point of connection to the system

(MV-LV networks), leading to the following definition: '*DG is an electric power source connected directly to the distribution network, or on the customer site of the meter*' [1]. CIGRE Working Group 37.23 has defined DG as electrical generation that is *not centrally planned, not centrally dispatched and connected to the distribution network* [21]. A DG unit usually produces electric power well below 100 MW [47]. However, this definition excludes non-CG units that are directly connected to the HV network, such as for example large wind parks. A large wind park cannot be considered as DG; but, on the other hand, it is non-CG, due to its non-dispatchability.

In essence, DG is a subset of non-CG. In this thesis, the terminology DG is used when referring to non-CG units that are connected to the distribution networks.

1.5.4 Stochastic Generation (SG)

Based on their dispatchability, two types of non-CG units can be distinguished:

- *Dispatchable non-CG*: several non-CG technologies enable the power units to be *locally dispatched*, i.e. the unit operator can regulate the power output of the unit by controlling the supplied primary energy sources (or fuels). Small fossil-fuel power plants, biomass power plants, geothermal power plants, fuel cells and CHP plants belong to this category.
- *Non-dispatchable (stochastic) non-CG*: power plants that make use of non-dispatchable prime energy sources for the electricity production. In this case, the operator cannot dispatch the units because the primary energy source can not be controlled (stochastic). The power output of the unit is defined by the availability of the prime mover[†]. The following types of generation can be considered as *Stochastic Generation (SG)*: small (run-of-the-river) hydro, wind turbines, photovoltaics, tidal power plants, wave power plants and CHP plants.

A note should be made with regard to geothermal and CHP power plants. The geothermal primary energy source is not as flexible as fossil fuels for dispatching the generator units [85]. Also CHP units may be not that flexible on their electricity output [21]. A CHP unit can be classified as either dispatchable or non-dispatchable depending on its operation scheme. As a dispatchable unit, a CHP is dispatched according to the electrical load while the heat production is non-dispatched. On the other hand, a CHP can be operated to meet the heat load with the electrical power output as the 'by-product'. In this case, the electrical power output is practically non-dispatchable [47]. The choice of the CHP operation scheme depends on commercial/economic considerations [76]. As a summary, Table 1.3 lists the various types of non-CG and their classification as dispatchable and stochastic generation.

We refer to SG as power generation driven by an uncontrolled prime mover. Although it mainly refers to RES, several types of renewable generation are not SG, such as large hydro, biomass and geothermal power plants. On the other hand, heat-load-driven CHP plants belong to SG. The definition of SG stresses the non-controllability of the prime mover and this is a fundamental difference when compared to CG; SG implies *power generation uncertainty*.

[†]Reducing the power production is, to some extent, possible. However, normally most of the stochastic generation is operated in such a way that the electricity production is maximized.

Table 1.3: Classification of non-CG based on dispatchability.

Generation Technology	Dispatchable	Stochastic
Small-scale Fossil-Fuel-Fired Power Plants	✓	
Microturbines	✓	
Combined Heat and Power (CHP) Plants	✓	✓
Fuel cells	✓	
Geothermal power plants	✓	
Biomass power plants	✓	
Small hydro-power plants		✓
Wind turbines		✓
Photovoltaics		✓
Tidal power plants		✓
Wave power plants		✓

1.6 Horizontally-Operated Power System

The large-scale implementation of non-CG units leads to a transition towards a new, 'horizontal' power system structure. In this system structure, in addition to the large CGs connected at the HV networks, medium- to small-scale units are also connected at the MV and LV networks (distribution systems). These units are either locally dispatched or stochastic. This uncontrolled power infeed from the lower voltage levels of the system leads to a radical change of the former vertical structure of the system. The power for example can flow not only 'vertically', i.e. from the higher to the lower voltage levels, but also 'horizontally', i.e. from one MV or LV network to another or from a generator to a load within the same MV or LV network. A power system that follows these operational principles is called a *horizontally-operated power system (HOPS)*.

In a HOPS, the distribution systems are *active* networks; besides distributed loads, they also have generation installed. Fig. 1.2 shows the transition from a passive to an active distribution network structure. The power flow between the transmission network and the active distribution network is no longer uni-directional (down) but can be bi-directional.

The transmission system acts as an energy bus that interconnects the different active distribution systems and the remaining large CG units. When the penetration level of non-CG increases to such an extent that a major part of the system load can be supplied by those units, the number of CG units in the system will decline. This 'Vertical-to-Horizontal' transformation of the power system is presented in Fig. 1.3.

1.7 Objectives and limitations

The analysis of the system operation under the 'vertical-to-horizontal' transformation of the power system, caused by a large-scale implementation of non-CG/SG units in the system necessitates the incorporation of generation uncertainty into the power system modeling. This thesis focuses on the development of new methods for the power system uncertainty analysis and the investigation of the operation of a horizontal power system. In particular, the following objectives are set:

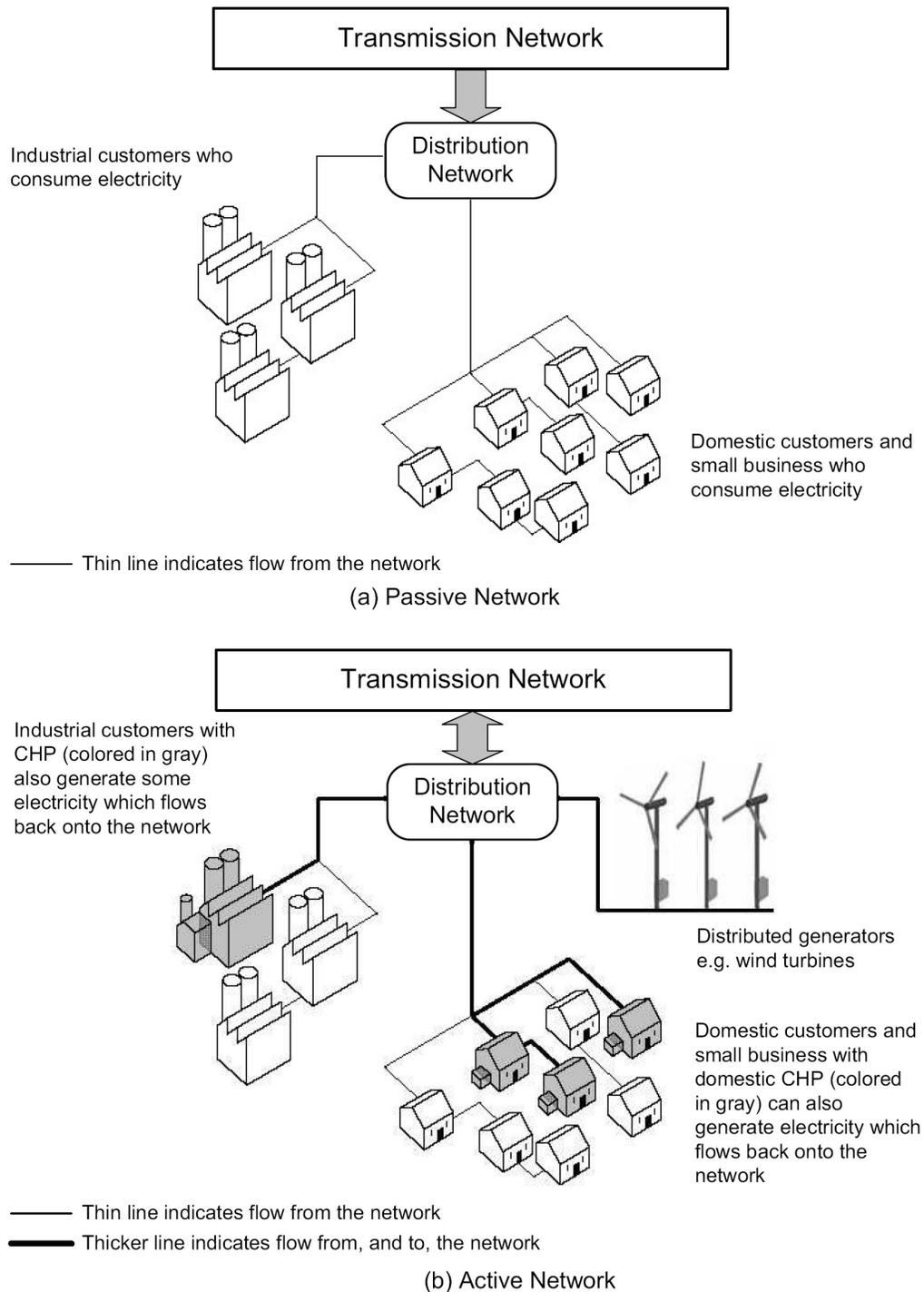


Figure 1.2: A conventional (passive) distribution network (upper graph - a) and an active distribution network (lower graph - b) [76].

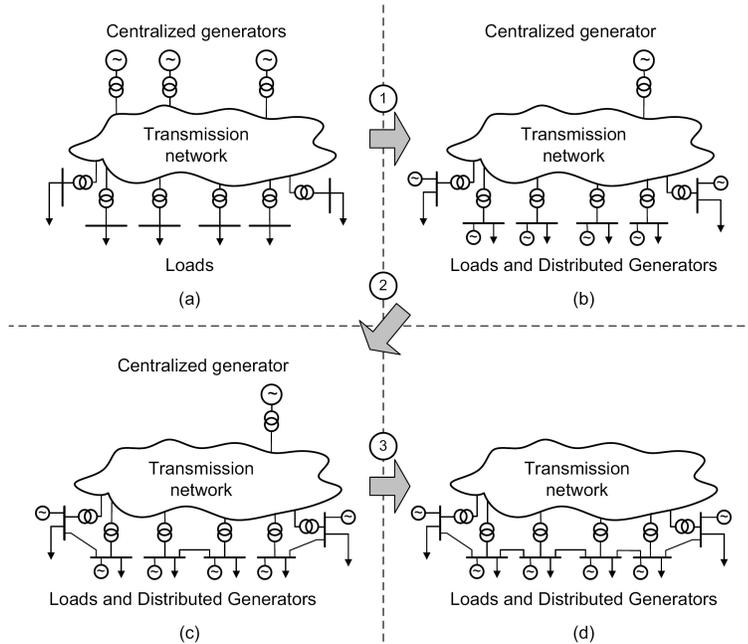


Figure 1.3: 'Vertical-to-Horizontal' transformation of a power system [76].

- The development of methods for the modeling of stochastic generation in the power system.
- The investigation of the impact of large-scale penetration of stochastic generation in the power system, in two parts:
 1. *Active distribution networks*: analysis of the impact of the large-scale incorporation of DG in MV and LV networks, investigation of the transformation to an active network and the impact on the voltage in the system.
 2. *Transmission system as energy bus*: analysis of the power flows in a horizontally-operated power system.

These issues will be addressed in this thesis. This research is unique, since it is the first time that the solution of the problem of correlated stochastic resources in the power system is addressed. The focus is on the steady-state system operation; dynamic system studies are beyond the scope of this thesis. For the steady-state analysis, a memory-less power system is considered, i.e. the state of the system on each moment does not depend on the previous state of the system. This corresponds to a power system without large-scale storage.

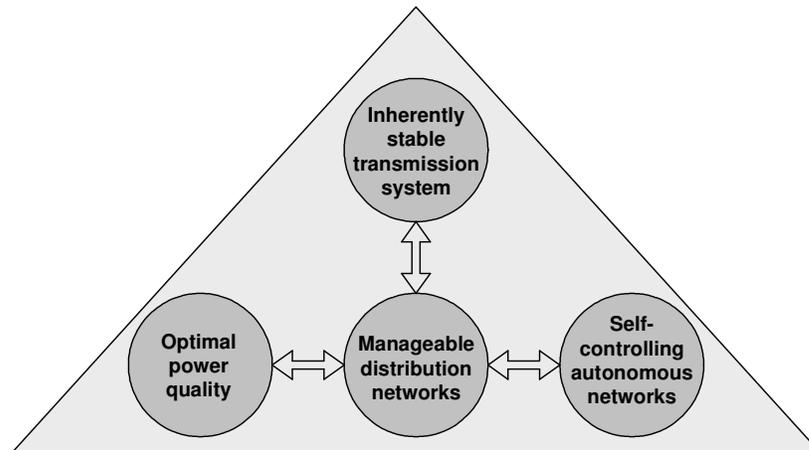


Figure 1.4: The four parts of intelligent power systems research project.

1.8 Research Framework: 'Intelligent Power Systems'

The research presented in this thesis has been performed within the framework of the 'Intelligent Power Systems' project [31]. The project is part of the IOP-EMVT program (*Innovation Oriented research Program - Electro-Magnetic Power Technology*), which is financially supported by SenterNovem, an agency of the Dutch Ministry of Economical Affairs. The 'Intelligent Power Systems' project is initiated by the Electrical Power Systems and Electrical Power Processing groups of the Delft University of Technology and the Electrical Power Systems and Control Systems groups of the Eindhoven University of Technology. In total 10 Ph.D. students, who work closely together, are involved in the project.

The research focuses on the effects of the structural changes in power system generation and consumption that are taking place, like for instance the large-scale introduction of distributed (renewable) generators [77]. Such a large-scale implementation of distributed generators leads to a gradual transition from the current 'vertically-operated power system', which is supported mainly by several big centralized generators, into a future 'horizontally-operated power system', having also a large number of small to medium-sized distributed (renewable) generators. The project consists of four parts (as illustrated in Fig. 1.4).

The first part investigates the influence of uncontrolled decentralized generation on the stability and dynamic behavior of the transmission network. As a consequence of the transition in the generation, less centralized plants will be connected to the transmission network as more generation takes place in the distribution networks, whereas the rest is possibly generated further away in neighboring systems. Solutions that are investigated include the control of centralized and decentralized generation, the application of power-electronic interfaces and monitoring of the stability of the system.

The second part focuses on the distribution network, which becomes 'active'. Technologies and strategies have to be developed that can operate the distribution network in

different modes and support the operation and robustness of the network. The project investigates how the power-electronic interfaces of decentralized generators, or between network parts, can be used to support the grid. Also, the stability of the distribution network and the effect of the stochastic behavior of decentralized generators on the voltage level are investigated.

In the third part autonomous networks are considered. When the amount of power generated in a part of the distribution network is sufficient to supply the local loads, the network can be operated autonomously, but as a matter of fact remains connected to the rest of the grid for security reasons. The project investigates the control functions needed to operate the autonomous networks optimal and secure.

The interaction between the grid and the connected appliances has a large influence on the power quality. The last part of the project analyzes all aspects of the power quality. The goal is to support the discussion between the polluter and the grid operator who has to take measures to comply with the standards. The realization of a power quality test lab is an integral part of the project.

1.9 Outline of the Thesis

The thesis is organized as follows:

- The 2nd chapter presents the modeling principles for the developed approach, i.e. the *steady-state uncertainty analysis*. The different approaches presented so far in literature for the solution of this problem are presented and compared. It is shown that the most suitable methodology for the system modeling is the use of stochastic simulations (Monte-Carlo simulation), due to the existence of complex interdependencies between the system inputs. The problem definition leads to a multivariate uncertainty analysis problem. The modeling procedure is decomposed into two basic components, the modeling of the marginal distributions and of the stochastic dependence structure. The impact of each of these components is discussed: the marginal distributions define the power output of each unit, while the stochastic dependence structure defines the sum of all units, acting as the stochastic generation dispatch.
- In the 3rd chapter the modeling principles for bivariate dependence are presented. The chapter is organized in terms of examples from the modeling of system loads (normal distributions) and stochastic generation. Through these examples, the fallacies concerning the extension of the 'traditional' methods used for load modeling to the modeling of stochastic generation are discussed. New concepts are presented, namely the *copula theory* for the bivariate modeling of dependent r.v.
- The 4th chapter presents the techniques for multidimensional dependence modeling. The *Joint Normal Transform Methodology* is presented as the main technique for the modeling of correlated inputs when related measurements are available. The application of the methodology is presented through a comprehensive example, namely the penetration of wind power in the power system of The Netherlands.

- In Chapter 5, the concept of *stochastic plants* is presented as a means of model reduction. Stochastic plants are clusters of stochastic generators/loads that behave in a similar manner. A simple methodology may be used for the modeling of such clusters, namely the concept of perfect correlation or comonotonicity. This case corresponds to the worst-case scenario for the aggregate power output of the cluster. The use of the concept of stochastic plants leads to a risk-averse model reduction.
- In Chapter 6, two basic power system planning problems are investigated: the integration of stochastic generation in a distribution system with emphasis to the impact to the voltage profile of the system and the integration of stochastic generation in a bulk power system. In the distribution system planning, the voltage control by stochastic generators is analyzed and different control strategies are investigated for the case of wind power generation. In the bulk power system planning, the horizontal operation of the power system is investigated for the case of wind power generation, namely the bi-directional power flows in the system.
- The conclusions and recommendations for future work are in Chapter 7.

Chapter 2

Power System Steady-State Uncertainty Analysis

As presented in chapter 1, the horizontal operation of the power system derives from the incorporation of non-CG power generators in the system. An appropriate methodology should be adopted for the unified modeling of the power generation uncertainty in the system. The modeling principles for such an approach are presented in this chapter, under the terminology '*power system steady-state uncertainty analysis*'.

2.1 Power system steady-state: a 'snapshot' approach

A power system is predominantly in steady-state operation or in a state that can be regarded with sufficient accuracy as steady-state. In practice, there are always small load/generation changes, switching actions and other transients occurring, so that in a strict mathematical sense most of the variables are varying in time. However, these variations are most of the time so small, that the use of an algebraic (non time-varying) model of the power system is justified. The system steady-state operation may therefore be regarded as the transition between consecutive steady-state 'snapshots'; each of these snapshots corresponds to a different set of inputs (load, generation).

The system analysis corresponds to the investigation of all respective snapshots, which is equivalent to the analysis of the system steady-state for the set of all possible inputs. Two modeling steps can be distinguished:

1. *Deterministic System Model (DSM)*: the DSM represents the way in which the system variables are combined to produce the system outputs, referring to all physical constraints and control actions that take place during the system operation. This model is the steady-state system model.
2. *Stochastic System Model (SSM)*: the SSM describes how the uncertain system inputs behave and interact; this relates to how mechanisms outside the system affect the system behavior (weather conditions, customer behavior, etc.).

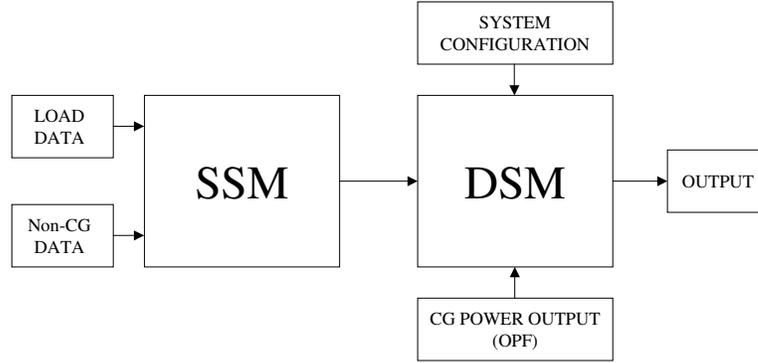


Figure 2.1: Schematic diagram of the system analysis.

In Fig. 2.1, a schematic diagram for the system analysis is presented. As may be seen, the data for the uncertain system inputs (loads and non-CG) are fed into the SSM, providing the set of stochastic system inputs, i.e. the different snapshots of operation. These are passed on to the DSM, together with the system configuration data and the non-stochastic system inputs (CG setpoints obtained by an optimal power flow algorithm), so that the respective output is obtained.

2.2 Deterministic System Model (DSM)

2.2.1 Steady-State System Model

The steady-state or *load flow* analysis is the fundamental study for power systems analysis. The exact formulation of this problem concerns the determination of real and reactive power flows in each line (branch) of the power system [46]. The data used are the active and reactive power consumed by each load, as well as the real power generation and voltage magnitudes at the CG buses. The non-CG units are generally modeled as negative loads, except when they are engaged to perform voltage control. The detailed load-flow formulation is presented in Appendix A.

Using this model, the power system is represented as a system of non-linear static equations. In particular, the system states \mathbf{X} (nodal voltage angles and magnitudes) and outputs \mathbf{Z} (active and reactive power flow in each system branch), can be calculated by a specific set of inputs \mathbf{Y} (active and reactive power injections at each system node, i.e. load \mathbf{L} minus generation \mathbf{G} : $\mathbf{Y} = \mathbf{L} - \mathbf{G}$) [46]:

$$\begin{cases} \mathbf{Y} = \mathbf{g}(\mathbf{X}) \\ \mathbf{Z} = \mathbf{h}(\mathbf{X}) \end{cases} \quad (2.1)$$

As mentioned, the steady-state system model (DSM) gives a ‘snapshot’ of the system operation, that corresponds to either a specific instant in time or, equivalently, to a potential outcome due to the stochasticity of the system inputs.

2.2.2 Deterministic analysis and vertical power system

Traditionally, the power system steady-state analysis is treated as a deterministic problem. In particular, instead of investigating all possible system states, the system design is based on specific worst-case scenarios. Since in the vertical power system the power flows in one direction (from the transmission towards the distribution level) and the uncertainty of the system inputs is due to the load, the worst-case operational scenario corresponds to the maximum system loading. The usual practice has been to investigate these situations and omit the SSM.

However, in the case of a horizontally-operated power system, such an approach may prove insufficient. The connection of stochastic power generation in different voltage levels in the system results in bidirectional power flows in the system lines; this complicates the system analysis. It is evident that the definition of the worst-case operational scenario for the system becomes a non-intuitive process, due to the additional power generation uncertainty. Furthermore, the stochasticity of both consumption and generation has an enormous impact on the system, that can be analyzed only by incorporating the SSM in the analysis.

2.3 Stochastic System Model (SSM)

With the SSM we try to quantify the uncertainty related to the system inputs. In general this requires a prohibitive amount of calculations; for a network of N loads and SG-units, each taking k different values, a total of k^N deterministic load flow calculations should be performed*. Therefore, an appropriate uncertainty analysis methodology should be employed.

2.3.1 Uncertainty analysis methodologies: general

The goal of an uncertainty analysis methodology is the quantitative representation of uncertainty. In general, the following steps may be recognized in such methodologies:

1. *Data collection*: respective data are collected corresponding to each uncertain system input.
2. *Representation/quantification of uncertainty*: an appropriate mathematical methodology is used for the representation and quantification of the inputs' uncertainty.
3. *Communication of uncertainty to DSM*: the uncertainty in the system inputs is propagated through the DSM and results are obtained.

2.3.2 Uncertainty in power system analysis

In power system analysis, two main sources of uncertainty may be recognized:

*For example, for a small system with a total of $N = 10$ loads and SG-units whose output takes $k = 5$ possible values, the number of load flow calculations rises to $5^{10} = 9765625!!!$

1. *Inputs uncertainty*: this is the uncertainty in the system load and generation. The generation uncertainty corresponds to both the output uncertainty due to the prime-mover stochasticity and the loss of a generator due to a random failure.
2. *Configuration uncertainty*: this uncertainty is related to the loss of one (or more) of the major transmission system components due to a random failure. This contingency leads to a new system topology and the problem degenerates to the modeling of the inputs uncertainty [25].

Starting from the early seventies, different computational methodologies were introduced for power system uncertainty analysis and a large number of research papers was published [67].

2.4 Literature review

The main approaches found in the related literature, are the *Probabilistic Uncertainty Analysis* [11] and the *Fuzzy Arithmetic Analysis (Possibilistic Uncertainty Analysis)* [97, 59]. These methods are used respectively for the modeling of the two basic types of uncertainty related to power system studies, the *quantitative uncertainty*[†] and the *qualitative uncertainty*[‡] [10].

2.4.1 Fuzzy arithmetic analysis

Fuzzy logic is derived from fuzzy set theory dealing with reasoning that is approximate rather than precisely deduced from classical predicate logic. Fuzzy arithmetic analysis is used for modeling qualitative uncertainty. In particular, using the concept of degree of membership of a value to a set, it is possible to establish the notion of *fuzzy sets* and *fuzzy arithmetic* as a representation of vagueness or uncertainty of a non-random nature [97].

The notion of fuzzy methods originates from an extension of the notion of a set, where membership in a set is permitted to be something other than a binary variable (member/non-member). A *set membership function (smf)* $p(x)$ is defined so that a value $0 < p(x) < 1$ denotes the *belief in the possibility (degree of belief)* that x belongs to the set. The extreme values 1 and 0 are used to denote that x belongs to the set or not. Many applications of this methodology can be found in the literature on power systems planning based on a fuzzy solution of the load flow problem [58, 94].

2.4.2 Possibilistic vs Probabilistic Analysis

The co-existence of these two methodologies has been the source of confusion in uncertainty analysis. Degrees of belief are often confused with probabilities. However, they are conceptually distinct; fuzzy truth represents membership in vaguely defined sets, not likelihood of some event or condition.

[†]Quantitative uncertainty is the uncertainty that is quantifiable in numerical terms by a mathematical function with deterministic parameters.

[‡]Qualitative uncertainty is uncertainty that is initially expressed in vague, non-numeric (usually verbal) terms such as '*approximately equal to*' or '*a small percentage*'.

To illustrate the difference, consider this example [95]: Bob is in a house with two adjacent rooms: the kitchen and the dining room. In many cases, Bob's status within the set of things 'in the kitchen' is completely plain: he's either 'in the kitchen' or 'not in the kitchen'. What about when Bob stands in the doorway? He may be considered 'partially in the kitchen'. Quantifying this partial state yields a fuzzy set membership. With only his big toe in the dining room, we might say Bob is 99% 'in the kitchen' and 1% 'in the dining room', for instance. No event (like a coin toss) will resolve Bob to being completely 'in the kitchen' or 'not in the kitchen', as long as he's standing in that doorway. Fuzzy sets are based on vague definitions of sets, not randomness.

In power system analysis, we may interpret this difference by an example of a planning problem that involves the analysis of the impact of the incorporation of two types of SG in the system, namely wind and solar power. The uncertainty concerning the system planning in this case is twofold:

1. *Qualitative*: this uncertainty refers to the vagueness concerning the type of SG in each generation site (solar or wind). The system designer may have a degree of belief concerning the possible type that is going to be installed and this type of uncertainty may be modeled using the possibilistic approach.
2. *Quantitative*: this uncertainty refers to the output of a certain stochastic generator on a specific generation site. This uncertainty may be quantified in numerical terms by the statistical analysis of respective data, or by expert judgement.

Although there is a distinctive line between these two types of uncertainty, it is a usual fallacy to interpret quantitative uncertainty using a possibilistic approach. A recent example may be found in [80], where a fuzzy approach is presented for the modeling of correlated stochastic inputs, referring to output of small hydro power plants, solar systems, wind parks and loads. In this case, a qualitative representation is used to quantitatively assess the power flows in the system, leading to fallacies concerning the interpretation of the results.

The confusion concerning the borderline between the two methodologies derives from the absence of sound operational definitions concerning possibility and fuzziness [22]. In the problem definition, it should be made clear what type of uncertainty is concerned. The uncertainty of the output of SG or load in a specific system is quantitative and probabilistic analysis is the appropriate modeling approach.

2.5 Probabilistic steady-state uncertainty analysis

According to the probabilistic uncertainty analysis, each uncertain system input is represented as a *random variable (r.v.)* with a specific *probability density function (pdf)* [71]. In the related literature, two main probabilistic approaches can be found, namely the *Analytical methods* and the *Stochastic Simulations (Monte-Carlo Simulation - MCS)*.

2.5.1 Analytical methods

Computational efficiency is the basic advantage offered by an analytical solution of the steady-state problem, compared to MCS. The lack of sufficient computational power

provided the impetus from the early 1970s for numerous research papers dealing with an analytical formulation of the probabilistic steady-state analysis [82]. However, in order to reach an analytical formulation of the problem, a number of simplifications are necessary:

1. *Linearization of the system model*: the linearization of the steady-state system model (equation 2.1) permits the representation of the system outputs as a linear combination (weighted sum) of the system inputs. The linearization is performed around an operating point that corresponds to the mean value of the system inputs: $Y_\mu = g(\mu(X))$ and $Z_\mu = h(\mu(X))$. This approximation is accurate for cases where the dispersion of the system inputs is limited around the mean value. When this is not the case, the input data will be transformed less accurately by this linear model, particularly in the tail regions as they are the furthest away from the point of linearization.
2. *Independence*: the system inputs are assumed to be statistically independent. This assumption, in combination with the previous one, permits the computation of the system outputs by the use of a series of convolutions or by the application of the Gram-Charlier expansion method and the computation of the cumulants of the system outputs by the ones of the system inputs, based on their invariance to linear transformations.
3. *Normality*: the system inputs are assumed to be normally distributed. This assumption permits the use of linear correlation for the representation of the dependence structure between the respective r.v. In this case, linearly dependent r.v. may be incorporated in the analysis and, in combination with the first assumption, the output distributions may be obtained analytically by convolution or using orthogonalization techniques (Gram-Schmidt orthogonalization).

The choice of assumptions determines the analytical solution. Hereunder, the basic literature on the problem is presented, together with the basic assumptions, while in the following chapter, the impact of these assumptions (especially normality and independence) is discussed in detail.

The initial probabilistic analytical formulation of the power system steady-state analysis was introduced in the early seventies [16, 5], under the terminology '*Probabilistic Load Flow (PLF)*'. The two main assumptions for this approach are linearization and independence. Due to the linearization of the system model, the system states (nodal voltages) and outputs (line power flows) can be expressed as a weighted sum of the system inputs (nodal power injections). Since the r.v. representing the load at each system bus are assumed to be independent, the pdf of the circuit flows can be computed with a series of convolutions, using appropriate numerical techniques, basically *Fast Fourier Transform (FFT)* algorithms [8]. The initial DC formulation of the problem was followed by a series of papers dealing with the more detailed AC formulation [4, 9, 3, 7]. Some typical extensions and improvements of the approach can be found in [25, 28, 26]. In [6], a multi-linear approximation for the system model is presented, in order to increase the accuracy of the algorithm.

The assumption of independence of the nodal loads is, however, quite unrealistic. In [9] and [2], a linear dependence between normally distributed nodal power injections in PLF is suggested, which was further investigated in [7] and [27]. According to this approach,

the dependence between demands in the system is assumed to be close to perfect linear. Therefore, each normal load can be split in two components, one that corresponds to mutual linear correlation between loads and one that represents a superimposed independent normal noise. All linearly correlated components are summed to one normal r.v. and convolution techniques can be further employed for the derivation of the output distributions.

Another formulation treating the same problem appeared in 1975 under the terminology '*Stochastic Load Flow (SLF)*' [37]. In SLF analysis, the load and generation at an instant of time are treated as r.v. The deterministic equations were modified by adding a random normally distributed noise vector whose covariance defines the variations of the power injections around their base values. Assuming applicability of the *Central Limit Theorem (CLT)*[§], the state and output distributions are considered to follow normal distributions; therefore, only their variance should be computed. This approach exhibits the advantage of using the detailed system model and also modeling the correlation between the loads at any two buses. However, Monte-Carlo simulations indicate that normality of the system states and outputs is an unreliable hypothesis [27, 7, 6]. For this reason, Sauer and Heydt [81] have proposed the use of higher moments (third and fourth) for a more accurate representation of the pdfs.

Another method for treating the correlation among bus loads and the generation dispatch procedure is proposed in [54]. The model assumes normal distribution of bus loads and a linearized economic dispatch model. The circuit flows and bus voltages are expressed as a linear combination of the bus loads only, and are assumed to be normally distributed. In [55] the approach was extended to address the issue of the nonlinearities of the power system model, the economic dispatch and availability of the generating units. In [56] a quadratic approximation of the steady-state system model was introduced. This approach was further developed in [90] and [91], where the concept of the *nonconforming stochastic electric load* is developed. According to the authors, a typical probabilistic load model is a conforming electric load model, i.e. a specific bus load is a fixed percentage of the total system load. Statistically, this means that the bus loads are correlated one hundred percent[¶]. For a more realistic representation of the electric load, it is necessary to represent the bus electric load as a non-conforming load. A non-conforming load model assumes variable correlation between various bus loads and is obtained as a linear combination of conforming loads. The basic assumption for the applicability of this approach is the normality of the power injections in the system nodes.

Another approach for the probabilistic solution of the steady-state problem has been presented based on the use of the concept of cumulants and Gram-Charlier expansion theory [98]. According to this approach, the output distribution may be approximated as a weighted sum of standard normal distributions; the weighting coefficients in this sum may be expressed as a function of the cumulants of the distribution. The method proposes further the linearization of the system model and independence between the system inputs. In this case, the output distributions are defined as a linear combination of the independent system inputs and based on the basic properties of cumulants, the cumulants of the output distribution may be calculated as a linear combination of the respective in-

[§]According to the CLT, under general conditions the distribution of the sum of n r.v. approaches a normal distribution as n increases [71].

[¶]This '*one hundred percent*' correlation actually corresponds to the case of linear dependence.

put ones. These cumulants are further used to derive the weighting coefficients for the Gram-Charlier expansion of the output distribution.

An alternative approach for the treatment of linearly correlated system inputs has been presented in [18] and further developed in [14]. Here, the Gram-Schmidt orthogonalization is used in order to transform the linearly correlated system inputs into a weighted sum of independent r.v. By linearizing the system model, these transformations of the system inputs may be used to express the system outputs as a weighted sum of independent (orthogonal) r.v. Further, the above methods may be used for the derivation of the system output distributions. The assumption of linearly correlated inputs is equivalent to the case of normality. This issue will be further discussed in the following chapter.

2.5.2 Stochastic Simulations (MCS)

The *Monte-Carlo Simulation (MCS)* method is the general designation for stochastic simulations using random numbers [11, 79]. MCS is a simulation procedure applied to problems involving r.v. The simulation process consists of producing pseudo-random samples from the SSM's underlying probability distributions and propagating these samples through the DSM, forming sets of samples for output quantities of interest, which can then be subjected to descriptive statistical analysis. This method offers significant advantages compared to the analytical methods, since the basic computational part is deterministic and there is no need to simplify the mathematical models to ensure applicability of the method. The main problem in the analysis is how to encode information about the joint probability distribution of the system inputs for the sampling of the system inputs. We deal with this problem extensively in chapter 4.

The main disadvantage of the method is the fact that in order to obtain proper results, the DSM has to be calculated a large number of times, leading to increased computation time. It should however be noted that the obtained information depends only on the sample size and not on the DSM size. This means that the same information is obtained from a 1000-sample MCS for the power flows either in a 10-bus system or in a 1000-bus system. On the other hand, with the present computational power, the DSM calculation time is limited to the order of seconds for large systems, making MCS an appropriate technique for system analysis.

2.5.3 Snapshot Approach and MCS: Memory-less system

MCS offers a natural representation of the snapshot approach: the sampling of the system inputs provides the different snapshots of system operation, that corresponds to different points in time.

A note should be made concerning the applicability of the method in the case that the system contains *energy storage devices*. Implementation of energy storage devices in the system leads to a transition from a *memory-less system* to one with memory: the output power of the energy storage devices is determined by the amount of stored energy (integral over a previous operational period) and the operational control. Therefore, in order to define the state of the system *now*, we need to know the system operation *in the past*. The issues concerning the modeling of such systems are discussed in [50]. The impact of large-scale storage in the system is out of the scope of this thesis. However,

the modeling principles presented here form the basis for the treatment of this problem. For a thorough analysis of the impact of stochastic storage in energy systems, one should refer to [49].

2.5.4 Literature review: conclusions

As a conclusion, we may infer that the applicability of the analytical methods is restricted by the linear approximation of the system model, the assumption of independence between the system inputs and the consideration of linear correlation for cases of normally correlated system inputs. These have been reasonable approximations for the modeling of the uncertainty of the system loads.

However, with SG in the system, the validity of these assumptions is endangered. The problem formulation leads to a multivariate uncertainty analysis problem that involves a large number of different types of non-standard distributions which present complex interdependencies. The indicated method to treat such problems is MCS.

The main differences between load uncertainty and SG uncertainty are illustrated in the following section.

2.6 Load Uncertainty vs SG Uncertainty

2.6.1 Load Uncertainty: time-dependent stochasticity

In Fig. 2.2, the daily load profile of a distribution network based on 15-minute load measurements for the period of one month is presented. As we may see, all measurements fall in a small region around the time-conditional mean (white line in the figure). This is due to the dependence of the human activities on a cyclic-deterministic phenomenon (time of day, day of week, season); the power consumption is not so stochastic, but presents a high time-dependence.

In Fig. 2.3, the daily system load for the year 2003 in The Netherlands is presented (15-minute averages for the period between January 1 and December 31, 2003)^{||}. The cyclic dependence on the time of day is apparent in the graph. In order to isolate the effect of the dependence on the season and the day of the week, in Fig. 2.4 the daily system load is presented for one month in this year. Three different daily load patterns may be distinguished: high loading level corresponding to workingdays (weekdays), medium corresponding to Saturday and low one corresponding to Sunday (Fig. 2.4a). In the second graph (Fig. 2.4b), the workingdays are isolated. As we may see, the uncertainty is limited around the time-conditional mean.

In Fig. 2.5, the same procedure is adopted for the different seasons of the year and the same results concerning the load stochasticity are obtained.

Such interdependencies appear in the stochastic behavior of the system loads, but also with several types of SG, e.g. solar energy. These interdependencies can be removed by performing the calculation separately for groups of hours with similar statistical characteristics (*Time-Frames Analysis* [18]). This procedure corresponds to a conditioning of

^{||}These measurements actually correspond to the aggregate of the power produced by the system CG units, which offer an adequate representation of the power consumed at the system loads.

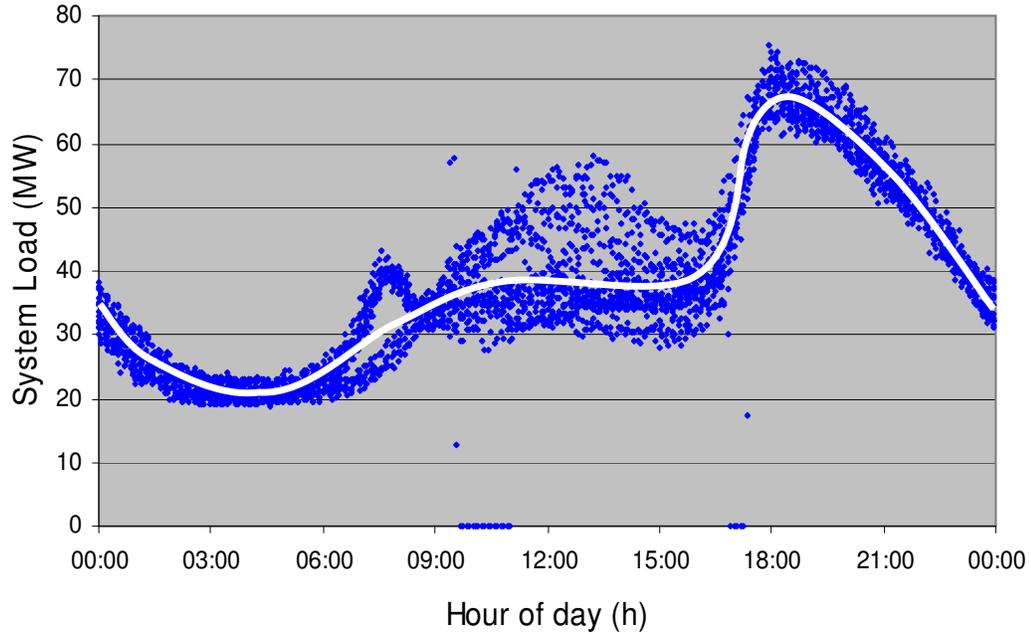


Figure 2.2: Daily load for a distribution system for one month.

the calculation in time. The load in each *time-frame* (TF) can be modeled by superimposing a random noise variable to the conditional mean. This has been the basic load modeling platform in power systems analysis. An aggregation procedure may be further applied in order to obtain the distribution for the whole period of concern**. The TF s chosen for the system analysis depend on the statistical properties of the random inputs. The usual practice is to use a normal distribution for modeling the stochasticity in each TF .

Load stochasticity is the main source of uncertainty in vertical power systems. In this case, the assumptions of independence and normality offer a good modeling platform, corresponding to the conditional load distributions. This is the main reason for the so extensive use of these assumptions in the analytical formulations. In the case of SG however, this approach may prove insufficient, since non-time-dependent uncertainty appears in the system modeling. In the following section we present this type of uncertainty by a basic example: the wind power uncertainty.

2.6.2 SG Uncertainty

In Fig. 2.6, the aggregate wind power output for the Irish power system is presented, for a period of one month (Fig. 2.6a) and for some typical days during that month (Fig. 2.6b) [12].

**In this case, the obtained r.v. are no longer normally distributed. In section 2.9.2, we elaborate on the modeling issues of this aggregation procedure.

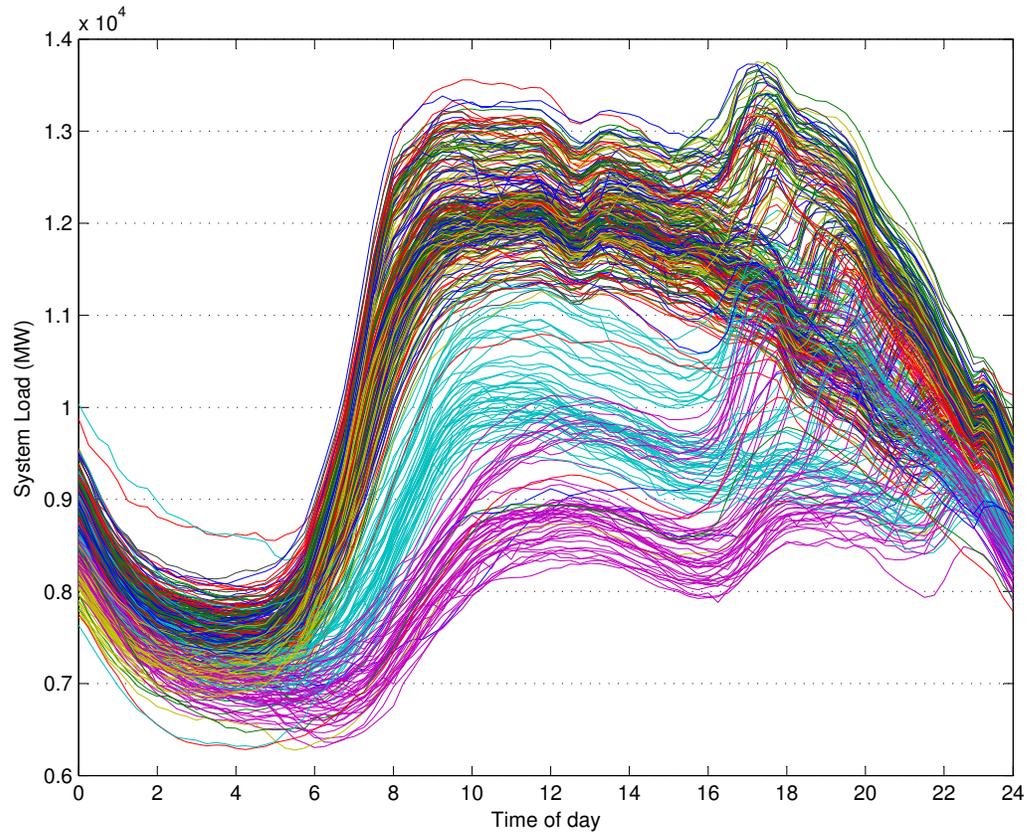


Figure 2.3: Daily system load in 2003 in The Netherlands.

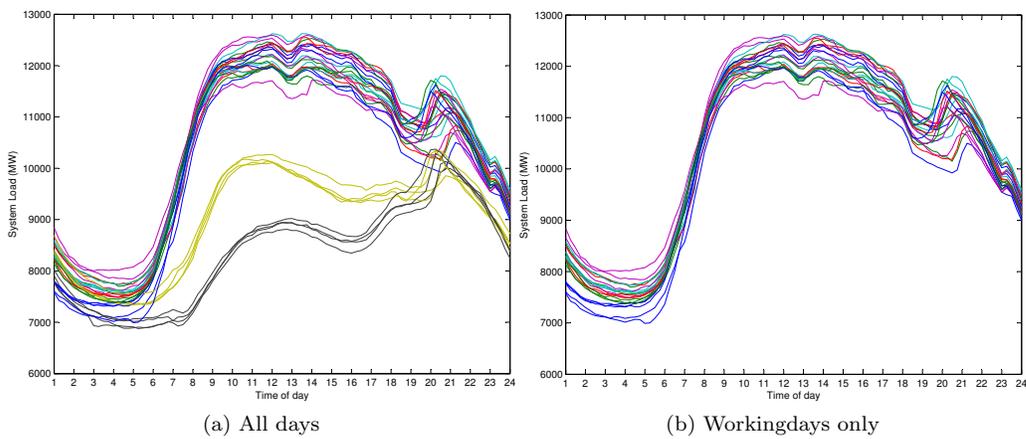


Figure 2.4: Daily system load in September 2003 in The Netherlands.

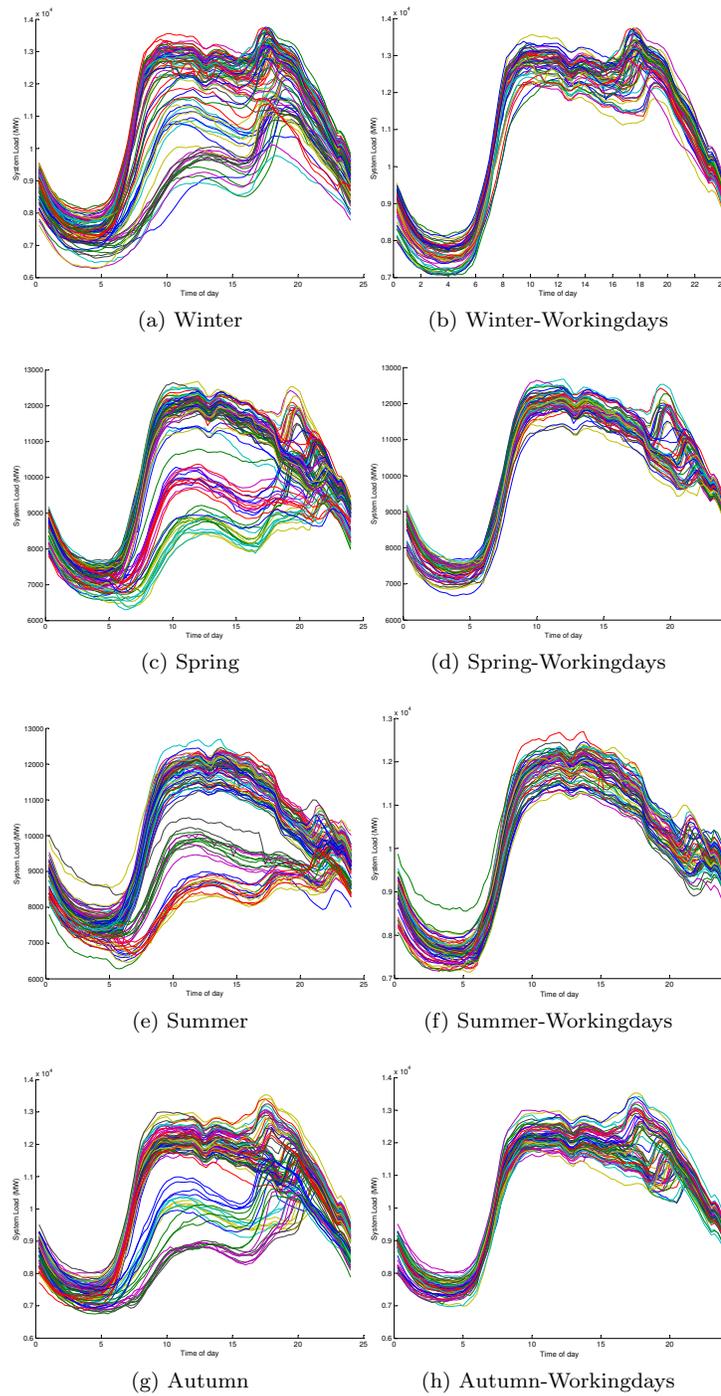


Figure 2.5: Seasonal daily system load in 2003 in The Netherlands.

We can see that the power output can be completely different from day to day; between 0 and 5% for a typical non-windy day (e.g. day 5), between 70% and 90% for a typical windy day (e.g. day 26), being low for the first half of the day and high for the second one (e.g. day 15), or the opposite (e.g. day 14). The variability of the output is so high, that the information obtained from the mean value during the day is insufficient to depict the stochasticity; the power output may vary between zero and maximum for every time of the day. In this case, the time-dependent modeling based on normal distributions fails to represent the system stochasticity.

In the literature, a time-conditional modeling of SG is presented in the case of the modeling of the operational uncertainty of a system with SG units. We will elaborate further on this issue in the following section.

2.7 Planning uncertainty vs operational uncertainty

The basic difference between planning and operational uncertainty lies in the degree of knowledge of the system analyst concerning the uncertainty of the system inputs. In particular, the stochastic prime mover uncertainty for a certain point in time in the future may present a small deviation around the forecasted value; this type of uncertainty may be modeled in a time-conditional manner, as presented in the previous section, by introducing a noise term around the forecasted value. In this case, the uncertainty corresponds to the forecast error and the concepts of normality and independence may provide a good approximation of reality [43, 20].

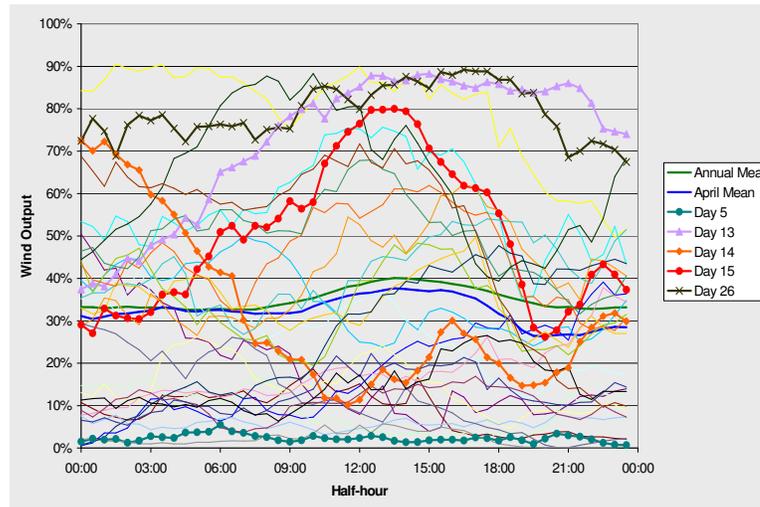
This problem is conceptually different from the planning problem. The planning uncertainty corresponds to the analysis of all possible combinations of the system inputs. In this case, a time-conditional modeling may lead to severe fallacies, especially due to the assumption of independence. We will elaborate further on this issue in the next chapter.

2.8 Problem Formulation

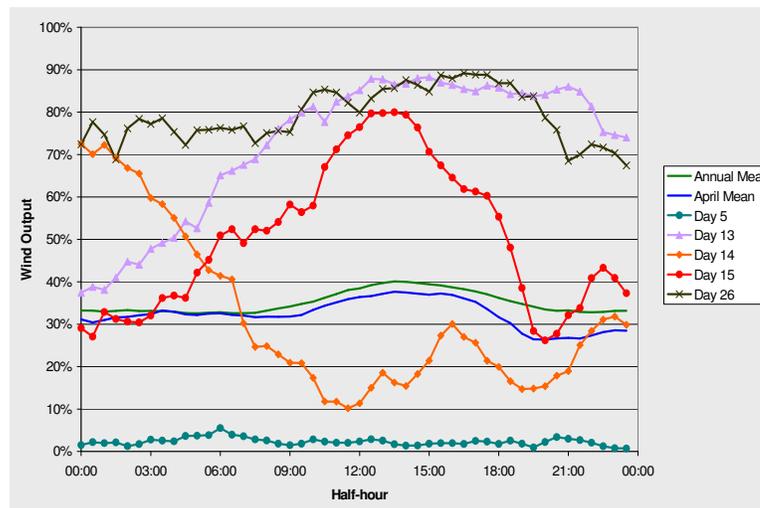
2.8.1 Types of r.v.

The types of input r.v. involved in the steady-state uncertainty analysis are the *load r.v.* $\{\mathbf{L}\}$, the *generation capacity availability r.v.* $\{\mathbf{X}\}$ and the *energy availability r.v.* $\{\mathbf{G}\}$. In general, the active power is considered as a r.v.; the reactive power can be omitted from the uncertainty analysis, since in most cases the load power factor can be considered to be constant, while the generated reactive power is either regulated for voltage control, or supplied under a constant power factor. In these cases, the reactive power is obtained from the active power based on a deterministic functional relationship.

Omitting the reactive power in the uncertainty analysis is a common assumption that is made in many cases, including actual utility practices and studies, and leads to a reduction of the problem dimension. In cases when this is not valid, reactive r.v. should be introduced. The approach presented remains valid.



(a) Windpower output in the Irish power system for one month



(b) Windpower output in the Irish power system for specific days of one month

Figure 2.6: Windpower uncertainty.

Load r.v.

The load r.v.'s $\{\mathbf{L}\}$ model the active power consumed by each load. The time-dependence of the load stochasticity is removed by performing the calculation for groups of hours with similar statistical characteristics (time-conditioning of the stochastic model - *Time-Frames Analysis* [18]). Although the usual practice is to use a normal distribution for the modeling of a time-conditioned load [28, 42], this strongly depends on the respective choice of the time-frame. When a larger time-frame is chosen, the accuracy of such an approximation is limited.

Energy Availability r.v.

The energy availability r.v.'s $\{\mathbf{G}\}$ model the power output of the SG units due to the stochastic prime mover stochasticity. The power distribution is obtained by a transformation of the prime mover distribution. In the case of CG (or dispatchable non-CG), these distributions $\{\mathbf{C}\}$ correspond to the output of the unit and are fixed-point distributions; their value is defined by energy trading mechanisms, economical operation of the system based on optimal power flow algorithms, fixed contracts, etc.

Generation Capacity Availability r.v.

The generation capacity availability r.v.'s $\{\mathbf{X}\}$ model the random unavailability of generating units (CG, SG) due to forced outages caused by unpredictable equipment failures or maintenance. They are binary state r.v. or, when more detail is needed, multi-state r.v. using multiple block representation [13]. These r.v. are multiplied with the energy availability r.v. for each unit.

Thus, in general the following groups of r.v. are involved in the system uncertainty analysis:

Energy Availability r.v.:

- c_T CG units C_1, C_2, \dots, C_{c_T} ,
- g_T SG units G_1, G_2, \dots, G_{g_T} ,
- l_T loads L_1, L_2, \dots, L_{l_T} ,

Generation Capacity Availability r.v.:

- CG units: $X_{C_1}, X_{C_2}, \dots, X_{C_{c_T}}$,
- SG units $X_{G_1}, X_{G_2}, \dots, X_{G_{g_T}}$.

2.8.2 Problem Decomposition

For the stochastic modeling of the system, the joint distribution F_{GLX} over all the input r.v. should be defined^{††}. The r.v. X_{Cg}, X_{Gg} , representing the generation capacity

^{††}By F_S , the *cumulative distribution function (cdf)* of a r.v. S is denoted.

availability, are considered to be mutually independent, as well as independent from the stochastic inputs G_g and L_l [13]. Thus, the problem can be decomposed as follows:

$$F_{GLX} = \underbrace{F_{GL}}_{\text{Inputs Uncertainty}} \cdot \overbrace{\prod_{c=1}^{c_T} F_{X_{C_c}} \cdot \prod_{g=1}^{g_T} F_{X_{G_g}}}^{\text{Generation Reliability}} \quad (2.2)$$

into two distinct parts, the definition of the joint distribution F_{GL} over the non-dispatchable system entities i.e. loads and SG (*system inputs uncertainty*), and the *system generation reliability*.

The system generation reliability can be performed based on the well-known concepts developed for the study of the reliability of CG units in vertical power systems [13]. In this case, the failure and maintenance rates of the SG units should be incorporated in the analysis (second term in equation 2.2). Since this modeling part has been extensively investigated in the literature, we will further concentrate on the definition of the other term in equation 2.2, the system inputs uncertainty.

This is the core of the power system uncertainty analysis and is the most cumbersome part of the stochastic modeling of the system. Generally, in such a problem the one-dimensional marginal distributions F_{G_g} and F_{L_l} are the most easy to assess through data analysis or expert judgement. It is a common fallacy to consider this information to be enough for the determination of F_{GL} . This information is sufficient if one assumes independence between G_g and L_l . Obtaining the joint distribution F_{GL} given the marginals F_{G_g} and F_{L_l} is however a non-trivial problem, since there exist infinite number of joint distributions F_{GL} with the same marginals F_{G_g} and F_{L_l} , corresponding to an infinite number of stochastic dependence structures between the r.v. G_g, L_l .

The arbitrary assumption of independence produces a systematic tendency to disregard the effect of the dependence structure. This generally is an underestimation of the system variability and can lead to serious pitfalls. The modeling of stochastic dependence has been a cornerstone in the research on multivariate uncertainty analysis in recent years. In [48, 63, 53, 62, 51] one may find some basic references on the treatment of this problem. The related research lead to a main approach, namely the splitting of the modeling effort in two separate tasks:

- *Marginals*: model the one-dimensional marginal distributions.
- *Stochastic dependence*: model the stochastic dependence structure between the inputs.

This approach is adopted in this thesis for the determination of the inputs uncertainty F_{GL} .

2.9 Marginal distributions

2.9.1 Sampling of a r.v. in MCS

By definition, for a r.v. X with an *invertible* cdf $F_X(x) = P(X \leq x)$, the r.v. $F_X(X)$ follows a uniform distribution on the interval $[0, 1]$. This relationship forms the basis for

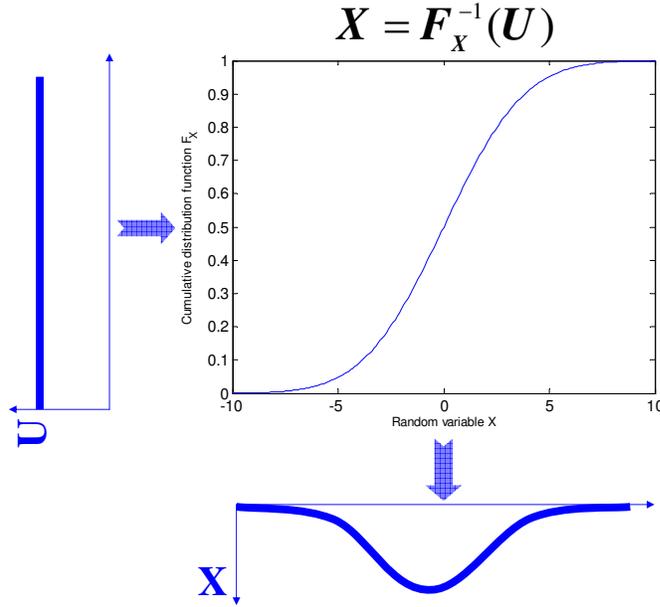


Figure 2.7: Sampling of a r.v. in MCS.

the sampling of any r.v. in MCS studies. The proof of this statement is as follows:

$$\text{For } r \in [0, 1] : P(F_X(X) \leq r) = P(X \leq F_X^{-1}(r)) = F_X[F_X^{-1}(r)] = r \quad (2.3)$$

Thus, if U is the uniform distribution, $F_X(X) = U \Leftrightarrow X = F_X^{-1}(U)$. Therefore, $F_X^{-1}(U)$ follows the distribution of X .

For the sampling of a single r.v. X with invertible cdf F_X , first a random realization u from a uniform r.v. U in $[0, 1]$ is generated and then the transformation $x = F_X^{-1}(u)$ is applied. In this case, the samples x follow the distribution F_X [51]. The above-mentioned procedure is presented in Fig. 2.7.

This method may also be applied for the sampling from measured data, using the empirical cdf F_X^e . In this case, the sampling between adjacent values may be performed by interpolation (the case of linear interpolation is presented in Fig. 2.8) or by choosing one of the border values. This feature offers significant advantages for the system modeling based on real data; instead of trying to fit some parametrical distribution, one may use the exact data distribution for the simulation.

2.9.2 Load

As mentioned, a normal distribution may be used for the modeling of time-conditioned r.v. in each TF. When the time period that is studied coincides with multiple TFs, an aggregation procedure can be applied to find the resulting distribution as a mixture of these normals. This mixture can be sampled by the use of an independent uniform r.v. U_{TF} as TF-indicator. In particular, based on the relative duration of each TF, each

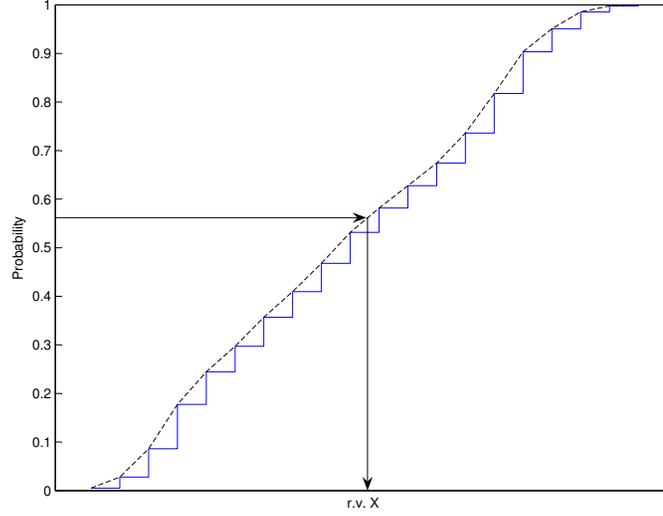


Figure 2.8: Sampling of a r.v. in MCS based on measured data.

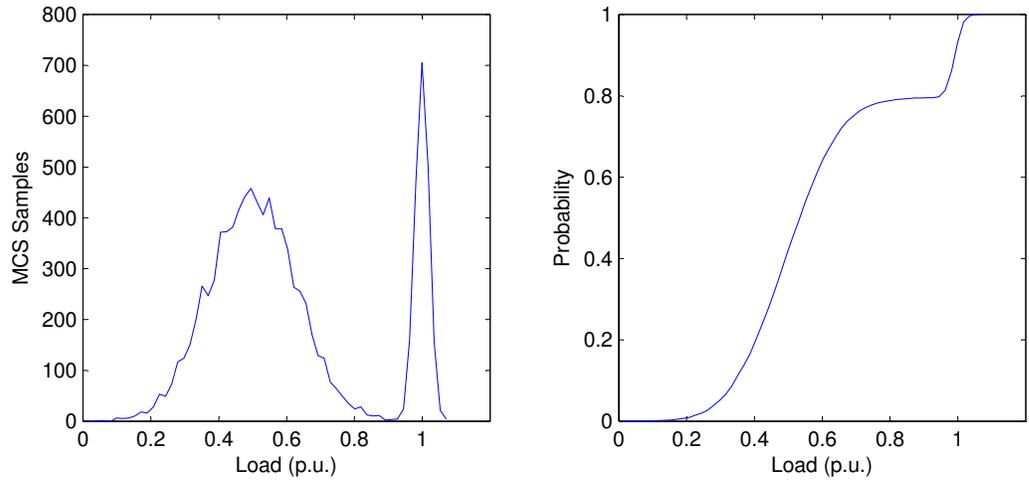
Table 2.1: Time-frames settings

	Time Ratio		Mean Load (% max Load)		St. Deviation (% Mean Load)	
	<i>2TF</i>	<i>3TF</i>	<i>2TF</i>	<i>3TF</i>	<i>2TF</i>	<i>3TF</i>
<i>TF1</i>	0.8	0.45	0.5	0.5	0.25	0.15
<i>TF2</i>	0.2	0.35	1	0.75	0.02	0.1
<i>TF3</i>	–	0.2	–	1	–	0.02

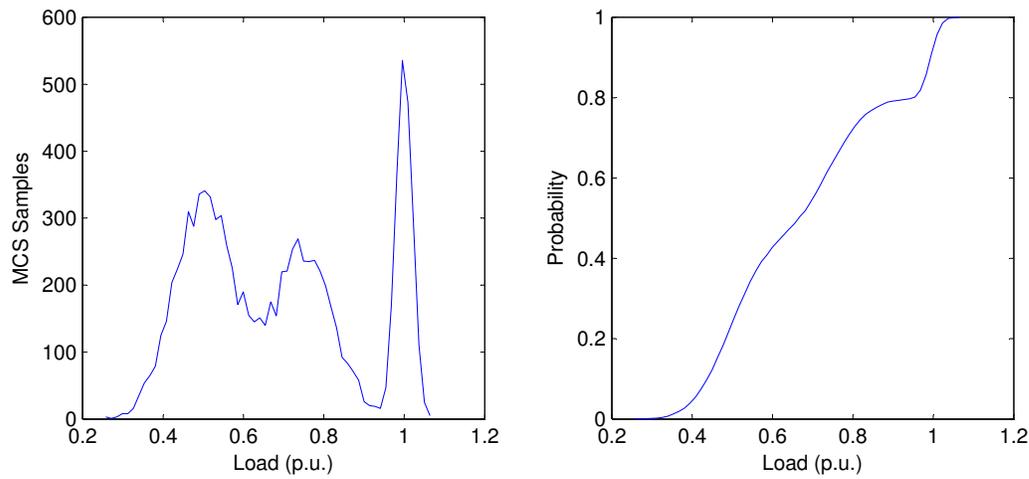
sample drawn from U_{TF} is matched to a specific TF and a sample is drawn from the normal distribution belonging to the respective TF. In table 2.1 example TF settings for a 2-TF and a 3-TF segmentation are presented. According to the time ratio of this table, in the case of 2-TF all samples lower than 0.8 correspond to TF1 and the rest to TF2.

In Fig. 2.9, the density and cumulative distributions for a 10000-sample MCS of a 2-TF and a 3-TF approximation are presented. We can see that by increasing the number of TFs, the cdf provides a better approximation of the load duration curve.

When the time-dependence is of no interest for the study and data are available, the method presented in the previous paragraph may be used for sampling based on the empirical distribution. We demonstrate the applicability of the methodology by sampling based on the load data presented in Fig. 2.3, representing the system load for The Netherlands for the year 2003 (35040 measurements). In Fig. 2.10 the pdf based on the measurements is presented, together with the respective simulated one. We can see that by using this method a high accordance between the distributions is achieved. More examples on the application of this method may be found in section 4.2.

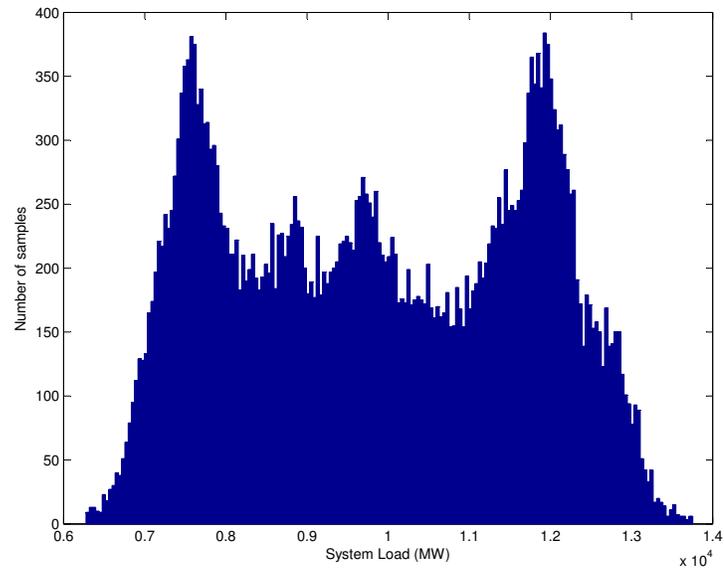


(a) 2-TF segmentation

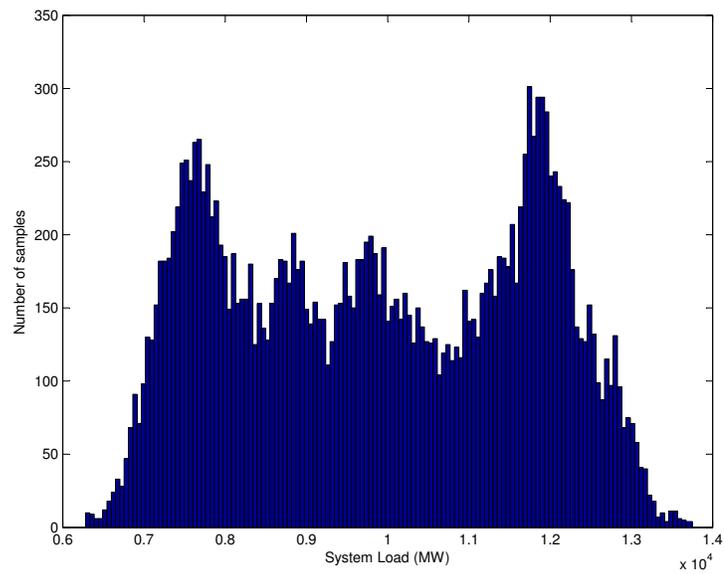


(b) 3-TF segmentation

Figure 2.9: Normalized load distribution as mixture of TF-distributions (10000-sample MCS).



(a) Measurements (35040 points)



(b) 20000-sample MCS

Figure 2.10: pdf for the system load in 2003 in The Netherlands.

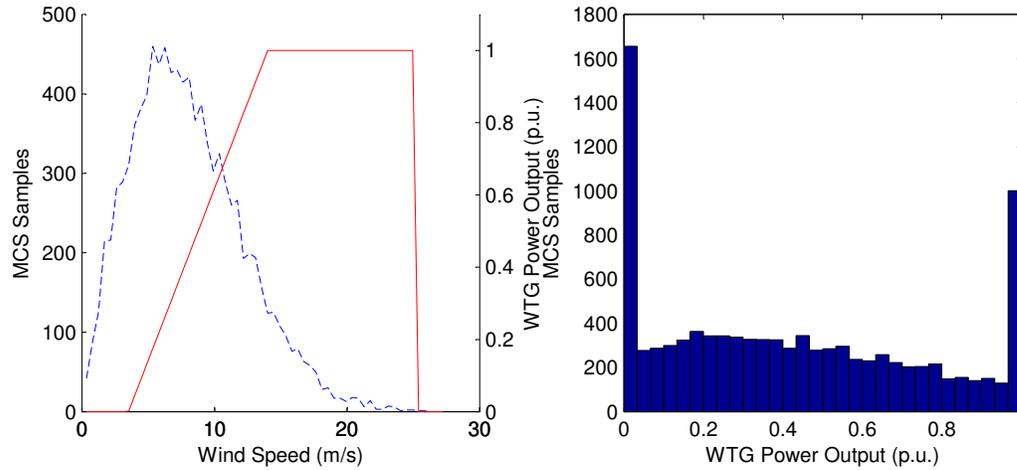


Figure 2.11: WTG wind speed/power characteristic and distributions.

2.9.3 Stochastic Generation

A stochastic generator is an energy conversion system that converts an uncontrollable primary energy source (wind-, solar-, hydro-, wave- energy, waste heat, etc.) into electrical power and feeds it into the power system. The power output of a stochastic generator is defined by two factors:

1. *Stochastic Prime Mover*: the type of primary energy source used for electrical power generation.
2. *Energy Conversion System*: according to the converter technology, the power output of the generator for each input value of the prime mover can be defined by a deterministic (functional) relationship.

Thus, the power output distribution is obtained as a transformation of the prime mover distribution. Due to this transformation, non-standard SG power distributions may be obtained. We will illustrate this issue by presenting the case of a *wind turbine generator (WTG)*. In this case, the prime mover, i.e. the wind activity in a specific location, follows a Weibull distribution [92]. The energy conversion system is modeled here by the wind speed/power output characteristic; a pitch-controlled WTG is considered here, with cut-in, nominal and cut-out wind speeds of 3.5, 14 and 25 m/s respectively. In Fig. 2.11, the results from a 10000-sample MCS for the WTG wind speed and the corresponding power output distribution are presented. The obtained power distribution presents a concentration of probability masses at the zero and nominal output power. This is due to the effect of the non-monotonic WTG characteristic [69]: for wind speeds lower than the cut-in and higher than the cut-out value, the power output is zero, while for wind speeds in between nominal and cut-out, nominal power is generated.

2.10 Stochastic Dependence Structure

The modeling of the stochastic dependence structure between the system inputs is the second task for the stochastic modeling of the system. Whereas the marginal distributions refer to the power output spectrum of single units, the dependence structure refers to the joint behavior of these inputs and has important implications for the properties of their sum. The following example will help to clarify this statement.

Consider the case of two loads L_1 and L_2 . Their consumption is a fixed integer between 1 and 5 MW. Three cases of operation are considered; their consumptions are co-fluctuating, thus they take the same values at the same time (case 1), their consumptions are fluctuating independently (case 2), and their consumptions fluctuate in the opposite way (case 3). In all cases, the total load in the system $L_1 + L_2$ is calculated:

<i>Case1 :</i>	$L_1(MW)$	1	2	3	4	5
	$L_2(MW)$	1	2	3	4	5
	$L_1 + L_2(MW)$	2	4	6	8	10
<i>Case2 :</i>	$L_1(MW)$	1	2	3	4	5
	$L_2(MW)$	2	5	4	3	1
	$L_1 + L_2(MW)$	3	7	7	7	6
<i>Case3 :</i>	$L_1(MW)$	1	2	3	4	5
	$L_2(MW)$	5	4	3	2	1
	$L_1 + L_2(MW)$	6	6	6	6	6

Although nothing changes in the single behavior of each load, their aggregated consumption is totally different. What we actually do is adding the outputs with a different order, which has a direct impact to the variability of the sum; in the first case a large fluctuation from 2 to 10MW is observed, in the second case a smaller one, while in the third case the aggregated consumption is fixed at 6MW. The variability of the sum determines the stochastic stress in the system and is different between the cases; case 1 is the worst operational scenario (highest variability), and case 3 is the best (minimum variability), with the maximum system loading reaching only 60% of the one in case 1. Although the distribution of the output is different, the mean value for all cases is the same (6MW), since the single behavior of the inputs is the same.

In the case of SG, the stochastic dependence refers to the coupling between the respective prime movers. In particular, the power output of stochastic generators situated in a small geographic area show similar fluctuations due to their mutual dependence on the same prime mover, which is not the case for stochastic generators situated in remote areas.

Modeling the dependence structure is the more cumbersome problem in the power system uncertainty analysis. The methods developed until now are limited to the treatment of correlated normal loads. However, new methods have to be employed for the modeling of the complex interdependencies introduced by the SG in the power system.

2.11 Conclusions

The incorporation of stochastic generation in the power system necessitates the use of a unified approach for the modeling of the inputs uncertainty. Traditionally, stochastic models based on independence and normality are used for modeling the time-dependent stochasticity of the loads in the system. Stochastic generation however introduces non-time-dependent stochasticity, which cannot be modeled based on this approach. The modeling of the stochastic dependence between the correlated resources is the cornerstone in the system analysis, since it has a direct impact to the determination of the sum of the stochastic inputs. The approach presented here proposes the splitting of the modeling effort in the modeling of the one-dimensional marginal distributions and the multidimensional stochastic dependence structure. We devote the next chapter in the presentation of the theory related to the modeling of stochastic dependence and investigate the fallacies deriving from the use of the assumptions of normality and independence for the modeling of stochastic generation.

Chapter 3

Models of Stochastic Dependence

As discussed in the previous chapter, modeling the stochastic dependence between the system inputs is a task of utmost importance, since it defines the aggregate power output from multiple stochastic inputs. The modeling of stochastic dependence has constantly been underestimated in the related literature. In this chapter we show the modeling principles for the stochastic dependence, based on the confrontation of two examples: the case of two normal loads and the case of two SG units. In particular, we investigate the role of dependence to the definition of the properties of the sum of r.v. (loads or SG units). We show that, the assumptions used so far i.e. normality and independence, when applied to the modeling of SG units, can lead to severe fallacies.

3.1 Mean value of the sum of r.v.

The sum of n r.v. X_1, X_2, \dots, X_n is a r.v. with a mean that equals the sum of the means of the respective r.v. [71]:

$$\mu \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n [\mu(X_i)] \quad (3.1)$$

Thus, the mean value of the distribution of the sum of r.v. always equals the sum of the mean values of the respective r.v. This however does not hold for other moments of the distributions. In particular, different dependence structures yield different distributions around the same central point. In the course of this chapter, we present the theory concerning the clarification of these issues.

3.2 Independence

The concept of independence is fundamental in probability theory and has been widely used in the stochastic modeling of power systems. We can say that two events A and B

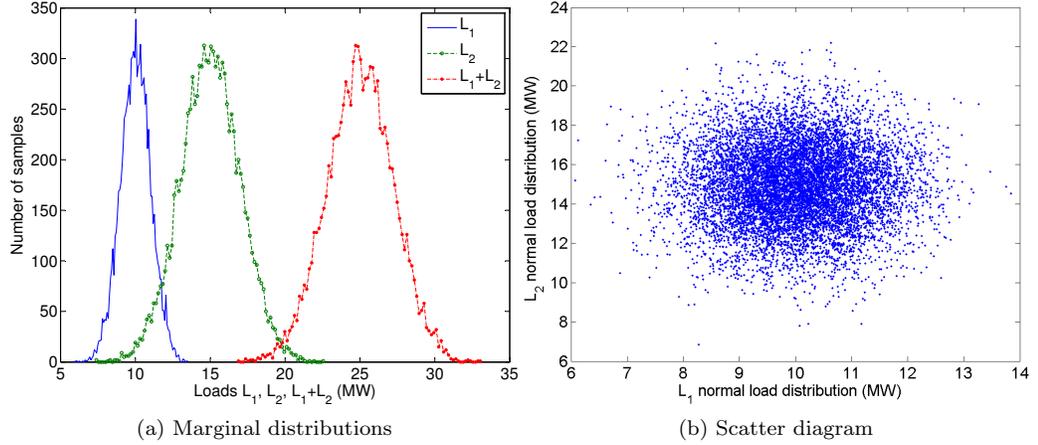


Figure 3.1: Independent normal load distributions.

are *independent* when the occurrence of one of them has no influence on the probability of the other. More precisely,

Definition 3.2.1 (Independence) *Random variables X_1, \dots, X_n are **independent** if for any intervals I_1, \dots, I_n ,*

$$P\{X_1 \in I_1, \text{ and } \dots X_n \in I_n\} = \prod_{i=1}^n P\{X_i \in I_i\} \quad (3.2)$$

The sum of n independent r.v. X_1, X_2, \dots, X_n is a r.v. of which the mean is given by the equation 3.1 and the variance/standard deviation is [71]:

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n [\text{Var}(X_i)] \Rightarrow \sigma \left(\sum_{i=1}^n X_i \right) = \sqrt{\left(\sum_{i=1}^n [\sigma^2(X_i)] \right)} \quad (3.3)$$

3.2.1 Independent normal loads

The sum of n independent normal loads L_1, L_2, \dots, L_n can be calculated analytically; it is a normal distribution with a mean and standard deviation given by the equations 3.1 and 3.3. In Fig. 3.1, the marginal distributions and the scatter diagram for two independent loads with mean values $\mu_1 = 10MW$ and $\mu_2 = 15MW$ and standard deviations $\sigma_1 = 1MW$ and $\sigma_2 = 2MW$ respectively, are presented, together with the distribution of the sum.

3.2.2 Independent stochastic generators: Central Limit Theorem

In the case of stochastic generators, independence refers to the decoupling of the prime movers. For WTGs, this refers to a mutual independence between the wind speed distributions at the different sites. This type of relationship can refer to remote sites in

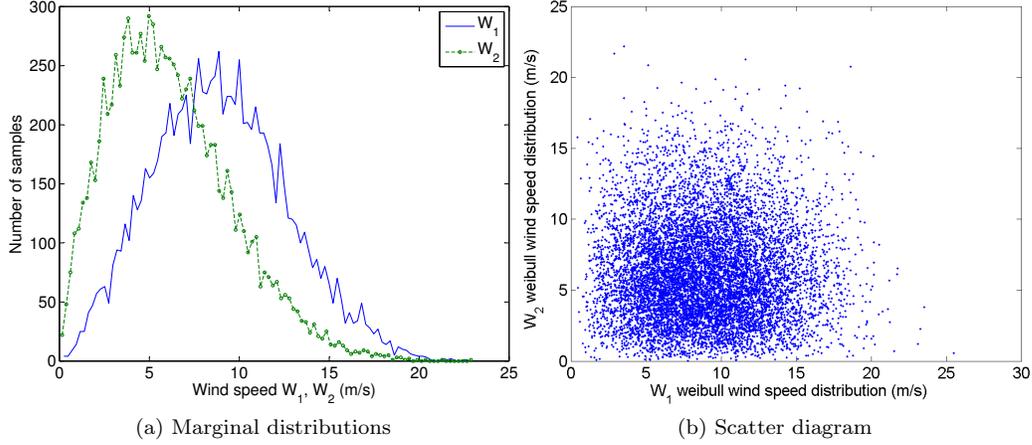


Figure 3.2: Independent Weibull wind speed distributions.

the system, but not to WTGs situated in a small geographic area. In general, the wind resources present significant correlation*. In the following, we will consider two pitch-controlled WTGs, according to the modeling principles presented in section 2.9.3. The respective Weibull parameters η and β for the wind speed distributions at the two sites are $\beta_1 = 2.6$, $\beta_2 = 1.9$ and $\eta_1 = 10$, $\eta_2 = 7^\dagger$.

In Fig. 3.2, the marginal distributions and scatter diagrams of 10000 MCS samples for independent wind in two sites are presented. Obviously, in order to comply with the Weibull marginals, a different scatter diagram than in the case of normal loads (Fig. 3.1b) results.

Due to the effect of the wind speed-power output characteristic of the WTG, the power output follows a non-standard distribution as the one presented in in section 2.9.3, with a concentration of probability masses at the zero and maximum values. Although the output of each WTG follows this distribution, their sum is different. In Fig. 3.3, the power output in the case of a single WTG is presented, next to the aggregate power output of 2, 5 and 10 independent WTGs. As can be seen, this summation leads to a 'smoothing' of the aggregate power output distribution; the spikes diminish and the probability distribution of the sum approaches a normal distribution (Fig. 3.3d) when the number of r.v. in the sum increases. This is due to the applicability of the *Central Limit Theorem (CLT)*[‡].

Thus, in the case of independence, irrespective from their type, a number of r.v. will always sum up to a normal distribution. The stochastic behavior of the sum is therefore determined by the assumption of independence, rather than the type of the marginal distributions. This is a general result: irrespectively from type, the aggregate of a large

*In section 4.2, the statistical analysis of the wind regime in The Netherlands is presented; the lowest measured correlations between wind speed distributions throughout the country is 0.47.

[†]A Weibull pdf with scale parameter η and shape parameter β is: $f(v|\eta, \beta) = \frac{\beta}{\eta} \left(\frac{v}{\eta}\right)^{\beta-1} \exp\left(-\left(\frac{v}{\eta}\right)^\beta\right)$.

[‡]According to the CLT, under general conditions the distribution of the sum of n independent r.v. approaches a normal distribution as n increases [71].

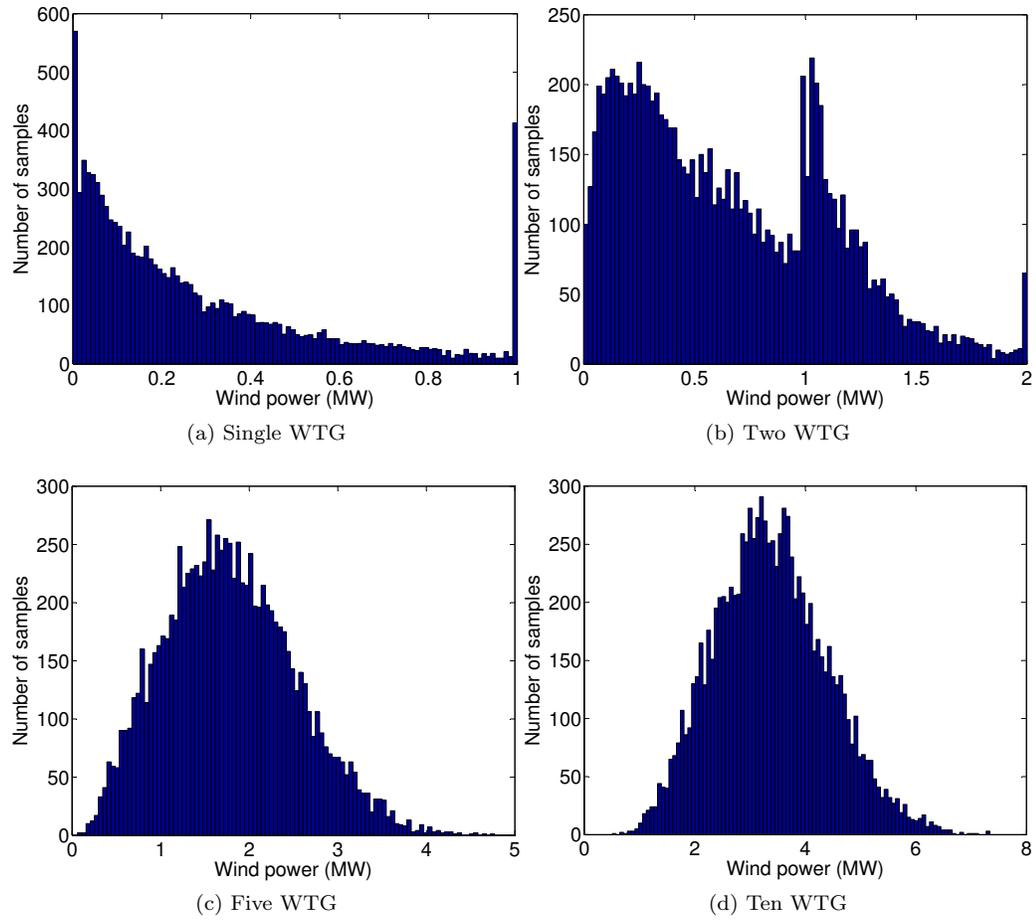


Figure 3.3: Independent wind turbine power output distributions.

number of independent inputs approximates a normal distribution.

Probabilistic load flow (PLF) analysis

The sum of n independent r.v. can be calculated analytically by using the convolution integral. As discussed in section 2.5.1, convolution forms the computational core of the *probabilistic load flow (PLF)* analysis. The cornerstone of the method is the linearization of the load flow equations around an operating point[§]. As a result of this linearization, the system states (nodal voltages) and outputs (line power flows) can be expressed as a weighted sum of the independent system inputs (nodal power injections). The output distributions can be obtained as the convolution of the system inputs, using appropriate numerical techniques, basically *Fast Fourier Transform (FFT)* algorithms. The applicability of such numerical methods is restricted to system inputs that are mutually independent.

Independence however is just one case. When the system inputs are not independent, then they are dependent. In this case, we need to evaluate the degree of dependence and employ new techniques for the system analysis.

3.3 Measure of Dependence: Product Moment Correlation

The first question to be answered is how do we measure the dependence between r.v. The most widely used measure is the *product moment correlation*, also called *linear* or *Pearson correlation*. For the bivariate case it is defined as:

Definition 3.3.1 (Product Moment Correlation) *The product moment correlation of r.v. X, Y with finite expectations $E(X), E(Y)$ and finite variances $\sigma^2(X), \sigma^2(Y)$, is*

$$\rho(X, Y) = \frac{E(XY) - E(X)E(Y)}{\sigma(X)\sigma(Y)} = \frac{E[(X - \mu(X))(Y - \mu(Y))]}{\sigma(X)\sigma(Y)} = \frac{Cov(X, Y)}{\sigma(X)\sigma(Y)} \quad (3.4)$$

If we are given N pairs of samples (x_i, y_i) from the random vector (X, Y) , we calculate the sample, or population product moment correlation $\hat{\rho}_{xy}$ as follows:

$$\hat{\rho}_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{Y})^2}} \quad (3.5)$$

where $\bar{X} = \frac{1}{N} \sum_{i=1}^N x_i$ and $\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i$.

The product moment correlation is specified only for r.v. with finite expectations and variances. For every r.v. X and Y that fulfill these preconditions the basic properties of product moment correlation are the following:

1. *Range:* $-1 \leq \rho(X, Y) \leq 1$

[§]The solution of the deterministic load flow problem for the mean values of the system inputs is generally chosen as the point of linearization.

2. *Independence*: if X, Y are independent, then $\rho(X, Y) = 0$
3. *Invariance under linear transformations*:
for $a, c \in \mathbb{R} \setminus \{0\}, b, d \in \mathbb{R}, \rho(a \cdot X + b, c \cdot Y + d) = \text{sgn}(a \cdot c) \rho(X, Y)$
4. *Linear dependence*: if $\rho(X, Y) = 1$ then for $a > 0, b \in \mathbb{R}, X = aY + b$

Product moment correlation is a measure of stochastic dependence among many, albeit the most popular one. It measures the linear dependence between the r.v., in other words it is the geometric mean of the best linear predictor of X (Y) given Y (X). However, it has often been a source of confusion by being used as representation of dependence. Some of the confusion may arise from the literary use of the word to cover any notion of dependence. To a mathematician (product moment) correlation is only one particular measure of stochastic dependence. It is the canonical measure in the world of multivariate normal distributions, and more generally for spherical and elliptical distributions [38]. Thus, for normal distributions, it provides the complete representation of stochastic dependence. However, when we leave the world of normals, it becomes just a measure of linear dependence, that may be misleading as representation of dependence, especially in cases of complex, non-linear dependencies. In the following we will discuss these issues in more detail.

The sum of n dependent r.v. X_1, X_2, \dots, X_n is a r.v. with a mean that is given by the same equation as for the case of independence (3.1). The variance and the shape of the distribution, however, depends on the dependence structure between them. This impact of dependence is presented in a comprehensive example in the following section.

3.4 Correlated normal loads: Joint Normal Distribution

3.4.1 Bivariate case

In Fig. 3.4, the scatter diagrams for the two normal loads of the previous section are presented, corresponding to different product moment correlations ranging from -1 to 1 . As can be seen, while the correlation increases from zero to one, the r.v. become more linearly dependent. Negative correlation means that the best linear predictor of L_1 (L_2) given L_2 (L_1) has a negative gradient while in the case of positive correlation the best linear predictor has a positive gradient.

Although the marginal distributions remain the same, the distribution of the sum of the r.v. is radically affected by the dependence between them. In Fig. 3.5, the distributions for the sum of the two r.v. for the correlations in Fig. 3.4 are presented. Different correlations result to normal distributions with different variances around the same central point. Minimum variance is obtained for maximum negative correlation ($\rho = -1$) and maximum variance in case of maximum positive correlation ($\rho = 1$). Thus, depending on the correlation, the sum may vary between 23 and 25MW ($\rho = -1$) or between 15 and 35MW ($\rho = 1$). This result may be understood by the scatter diagrams of Fig. 3.4, especially the extreme ones for $\rho = -1$ (Fig. 3.4a) and $\rho = 1$ (Fig. 3.4f). Negative correlation means that high values of the one r.v. are matched with low values of the other, while positive correlation means that high values of the one r.v. are always

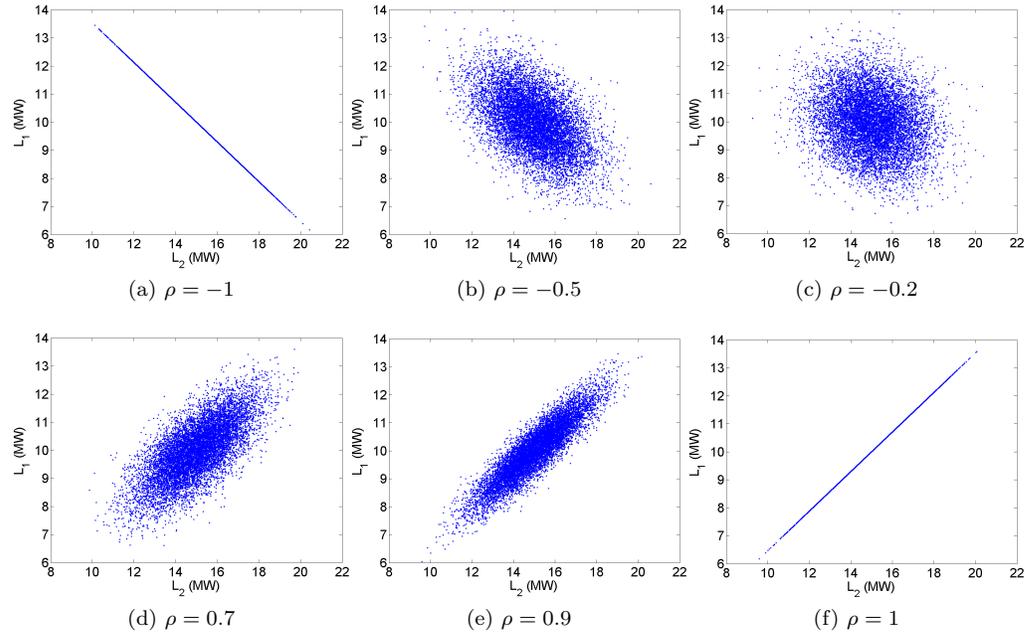


Figure 3.4: Scatter diagrams for correlated normal loads.

matched with high values of the other. This matching has a direct impact on the behavior of the sum: negative correlation prevents extreme values from happening at the same time, while positive correlation urges coincidence of extreme events.

As can be seen from Fig. 3.4a and 3.4f, $\rho = \pm 1$ results in a perfect linear relationship between the loads. This feature has been used for the extension of the PLF analysis in the case of correlated inputs. In [9], Allan et al. proposed a new formulation of the method in order to include correlated demands, that was extensively used [2, 27]. According to this approach, the dependence between demands in the system is assumed to be close to perfect linear. Therefore, each normal load can be split in two components, one that corresponds to mutual linear correlation between loads and one that represents a superimposed independent normal noise. All linearly correlated components are summed to one normal r.v. and convolution techniques can be further employed for the derivation of the output distributions.

The applicability of the technique is however restricted by the assumption of perfect linear dependence. When this assumption is not valid, simulation techniques that make use of the theory of the *multivariate (joint) normal distribution* should be adopted. The respective theory is presented in the following section.

3.4.2 Multivariate case

This case corresponds to the modeling of n correlated normal loads (more than two). The *multivariate normal distribution* can be used as the joint distribution of these r.v. In this

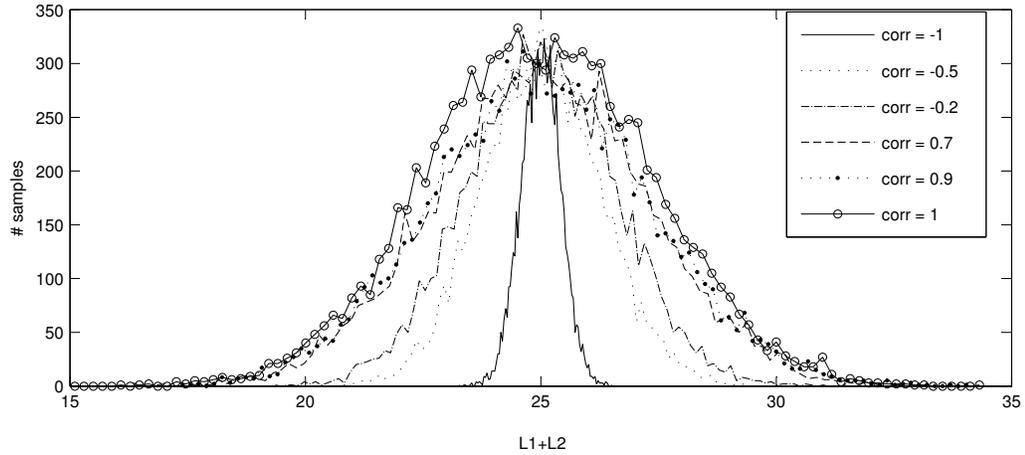


Figure 3.5: Aggregate distribution of correlated normal loads.

case, all mutual correlations are grouped in the form of a matrix, the *product moment correlation matrix* \mathbf{R} . In the case of normal loads $\mathbf{L} = [L_1, \dots, L_n]$ with mean values $\mu_{\mathbf{L}} = [\mu_1, \dots, \mu_n]$ and standard deviations $\sigma_{\mathbf{L}} = [\sigma_1, \dots, \sigma_n]$, the simulation algorithm 3.1 may be used. The application of this algorithm is straightforward: starting from an n -vector of independent standard normals \mathbf{Z} we can construct an n -vector of normals \mathbf{L} , correlated according to the product moment correlation matrix \mathbf{R} .

Algorithm 3.1 Joint Normal simulation algorithm for modeling of correlated normals.

1. Simulate an n -vector of independent standard normals $\mathbf{Z} = [Z_1, \dots, Z_n]$
 2. Calculate the matrix product $y = \mathbf{T} \times z$, where \mathbf{T} is a lower triangular matrix such that $\mathbf{R} = \mathbf{T} \times \mathbf{T}^T$ (*Cholesky decomposition*). The n -vector y forms a sample drawn from a random vector \mathbf{Y} that follows a standard multivariate normal distribution with correlation matrix \mathbf{R} : $\mathbf{Y} = \mathbf{T} \times \mathbf{Z}$
 3. Transform the correlated standard normals \mathbf{Y} to normals \mathbf{L} with mean $\mu_{\mathbf{L}}$ and standard deviation $\sigma_{\mathbf{L}}$: $\mathbf{L} = \sigma_{\mathbf{L}}^{\mathbf{T}} \cdot \mathbf{Y} + \mu_{\mathbf{L}}$
-

The results presented for the bivariate case, concerning the impact of correlation to the aggregate of the r.v. may be extended to the multivariate case. Thus, for different dependence structures, the aggregate of a number of correlated normals is a normal distribution with different spreading around the same central point.

3.4.3 Positive-semidefiniteness

A necessary condition in order to carry out the factorization $\mathbf{R} = \mathbf{T} \times \mathbf{T}^T$ used in algorithm 3.1 is that \mathbf{R} has to be *positive-semidefinite*; that is, for all $x \in \mathbb{R} \setminus \{0\}$, $x^T \mathbf{R} x > 0$ [¶]. The construction of the positive-semidefinite product moment correlation matrix is the most difficult part in the modeling procedure. In section 4.3, we elaborate further on this important issue.

Up to this point, we discussed about how we can model n normal loads, either independent or dependent, and m independent stochastic generators. What remains is to deal with the case of m dependent stochastic generators, i.e. the modeling of non-dependent, non-normal r.v. This is the most difficult problem to be faced, yet extremely relevant in power system modeling. Here, the measures of dependence applied in the world of normal distributions fail and new methods have to be employed.

3.5 Correlated stochastic generators: where product moment correlation fails

3.5.1 Bivariate case

In order to illustrate the problems arising from the modeling of correlated non-normal inputs, we will start the discussion with the bivariate case. In particular, we investigate the power output from two WTGs when their prime movers (wind speed modeled as Weibull distributions) are correlated.

According to the previous section, the product moment correlation can be used for the representation of dependence between normal loads. In the case of non-normals however, it fails to offer a good representation of the dependence. In order to illustrate the pitfalls concerning the use of product moment correlation, we will use a simple example, i.e. the case of perfect dependence. Perfect dependence between two r.v. corresponds to the case when the one follows perfectly the fluctuations of the other. In Fig. 3.6 the scatter diagrams and the generated sample sequences are presented, for the cases of (a) perfectly dependent normal loads, (b) Weibull wind speeds and (c) the case of a normal load perfectly dependent to a Weibull wind speed. As can be seen in Fig. 3.6a, 3.6b and 3.6c, perfect dependence corresponds to a co-fluctuation of the generated samples.

As shown in Fig. 3.6d, normals correlated with $\rho = 1$ are linearly dependent; thus they may be written as $L_1 = aL_2 + b$, where $b \in \mathbb{R}$ and $a > 0$. This linear dependence corresponds to the case of perfect dependence. The same type of perfect dependence between two Weibull wind speed distributions W_1 and W_2 however does not correspond to linear dependence, since a linear relationship of the form $W_1 = aW_2 + b$ between W_1 and W_2 is impossible due to the formula of the Weibull distribution. As may be seen in Fig. 3.6e and 3.6f, this co-fluctuation is the result of a non-linear dependence. This non-linear dependence leads to product moment correlation values less than 1; in our example $\rho_{L-L} = 1$, but $\rho_{W-W} = 0.9950$ and $\rho_{W-L} = 0.9954$. Thus, two r.v. may be perfectly dependent, but only for some types of r.v. (i.e. normals) this perfect dependence corresponds to a linear relationship between them. In this case the product

[¶]Equivalently, a matrix is positive-semidefinite if all its eigenvalues are non-negative.

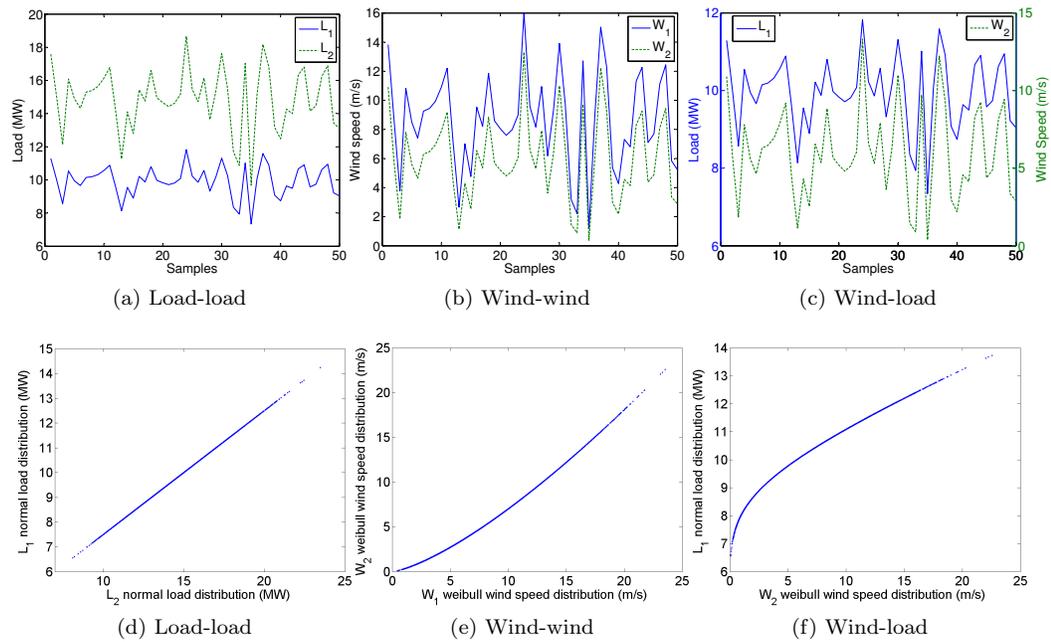
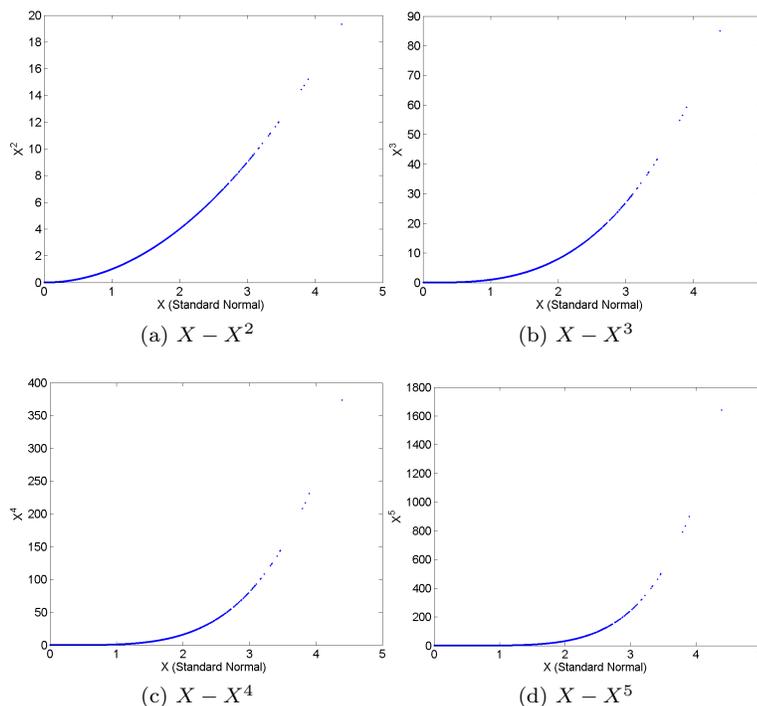


Figure 3.6: Time-series and scatter diagrams for perfect correlation between loads and wind speeds.

Figure 3.7: Scatter diagrams between X and powers of X .

moment correlation reaches the extreme values $\rho = 1$ and $\rho = -1$; in all other cases it lies in a closed interval $[\rho_{min}, \rho_{max}]$ and $\rho_{min} < 0 < \rho_{max}$ (Hoeffding theorem, [44]).

Although in the previous example ρ gives an estimate close to 1, in many cases this difference may be significant. For example, let X be a standard normal variable with support $\Omega = [0, +\infty]$. In Fig. 3.7, the scatter diagrams between X and the powers of X (from the second until the fifth) are presented. As may be seen, the linear relationship between the r.v. is reduced as we move to higher powers of X , although they continue to be perfectly dependent. The product moment correlation matrix for the random vector (X, X^2, X^3, X^4, X^5) is:

$$\mathbf{R}_\rho = \begin{pmatrix} 1 & 0.939 & 0.828 & 0.713 & 0.606 \\ 0.939 & 1 & 0.965 & 0.890 & 0.801 \\ 0.828 & 0.965 & 1 & 0.976 & 0.920 \\ 0.713 & 0.890 & 0.976 & 1 & 0.982 \\ 0.606 & 0.801 & 0.920 & 0.982 & 1 \end{pmatrix}$$

In this case, the insufficiency of ρ to depict dependence is evident: since the r.v. are perfectly dependent, one would expect a correlation close to 1.

Thus, when we leave the world of normal distributions, the product moment correlation fails as a measure of dependence and may lead to counter-intuitive results. The

mechanism of perfect correlation actually corresponds to all *monotonic* (and not necessarily linear) relationships between r.v.: 'two r.v. are perfectly correlated if the one follows the fluctuations of the other'. This is depicted in the scatter diagrams in Fig. 3.6 and Fig. 3.7. An adequate measure should be able to capture the monotonic relationship between the r.v. This leads to the introduction of the *Spearman-* or *rank-* correlation.

3.6 Rank correlation: transformation of marginals into ranks

3.6.1 Rank correlation

The idea behind rank correlation is simple: instead of measuring the correlation between the real values of the r.v., we first rank the samples from lowest to highest, and then measure the product moment correlation for the respective ranks. This captures the monotonic relationship between the r.v. In this case, the marginal distributions have no effect on the value of the measure.

Based on this idea, the calculation of the rank correlation from a population of N pairs of samples (x_i, y_i) for the random vector (X, Y) is straightforward. First we replace the value for each x_i by the value of its rank among the other x_i 's in the sample, that is $1, 2, \dots, N$. If the x_i 's are all distinct then each integer will occur precisely once. If some of the x_i 's have identical values, then we assign to these 'ties' the mean of the ranks that they would have had if their values were slightly different (for more information about how to deal with ties, see Press et al. [73]). We apply the same procedure for the y_i 's. Let R_i be the rank of x_i and S_i the rank of y_i among the other y 's. Then the rank correlation can be calculated as the product moment correlation of the N pairs of samples S_i and R_i , using the population product moment correlation of eq. 3.5.

The transformation of the marginal distributions into ranks is achieved by the application of the cumulative density function $F(\cdot)$ to the r.v. This transformation yields a uniform distribution. In particular, by definition $F_X(X)$ follows a uniform distribution U on the interval $[0, 1]$. For $r \in [0, 1] : P(F_X(X) \leq r) = P(X \leq F_X^{-1}(r)) = F_X[F_X^{-1}(r)] = r \Leftrightarrow F_X(X) = U$ ^{||}.

When applying this transformation in the case of two correlated r.v. X and Y , the uniform distributions $U_X = F_X(X)$ and $U_Y = F_Y(Y)$ correspond to ranks that retain the dependence structure between X and Y . Based on this transformation, the rank correlation may be defined as:

Definition 3.6.1 (Rank Correlation) *The rank correlation of r.v. X, Y with cdf F_X and F_Y , is*

$$\rho_r(X, Y) = \rho(F_X(X), F_Y(Y)) \quad (3.6)$$

Thus, the rank correlation is the product moment correlation of the ranks of the respective r.v. or a correlation of r.v. transformed to uniforms. For every r.v. X and Y , the basic properties of rank correlation are the following:

^{||}We already presented this basic transformation in the section 2.9.1, for the sampling of a r.v. X . As mentioned, $F_X^{-1}(U)$ follows the distribution of X , where U is a uniform distribution in the interval $[0, 1]$.

1. *Range*: $-1 \leq \rho_r(X, Y) \leq 1$
2. *Independence*: if X, Y are independent, then $\rho_r(X, Y) = 0$
3. *Invariance under non-linear monotonic transformations*:
 if $G : \mathbb{R} \rightarrow \mathbb{R}$ a strictly increasing function, then $\rho_r(X, Y) = \rho_r(G(X), Y)$
 if $G : \mathbb{R} \rightarrow \mathbb{R}$ a strictly decreasing function, then $\rho_r(X, Y) = -\rho_r(G(X), Y)$
4. *Monotonic dependence*: if $\rho_r(X, Y) = 1$ then there exists a strictly increasing function $G : \mathbb{R} \rightarrow \mathbb{R}$, $X = G(Y)$

In contrast to the product moment correlation, the rank correlation always exists. For uniform variables, product moment and rank correlations are the same ($\rho = \rho_r$), but in general they are different. For the *joint normal distribution*, the relationship between ρ and ρ_r is known:

Proposition 3.6.2 *Let (X, Y) be random vectors with joint normal distribution, then $\rho(X, Y) = 2 \sin\left(\frac{\pi}{6} \rho_r(X, Y)\right)$*

Thus, the rank correlation is a suitable measure of dependence that is not affected by the marginal distributions. We may see it by measuring the rank correlation for the the random vector (X, X^2, X^3, X^4, X^5) of the previous example; in all cases the same value $\rho_r = 1$ is obtained, corresponding to the case of perfect dependence.

3.6.2 Decoupling of dependence structure and marginals: towards copulas

The basic principle for the transition from the product moment correlation to the rank correlation is the decoupling of the dependence structure from the marginals. This is achieved by the transformation of the marginals into respective ranks by the application of the cumulative density function transformation; these ranks maintain the information concerning the dependence structure between the r.v. This dependence may be depicted by the scatter diagrams of the ranks, i.e. the transformed uniforms. In Fig. 3.8 we present these scatter diagrams for the case of independence (Fig. 3.8c) and perfect dependence (Fig. 3.8f), as they are obtained for the case of two normal distributions (loads) and two Weibull distributions (winds). In both cases, although the scatter diagrams of the non-transformed marginals are different, they yield the same bivariate distribution for the ranks. This bivariate distribution corresponds to the dependence structure. In the case of independence, the samples are distributed uniformly in the unit square, corresponding to the definition of independence between the r.v.: 'the occurrence of one of them has no influence on the probability of the other'. Indeed, for any rank of the one r.v., all ranks between 0 and 1 for the other may occur. In the case of perfect dependence, the samples are ordered and are distributed in the diagonal of the unit square, from the points $[0, 0]$ to $[1, 1]$. In this case, the occurrence of one rank for the one r.v. commits the occurrence of the same rank for the other r.v.

These two cases correspond to the extremes of total randomness (independence) and total order (perfect dependence). Reality however in most cases falls somewhere in between those extremes; this motivated the development of other distributions with uniform

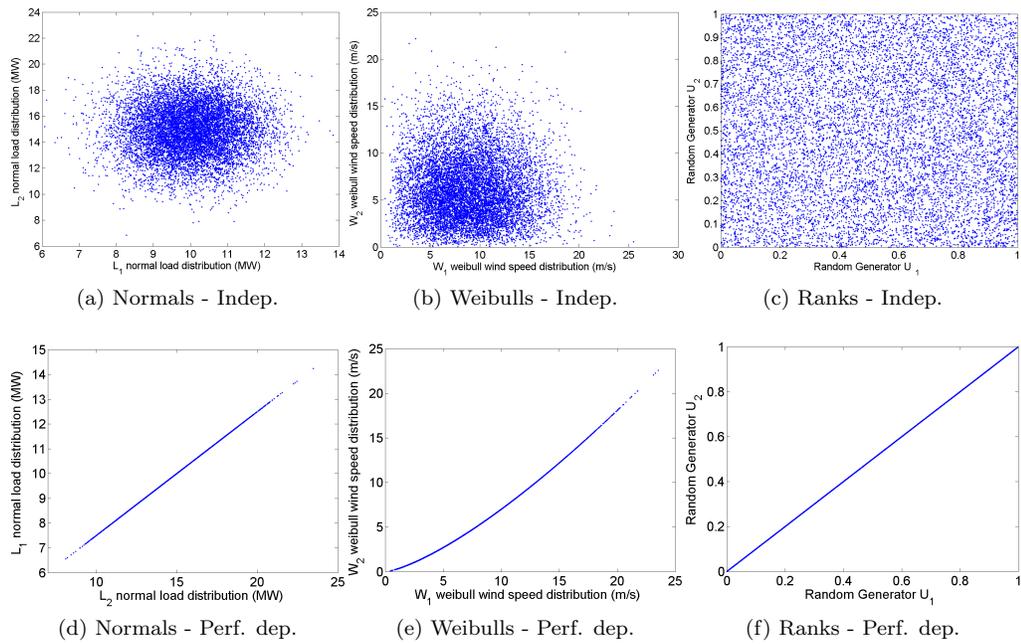


Figure 3.8: Scatter diagrams for independence and perfect dependence between normal and Weibull distributions and their respective ranks.

marginals for the modeling of these dependencies. These are the distributions of the ranks of the r.v. and offer the modeling platform for dependence analysis, by separating the dependence structure from the one-dimensional marginal distributions. In particular, the stochastic analysis may be decomposed in two basic components:

- the one-dimensional marginal distributions
- the stochastic dependence structure between the input r.v.

By applying the transformation $F_X(X)$ and $F_Y(Y)$, we no longer have to deal with the exact numbers, but with ranks of the respective distributions. Then we may use these distributions on ranks and model the functional dependence between X and Y . The investigation of the properties of these distributions led to the introduction of the notion of *copulas* for the modeling of stochastic dependence in uncertainty analysis.

3.7 Copulas

Copulas are functions that join or 'couple' multivariate distribution functions to their one-dimensional marginals. Alternatively, copulas are multivariate distribution functions whose one-dimensional marginals are uniform on the interval $[0, 1]$ [63].

3.7.1 Sklar's Theorem

The word *copula* was first employed in a mathematical or statistical sense by Abe Sklar in 1959 [87] in the theorem (which now bears his name) describing the functions which 'join together one-dimensional distribution functions to form multivariate distribution functions'. The theorem that bears his name offers the bridge between copula and joint distribution of a number of r.v.:

Definition 3.7.1 (Copula (Sklar's Theorem [87])) *The r.v. X and Y with cdf F_X and F_Y , are joint by copula C if their joint distribution can be written:*

$$F_{XY}(x, y) = C(F_X(x), F_Y(y)) \quad (3.7)$$

If F_X and F_Y are continuous, then C is unique; otherwise, C is uniquely determined on $[\text{Ran}(F_X) \times \text{Ran}(F_Y)]$. Conversely, if C is a copula and F_X and F_Y are distribution functions, then the function F_{XY} defined by 3.7 is a joint distribution function with marginals F_X and F_Y .

As Sklar notes however, the functions themselves predate the use of the term copula. They appear in the work of Fréchet, Dall' Aglio, Féron and many others in the study of multivariate distributions with fixed univariate marginals. Indeed, many of the basic results about copulas can be traced to the early work of Wassily Hoeffding. In [44], one finds bivariate '*standardized distributions*' whose support is contained in the unit square $[-\frac{1}{2}, \frac{1}{2}]^2$ and whose marginals are uniform on the interval $[-\frac{1}{2}, \frac{1}{2}]$. It seems that if Hoeffding had chosen the unit square $[0, 1]^2$ instead of $[-\frac{1}{2}, \frac{1}{2}]^2$ for his normalization, he would have discovered copulas.

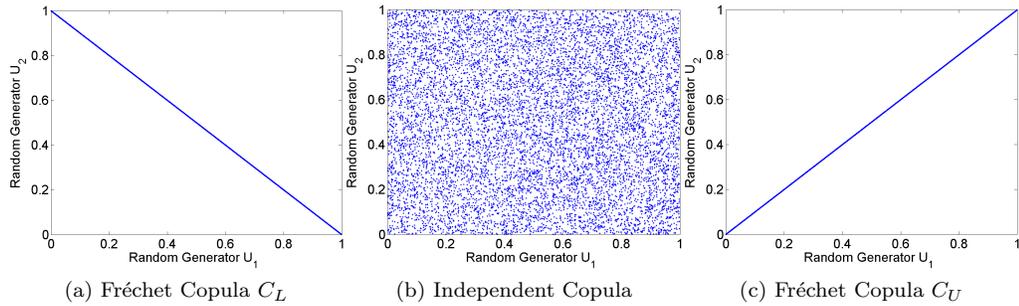


Figure 3.9: Scatter diagrams for the Independent and Fréchet copulas.

3.7.2 Fréchet-Hoeffding bounds

Hoeffding also obtained the basic best possible bounds inequality for these functions, characterized the distributions corresponding to those bounds and studied measures of dependence which are 'scale-invariant', i.e. invariant under strictly increasing transformations. Until recently, Hoeffding's work didn't receive the attention it deserved, due primarily to the fact that his papers were published in relatively obscure German journals at the outbreak of the Second World War. Unaware of Hoeffding's work, Fréchet in 1951 [40] independently obtained many of the same results, which has led to terms such as 'Fréchet-Hoeffding bounds'.

In particular, it can be shown that for any copula C the following inequality holds [40]:

$$C_L \leq C \leq C_U \quad (3.8)$$

where C_L and C_U are bivariate distributions such that the mass is spread uniformly on the main anti-diagonal and diagonal respectively. The copulas C_L and C_U describe perfect negative and positive dependence respectively:

Proposition 3.7.2 *If X, Y are joined by copula C_U (C_L), then $\rho_r(X, Y) = 1(-1)$.*

Thus, the copula inequality (3.8) may be interpreted in the following way: "all possible dependence structures between two r.v. are bounded between the cases of perfect negative (lower bound) and perfect positive dependence (upper bound)". These bounds may be used as worst-case scenarios for the power system stochastic modeling and the clustering of the random system inputs. The thorough investigation of the properties of these bounds and the application of the Fréchet inequality to power system stochastic modeling is the subject of chapter 5, where the concept of *stochastic plants* is introduced. In Fig. 3.9 the scatter plots for the Fréchet copulas C_L and C_U are presented next to another basic copula, the *independent copula*.

3.7.3 Families of Copulas

After Hoeffding, Fréchet and Sklar, the functions now known as copulas were rediscovered by several other authors, referring to them with different names, as *uniform representations* or *dependence functions*. The research in this field yielded important results about

several families of copula functions. In Fig. 3.9, three basic copulas were presented: the independent copula and the Fréchet copulas C_L and C_U . In the following we present an overview of some other basic families of copulas. For details on the properties of the different families of copulas, one may refer to [48, 63, 53, 51].

Normal (Gaussian) Copula

The *Normal or Gaussian copula* is obtained from the bivariate Normal distribution, by the transformation of the standard normal marginals into uniforms (ranks)**. In Fig. 3.10a, the scatter diagram for this copula is presented for rank correlation $\rho_r = 0.6$.

Its multidimensional extension is obtained in the same way, by transforming the standard normal marginals into uniforms, starting from the multivariate normal distribution. The normal copula is one of the most widely used copulas. It is uniquely determined by the product moment correlation matrix and it forms the basis for techniques for multi-dimensional dependence modeling, e.g. the *Joint Normal Transform* methodology. We elaborate further on this copula in section 4.1, where this methodology is presented.

The main disadvantage of the normal copula is that it fails to represent asymptotic tail dependence, i.e. dependence between the extreme values of the r.v. In particular, as shown by Embrechts et al. in [38], the Normal copula gives asymptotic independence, provided that $\rho < 1$. Regardless of how high a correlation we choose, if we go far enough into the tail, extreme events appear to occur independently in each margin. In order to represent tail dependence, adequate extensions of the Normal distribution should be used. These extensions form the family of *elliptical distributions*.

Elliptical Copulas

The elliptical distributions constitute generalizations of the multivariate normal distribution, which allow for the presence of heavy tails and asymptotic tail dependence. Mathematically they are the affine maps of spherical distributions in \mathbb{R} [38, 52]. As in the case of the Normal copula, the bivariate (and multivariate) elliptical copula is obtained by the transformation of the marginals of the respective elliptical distribution into uniforms and is uniquely determined by the correlation matrix. The advantage offered by the elliptical distributions is that some of the members of this family permit the modeling of tail dependence.

The *t copula* is one of the most widely used members of this family. It derives from the multivariate t distribution and, in contrast to the normal copula, the t copula is asymptotically dependent, provided $\rho > -1$ [38]. As presented in [33], this copula forms the basis for the construction of *t-related copulas*, that present different types of tail dependencies.

Archimedean Copulas

Another very popular family of copulas are the *Archimedean copulas*. Properties of this family have been studied extensively [63, 48]. They find a wide range of applications for three main reasons: (1) the ease with which they can be constructed; (2) the great

**This transformation is performed by the application of the standard normal cdf $\Phi(\cdot)$.

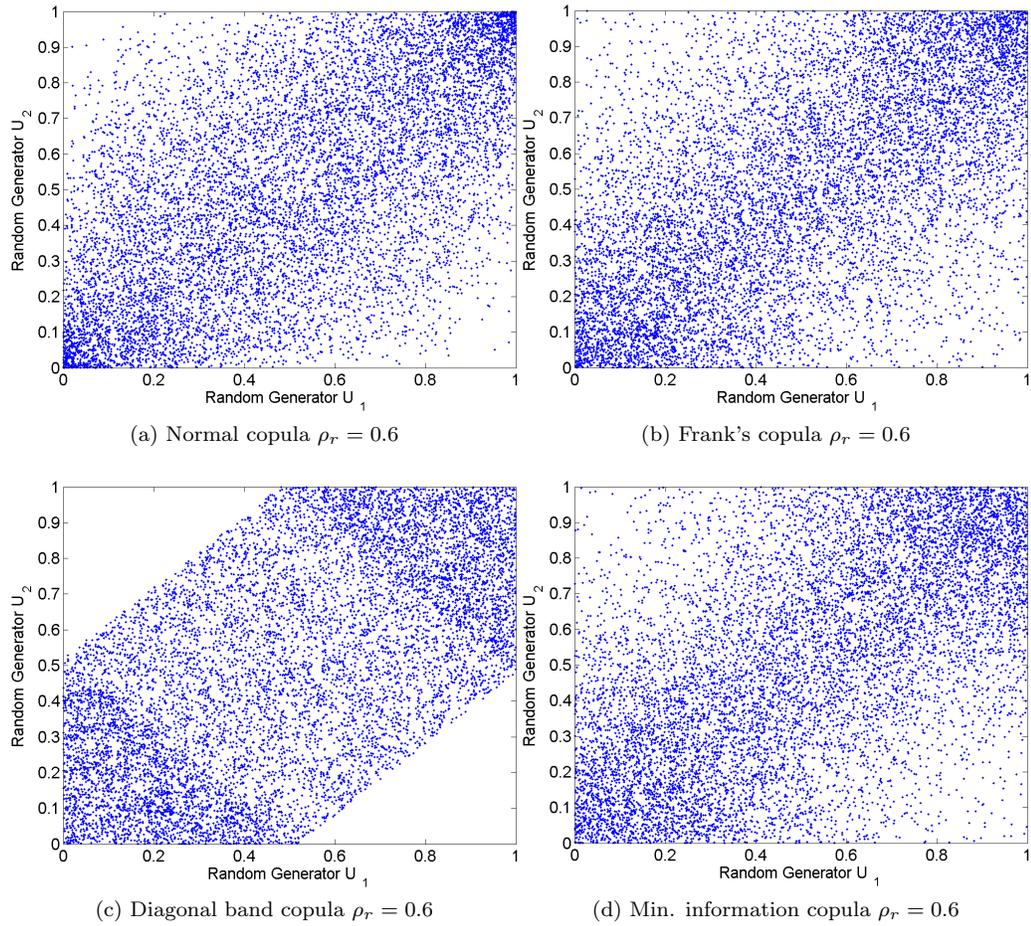


Figure 3.10: Scatter diagrams for four types of copulas under the same rank correlation.

variety of families of copulas which belong to this class; and (3) the many nice properties possessed by members of this class.

Two of the most widely used members of this family is the *Gumbel copula* and the *Frank's copula*. For a detailed presentation of the Archimedean copulas one may refer to [63]. In Fig. 3.10b the scatter diagram for the Frank's copula is presented, for rank correlation $\rho_r = 0.6$.

Diagonal Band Copula

The *diagonal band copula* is a natural generalization of the Fréchet copula [51]. In contrast to the Fréchet copula, for positive correlation the mass is concentrated on the diagonal band with vertical bandwidth $\beta = 1 - \alpha$. Mass is distributed uniformly on the inscribed rectangle and is uniform but 'twice as thick' in the triangular corners. For negative

correlation the band is drawn between the other corners. In Fig. 3.10c the scatter diagram for the diagonal band copula is presented, for rank correlation $\rho_r = 0.6$.

The construction of the diagonal band copula can be seen as putting the uniform density on $[-\beta, \beta]$ and translating this along the diagonal, always folding the mass that lies outside the unit square inwards. This procedure can be generalized by putting a non-uniform density $G(z)$ along lines perpendicular to the U axis such that $G(0)$ lies on the diagonal. In this way, a band of densities $G(z)$ will be constructed. This is the procedure for obtaining the *generalized diagonal band copula*. For further details, one may refer to [51].

Minimum Information Copula

The development of the *minimum information copula* came from the application side: for the simulation of two arbitrary continuous marginal distributions F_X, F_Y correlated with $\rho_r \in (-1, 1)$, when no further information on their joint distribution is available, we need to use a joint distribution that has minimal information with respect to the independent distribution with F_X and F_Y . This joint distribution is proved to be unique and is calculated as the solution of an optimization problem. For the case of uniform marginals this unique solution corresponds to the minimum information copula.

This copula is attractive because it realizes a specified rank correlation by adding 'as little information as possible' to the product of the margins, i.e. to the joint distribution corresponding to independence. Its main disadvantage is that it doesn't have a closed functional form. All calculations with this copula must involve numerical approximations. Further details on the minimum information copula are presented in [51]. In Fig. 3.10d, the scatter diagram for the minimum information copula is presented, for rank correlation $\rho_r = 0.6$.

3.7.4 Copulas and rank correlation

In Fig. 3.10 the scatter diagrams of four different copulas are presented for the same rank correlation. Although we keep the same rank correlation, the scatter diagrams present basic differences, corresponding to different dependence structures. As may be seen, rank correlation is not enough for the representation of dependence; this representation comes from the use of copulas.

As the rank correlation changes from -1 to 1, the copulas provide different representations of the dependence. In Fig. 3.11 we show how this representation changes, by presenting the scatter diagrams for the diagonal band copula for different values of ρ_r . For $\rho_r = 1$ and $\rho_r = -1$, they correspond to the Fréchet bounds C_L and C_U , as mentioned in proposition 3.7.2. While shifting towards $\rho_r = 0$, the dispersion of samples around the main diagonals increases. The case of $\rho_r = 0$ however does not always correspond to independence, as it was presented by the independent copula in Fig. 3.9b. As mentioned in the basic properties of rank correlation, independence corresponds to $\rho_r = 0$, but not vice-versa.

Coming back to our specific application, in Fig. 3.12, the power output from two WTGs is presented for two cases, where the wind at the respective sites presents the same rank correlation but is subject to different copula. Despite the same rank correlation, different power output distributions are obtained.

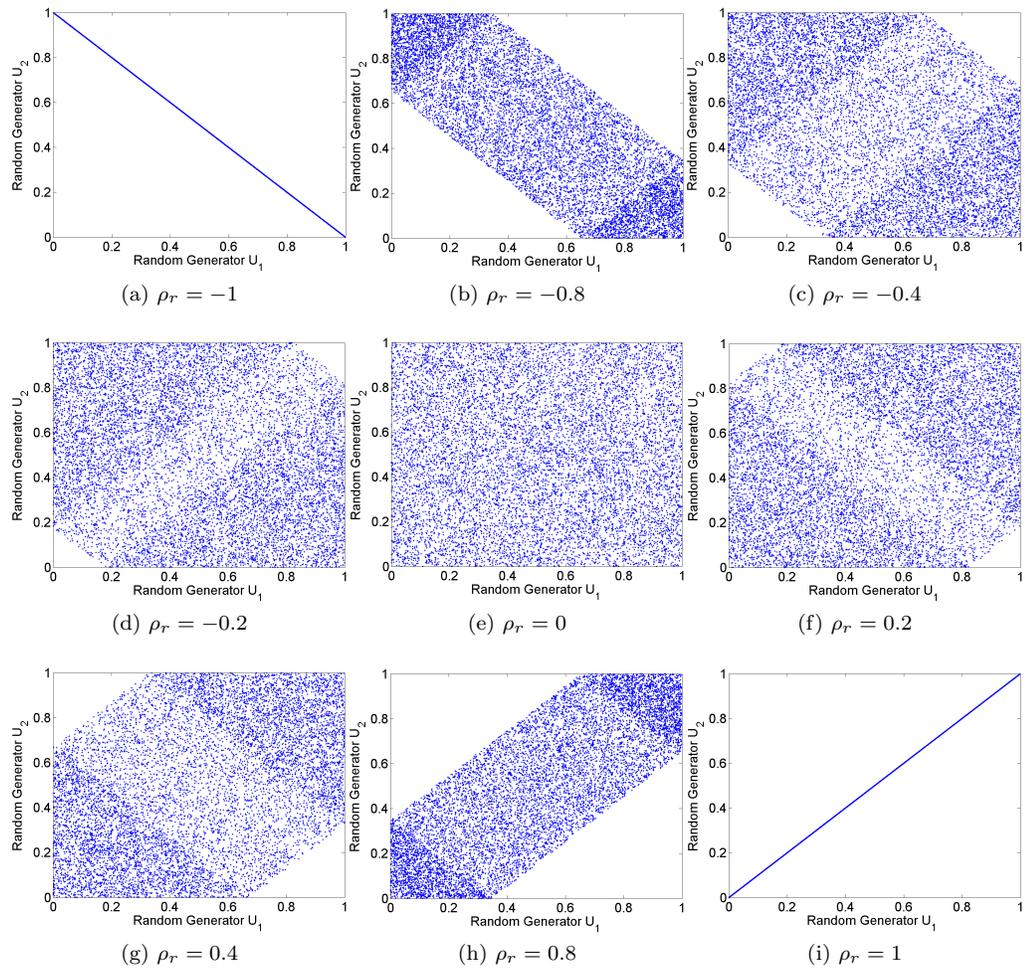


Figure 3.11: Scatter diagrams for the diagonal band copula for different rank correlations.

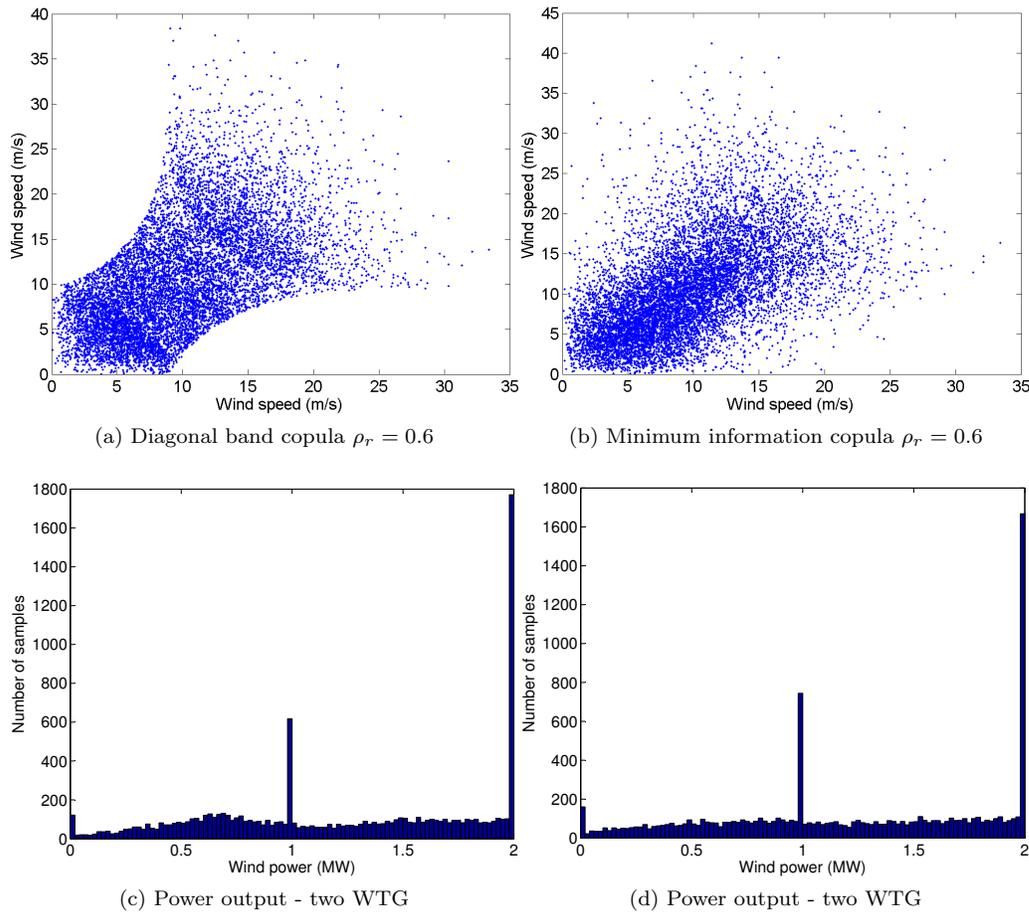


Figure 3.12: Wind scatter diagrams and power output distributions for two WTGs under the same rank correlation but with different dependence structures.

One may argue that this difference is quite small for using such complex models. This is true. The basic advantage of using copulas comes from their application in stochastic simulations; in particular, the use of copulas enables the simulation of correlated r.v.

3.7.5 Sampling of two correlated r.v. using copula

Summarizing the preceding analysis, we may see that for the modeling of two correlated r.v. X_1, X_2 , three basic elements are necessary: the one-dimensional marginal distributions F_1 and F_2 , the rank correlation ρ_{r12} and the copula $C_{12|\rho_{r12}}$. The sampling algorithm for the modeling of two r.v. is presented in algorithm 3.2. The core in the modeling is the generation of two uniform (rank) distributions U_1 and U_2 correlated according to a given dependence structure defined by $C_{12|\rho_{r12}}$. These rank distributions are later trans-

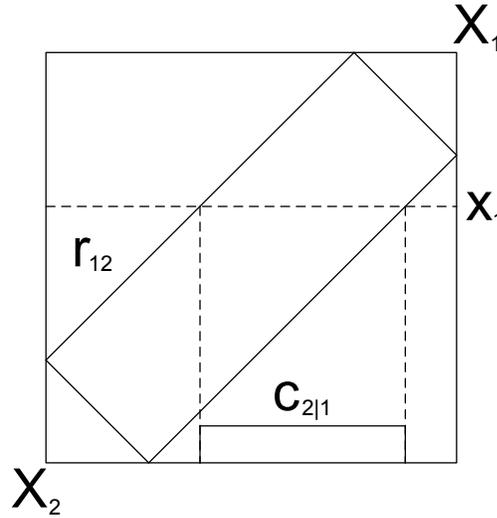


Figure 3.13: Sampling procedure for the diagonal band copula.

formed to the respective marginals. For the generation of these correlated uniforms we start by sampling two uniform r.v. U_{r_1} and U_{r_2} independently. In order to impose the dependence structure, we sample U_2 conditional to $U_1 = U_{r_1}$ by using the conditional copula $u_2 = C_{12|u_1, \rho_{r12}}^{-1}(u_{r_2})$.

Algorithm 3.2 Simulation algorithm for modeling two correlated r.v.

1. Sample two independent uniform r.v. U_{r_1} and U_{r_2} . Denote realizations u_{r_1} and u_{r_2} .
 2. $u_1 = u_{r_1}$: the sampling of the rank distribution U_1 .
 3. Find $C_{12|u_1, \rho_{r12}}$. This is the conditional distribution of U_2 given u_1 for rank correlation ρ_{r12} .
 4. By the inverse conditional copula we sample the rank distribution U_2 : $u_2 = C_{12|u_1, \rho_{r12}}^{-1}(u_{r_2})$.
 5. Transform the rank distributions according to the marginals: $x_1 = F_1^{-1}(u_1)$ and $x_2 = F_2^{-1}(u_2)$.
-

In Fig. 3.13, the sampling procedure for a diagonal band copula C_{12} based on this algorithm 3.2 is presented. As may be seen, the conditional diagonal band copula corresponds to a distribution $C_{2|1}$ that is constant for the cross section of $C_{2|1}$ and the ordinate $u_1 = u_{r_1}$, and zero outside. The 'dependent' samples u_2 are obtained by sampling the conditional copula $u_2 = C_{2|1}^{-1}(u_{r_2})$. Therefore, the outcome in this case is either 0 or constant, depending on the value of the independent sample u_{r_2} .

3.8 Conclusions

Studying the impact of stochastic generation in power systems requires the modeling of correlated non-normal distributions. Modeling the stochastic dependence is a main part in this procedure. Although the product moment correlation ρ is generally considered as a good representation of dependence, when we leave the world of normal distributions it fails as a measure of dependence, since it cannot capture non-linear, monotonic relationships, which is exactly what one usually expects to know about a relationship between variables. In order to capture such relationships, we should transform the one-dimensional marginal distributions into ranks. The product moment correlation of ranks of the respective variables, i.e. the rank correlation ρ_r , corresponds to an adequate measure of dependence, since it captures monotonic relationships between variables. This transformation of the marginals into ranks is the cornerstone in the proposed stochastic modeling, offering a decoupling between the modeling of the marginal distributions and the stochastic dependence structure. Based on this decoupling, specific functions are used for the modeling of the functional relationship between these ranks, offering significant advantages for the stochastic simulation. These are the copula functions. The properties of different copula families were presented, together with the basic simulation algorithm for the use of these functions for the stochastic modeling of two correlated random variables.

Hence, the simulation of bivariate relationships between dependent r.v. is performed using the marginal distributions, the rank correlation and the respective copula functions. What remains is the extension of these principles to the multidimensional case. The solution to this problem is presented in the following chapter.

Chapter 4

Multidimensional Dependence Modeling

In this chapter we present the extension of the bivariate stochastic dependence models to the multidimensional case, in particular the representation of the joint distribution of these r.v. based on the information of their mutual (bidimensional) stochastic behavior.

The approach presented in the previous chapter, i.e. the splitting of the modeling effort in the two tasks of modeling the marginals and the bivariate dependence structure may also be applied to the multivariate case. In general, the one-dimensional marginal distributions are the easiest to assess from data or expert judgement. The definition of the dependence structure is however a much more complex problem. In the case of N r.v., there are $2^N - 1$ distinct non-empty subsets of variables, each of which could have a distinct dependence structure. In this chapter, appropriate methodologies are employed for the stochastic modeling of such problems.

Data availability and adequacy is a basic problem faced in real applications. Driven from this perspective, we can distinguish two basic types of problems:

1. problems where adequate data are available, and
2. problems that are poorly or not supported by data.

In the first type of problems, one can calculate the product moment/rank correlation matrix by the time-series data describing the stochastic inputs in the system and use a multidimensional copula for the dependence modeling. The *Joint Normal Transform* method is presented as the basic methodology for the treatment of such problems. The second type of problems involves high uncertainty, where it is not possible to define all mutual correlations between the system stochastic inputs. In such problems, we typically try to capture the most important dependence relations and leave others unspecified. For this, we present graphical methods for specifying dependence structures which can deal with incomplete specifications and sampling routines that give *exactly* what we specify, up to sampling error, with minimal additional information. These are the *Copula-Tree* and *Copula-Vine* methods.

4.1 Joint Normal Transform methodology

For most statisticians, the starting point for dependence modeling is the joint normal distribution. The *Joint Normal Transform (JNT)* methodology is a natural extension of the method for the modeling of correlated normals presented in section 3.4. This method involves two main tasks for the modeling of an n -vector of r.v. \mathbf{X} , correlated according to a rank correlation matrix \mathbf{R}_r :

1. *Dependence structure*: generate an n -vector of correlated uniform rank distributions \mathbf{U}_C using a multivariate normal copula, according to the theory presented in section 3.7.3.
2. *Marginals*: transform the correlated ranks \mathbf{U}_C into the desired marginals X_i , by applying the inverse cdf transformation $X_i = F_i^{-1}(\mathbf{U}_C)$.

A note should be made concerning the use of a multivariate normal copula for the generation of the n -vector of correlated ranks \mathbf{U}_C . According to the theory of the joint normal distribution (section 3.4), we may construct a n -dimensional vector of standard normals \mathbf{Y} that realize a given *product moment correlation matrix* \mathbf{R} . Since product moment and rank correlations for the joint normal distribution are not equal, the rank correlation of the output samples will be different. In order to obtain the desired rank correlation matrix \mathbf{R}_r , we should start from an appropriate matrix \mathbf{R} . This matrix can be calculated from the relationship between product moment and rank correlation for the normal distribution (proposition 3.6.2):

$$\mathbf{R} = 2 \sin\left(\frac{\pi}{6} \mathbf{R}_r\right) \quad (4.1)$$

In this case, sampling the joint normal based on \mathbf{R} that fulfills this relationship, yields the desired \mathbf{R}_r at the output. According to these points, we may summarize the JNT simulation algorithm as follows [41, 19]:

As mentioned in section 3.4.3, a necessary condition in order to carry out the factorization $\mathbf{R} = \mathbf{T} \times \mathbf{T}^T$ used in this algorithm is that \mathbf{R} has to be *positive semi-definite*. By definition, a correlation matrix fulfills this precondition. However, in real applications it often happens that correlations are estimated by noisy procedures. It may thus arise that a '*measured*' correlation matrix is non-positive semi-definite. This hinders the applicability of the method. In this case, the method may be applied if we repair the violations of positive semi-definiteness.

In the following we demonstrate the applicability of the JNT methodology and elaborate further on these issues by the presentation of a comprehensive example: the impact of large-scale integration of wind power in the power system of The Netherlands. In this example we present how measurement data from different data systems can be combined with expert data for the system uncertainty analysis and how violations of positive semi-definiteness appear in the analysis and are being repaired.

4.2 Wind power integration in The Netherlands

The study case involves the evaluation of the impact of the integration of a total capacity of 1.5GW onshore and 5GW offshore wind power in the power system of The Netherlands

Algorithm 4.1 Joint Normal Transform simulation algorithm.

1. *Dependence structure*: Normal Copula
 - Calculate the transformed correlation matrix \mathbf{R} based on equation 4.1.
 - Simulate an n -vector of independent standard normals $\mathbf{Z} = [Z_1, \dots, Z_n]$.
 - Calculate the matrix product $y = \mathbf{T} \cdot z$, where \mathbf{T} is a lower triangular matrix such that $\mathbf{R} = \mathbf{T} \cdot \mathbf{T}^T$ (*Cholesky decomposition*). The n -vector y forms a sample drawn from a random vector \mathbf{Y} that follows a standard multivariate normal distribution with product moment correlation matrix \mathbf{R} : $\mathbf{Y} = \mathbf{T} \cdot \mathbf{Z}$.
 - Transform the correlated standard normals \mathbf{Y} to correlated ranks using the normal cdf transformation $\Phi(\cdot)$: $\mathbf{U}_C = \Phi(\mathbf{Y})$.
2. *Marginals*:
 - Transform the correlated ranks \mathbf{U}_C into the given marginals $\mathbf{F} = F_i^{-1}(U_C)$.

in 2020, according to the basic scenarios presented in [84]. The r.v. involved in the analysis are the wind speeds at each wind generation site and the system load.

For the analysis, four different data sets are used: measurements for onshore locations, measurements for offshore locations, data from experts concerning one offshore location and load data. The availability and adequacy of the respective data may be categorized as follows:

1. *Well-documented data*: in the specific study case, we consider 10 locations for the onshore wind parks (sites 1-9 and 13 in Fig. 4.1); for each location, hourly wind speed values are available for a period of 20 years from January 1, 1984 until December 31, 2003 [78]. This data set forms a well documented entity, with a small number of missing values. The planned capacities for each onshore wind park are the following:

Wind Park	1	2	3	4	5	6	7	8	9	13
Capacity (MW)	100	100	100	100	100	300	100	100	200	300

The total planned onshore capacity is 1500MW.

2. *Not well documented data*: measurements for 5 offshore sites are available, namely for the sites 11, 12, 14 (near-shore wind parks) and 10 and 15 (offshore wind parks) in Fig. 4.1. The available data set corresponds again to hourly measurements for the same period as for the onshore wind parks. This data set however is not so well documented, presenting a relatively large number of missing values. The total planned offshore capacity is 3500MW. In the following matrix, the planned capacities for each of these locations are presented together with the percentage of missing values for each data set:

Wind Park	10	11	12	14	15
Capacity (MW)	1000	500	500	500	1000
Missing values (%)	8.3	13.8	5.8	6.4	8.8

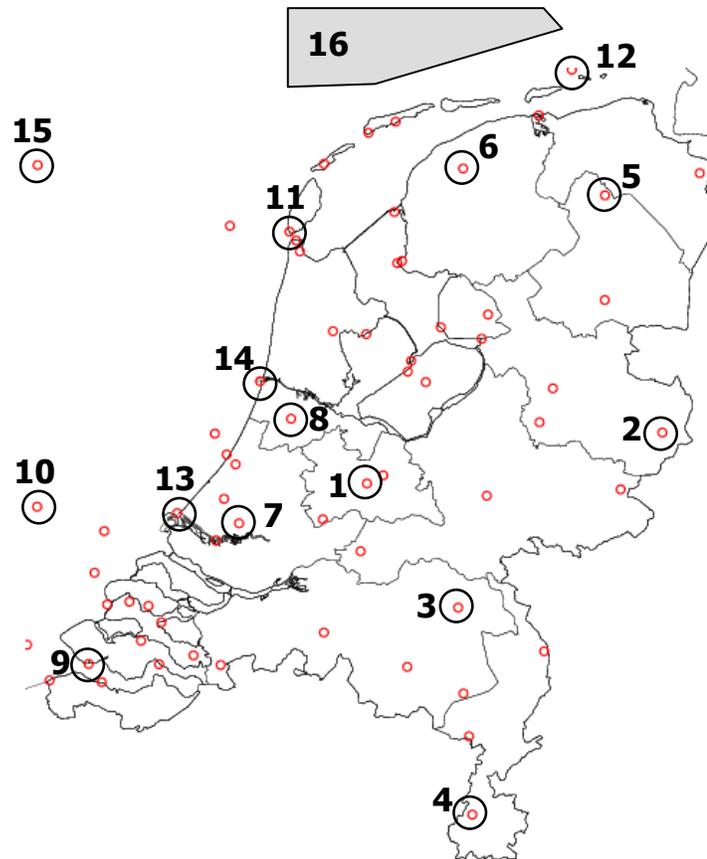


Figure 4.1: Wind park locations in The Netherlands

3. *Expert data*: the second stage in the project planning involves the expansion of the installed capacity by the installation of 1000MW in the shaded area (site 16) in Fig. 4.1. However, no measurements are yet available for this region. For this, experts can be asked to assess the wind in this area. In the following we present how such data can be incorporated in the analysis.
4. *Different data system*: the load data set corresponds to a different recording period and sampling frequency, in particular 15-minute averages for a period of one year, i.e. between January 1 and December 31, 2003. In order to simulate the load for the year 2020, we extrapolate the recorded values based on an annual load growth rate of 2%.

As may be seen, the data sets involved in the analysis are not uniform; they come from different data systems and span different periods with different sampling frequencies. Such a non-uniformity of the data sets is usual for real applications, due to the use of different databases and measuring systems. In such cases, the analysis should be based on extraction and utilization of maximum information from the available data set.

The application of the JNT methodology requires the computation of the following statistics:

1. *Marginal distributions*: the wind speed distributions at each generation site and the system load distribution.
2. *Dependence structure*: the rank correlation matrix between the wind speed r.v. and the system load.

4.2.1 Marginal distributions

Wind speed distributions

The wind speed distributions for the sites where data are available are obtained after adapting the measurement height (10m) to the wind turbine hub height. This adaptation is performed by the use of the *wind shear formula* presented in [29]. Typical pitch-controlled wind turbine generators (WTG) are considered for this project, with a hub height of 80m, a nominal power of 2MW and cut-in, nominal and cut-out wind speed values of 3.5, 12 and 25 [ms^{-1}].

We assume that the elicitation of experts yielded a wind speed distribution in area 16 that is best described as a Weibull distribution with scale parameter $\eta=11.8$ and shape parameter $\beta=2.1$.

In Fig. 4.2, the cumulative distributions for the different sites are presented. Such a graph permits a direct comparison of the wind potentials between different locations; in particular, the more to the right a cdf appears in the graph, the larger the wind potential at the respective location. In the graph, discontinuous lines are used for the onshore sites, continuous lines for the offshore sites and circular markers to denote the expert data cdf. The cdf's for the onshore locations are ordered according to their numbering from left to right (location 1 corresponds to the discontinuous cdf at the left, while location 13 to the one at the far right). As may be seen, the onshore wind potential is in general lower than the offshore and increases the closer the location is to the coast. The coastal location 13

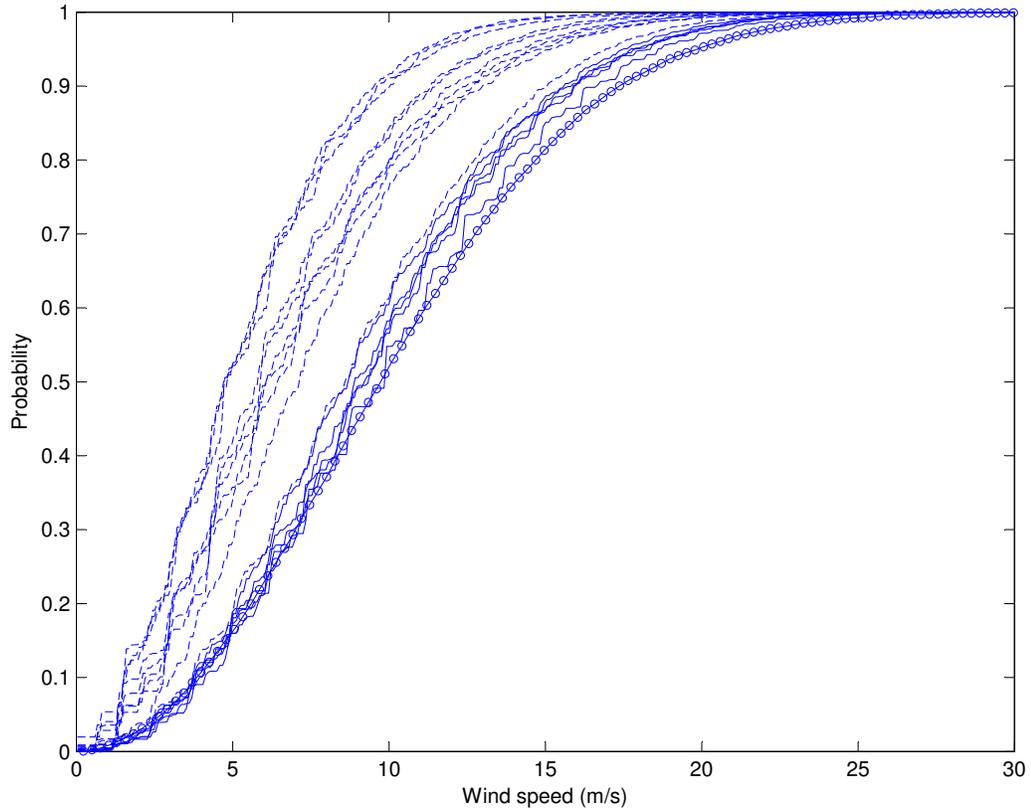


Figure 4.2: Wind speed cdfs for the wind park locations.

presents a wind potential similar to the near-shore locations 14, 12 and 11 (the respective order of the first three continuous cdfs). The offshore locations 10 and 15 correspond to the far-right continuous cdfs, while the wind potential at location 16 appears to be the highest, at the far-right of the graph.

The sampling of the wind speed distributions has been performed based on the empirical distribution obtained from the data analysis, according to the methodology presented in section 2.9.1. In Fig. 4.3, the measured wind speed distributions for two characteristic sites, the onshore site 1 (Fig. 4.3a) and the offshore site 14 (Fig. 4.3b) are compared to the simulated ones obtained from a 20000-sample MCS. In Fig. 4.3e and 4.3f, the respective cdfs are presented. We can see that the simulated distributions are very accurate approximations of the measured ones.

System load distribution

In Fig. 4.4, the pdf and cdf for the extrapolated system load for 2020 are presented. The extrapolation is performed based on a yearly load growth rate of 2%. This results in an expected system load variation for 2020 ranging approximately from 9GW to 19GW. As

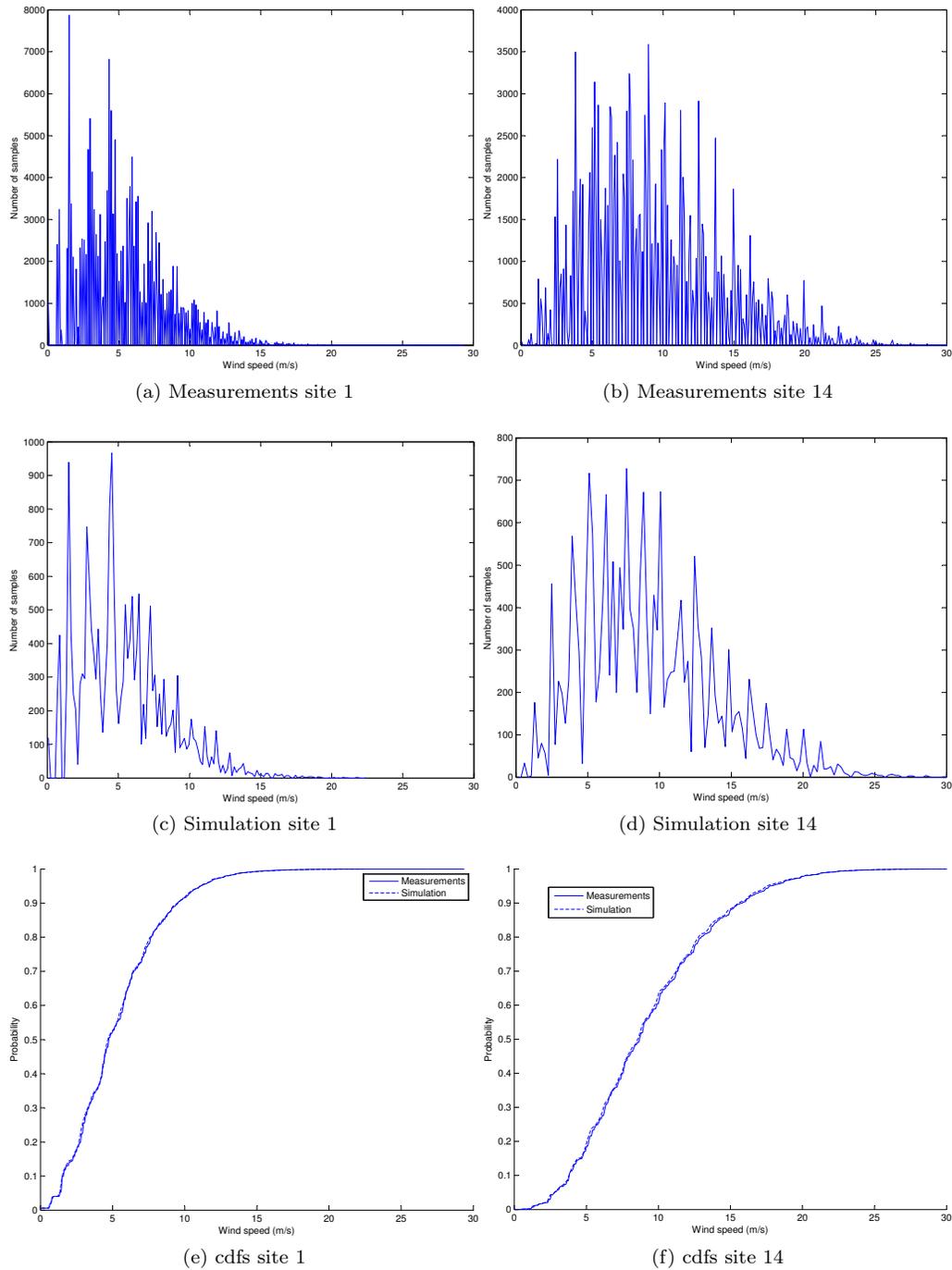


Figure 4.3: Measured and simulated wind speed distributions at site 1 (onshore) and 14 (offshore).

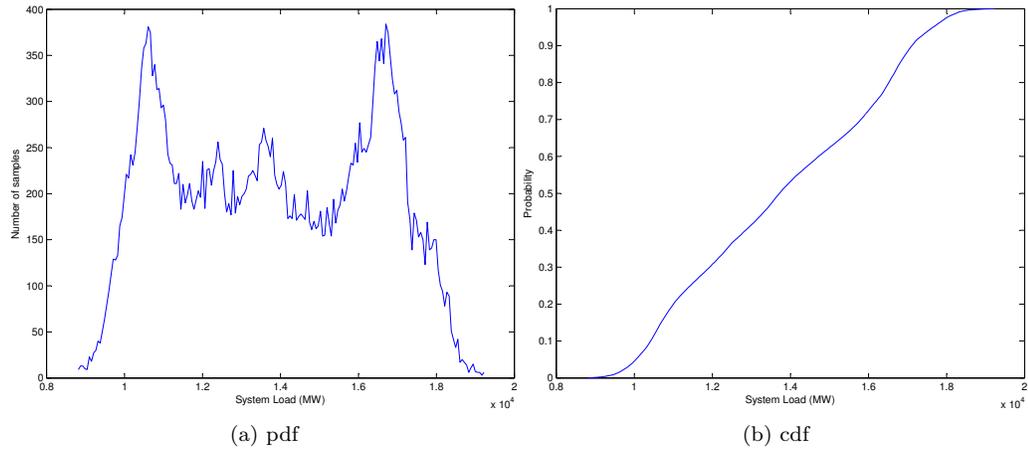


Figure 4.4: System load distributions for the year 2020 obtained by extrapolation of the 15-min recordings of year 2003.

may be seen, the planned wind power capacity corresponds to a significant part of the system demand.

4.2.2 Dependence structure: correlation matrices

For the application of the method, the transformed correlation matrix \mathbf{R} should be calculated from the rank correlation matrix \mathbf{R}_r , based on equation 4.1. Since different datasets are involved in the analysis, the matrix \mathbf{R}_r is calculated in parts, based on the maximum time-intersection of the respective datasets. In particular, the data for the period of 20 years are used for the derivation of the mutual wind speed rank correlations whereas a period of 1 year is used for the wind-load rank correlations.

Wind-load rank correlations

The rank correlations between the wind speeds and system load are calculated for the common year of measurements, according to the lower measurement frequency, i.e. hourly mean values. The obtained rank correlations are presented in the Appendix C.1.

The analysis shows a low correlation between the wind activity and the system load, ranging from zero to 0.24. The onshore wind presents a low rank correlation to the system load; the offshore wind is practically independent from the system load.

Mutual wind speed rank correlations

The rank correlations between the wind speeds at the 15 different onshore and offshore sites are calculated based on the whole data set, consisting of 20 years of hourly measurements. The obtained rank correlations are presented in the Appendix C.1. The results

show that the rank correlations are in between 0.47 and 0.91. Hence, the wind resources are correlated even for distant locations throughout the country.

By merging these submatrices, we obtain the 16×16 wind speed/system load rank correlation matrix $\mathbf{R}_r^{\text{WL}*}$.

Rank correlations assessed by experts

The assessment of experts for the rank correlations between the wind activity at location 16 of Fig. 4.1 and the rest of the r.v. involved in the analysis is assumed to yield the rank correlations presented in the Appendix C.1. [†] According to the expert data, the wind resource in area 16 appears to be independent from the system load. The correlation with the wind resources at the other locations ranges from 0.4 to 0.9. The incorporation of the rank correlations obtained by experts in the matrix \mathbf{R}_r^{WL} , yields the 17×17 rank correlation matrix $\mathbf{R}_r^{\text{WLe}}$.

Two stages in the project planning are investigated separately:

1. *5GW-integration scenario*: the incorporation of a total wind power capacity of 5GW, i.e. without the extra capacity of area 16, and
2. *6GW-integration scenario*: the installation of 1GW of wind power in area 16 is included as well.

In the first scenario, the matrix \mathbf{R}_r^{WL} is used to model the system dependence structure, while in the second one the matrix $\mathbf{R}_r^{\text{WLe}}$ is used.

4.2.3 5GW-integration scenario: positive semi-definite matrix

According to the simulation procedure presented in algorithm 4.1, first the transformed correlation matrix \mathbf{R}^{WL} is calculated from the rank correlation matrix \mathbf{R}_r^{WL} according to equation 4.1. Both matrices are positive semi-definite, so the method can be applied directly.

In Fig. 4.5, the aggregate onshore and offshore wind power for this scenario are presented. As may be seen, the different resources and dependence structures of onshore and offshore wind power lead to completely different distributions for the aggregate wind power. An accumulation of probability mass in the nominal power output of the offshore wind parks is observed, due to the operation in the nominal wind power area. In Table 4.1, the capacity factor η_W for the wind parks is presented. As may be seen, the onshore wind parks present a lower capacity factor than the offshore ones. The maximum capacity factor of 0.54 is obtained in wind park location 15.

In Fig. 4.6, the aggregate wind power for the system is presented, together with the net load distribution. The onshore and offshore wind power distributions add up to a distribution that has a uniform shape. The system net load distribution (Fig. 4.6b) alters drastically compared to the situation before the incorporation of 5GW of wind power (Fig. 4.6a).

*The notation $\{WL\}$ here refers to the correlation matrix containing the correlations between wind speeds (mutual) and system load.

[†]For details on expert elicitation and expert judgement, one may refer to [23] and [51].

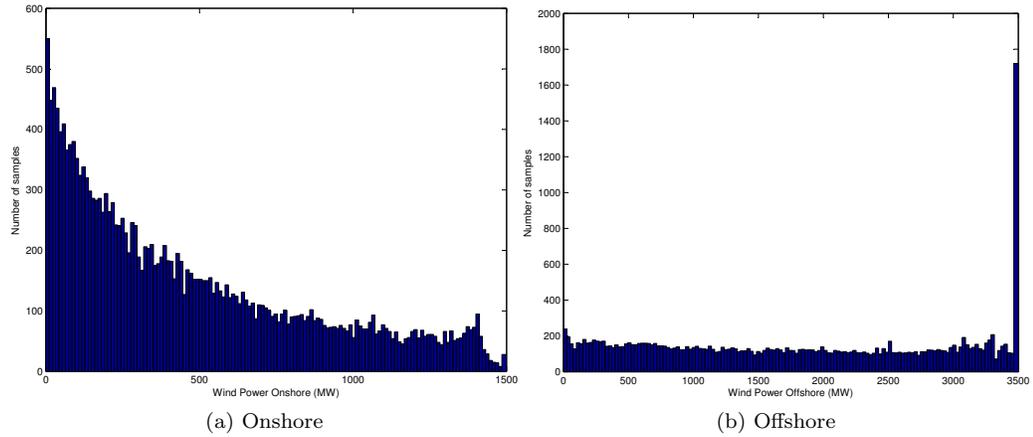


Figure 4.5: Onshore and offshore wind power distributions for the 5GW-integration scenario.

Table 4.1: Capacity factor for the wind parks in The Netherlands.

Windpark	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8
η_W	17.8	18.1	18.3	24.7	25.9	29.1	29.1	31
Windpark	W_9	W_{10}	W_{11}	W_{12}	W_{13}	W_{14}	W_{15}	
η_W	34.4	51.8	50.2	51.8	47.1	47.7	54.2	

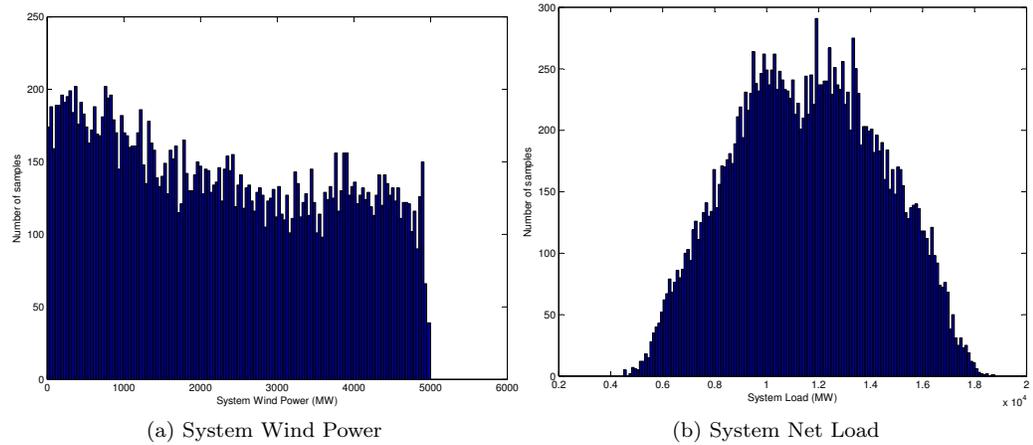


Figure 4.6: Total system wind power and system net load distributions for the 5GW-integration scenario.

4.2.4 6GW-integration scenario: non-positive semidefinite matrix

The second stage in the project planning involves the incorporation of an extra 1GW of wind power, resulting in a total of 6GW wind power in the system. In this case, the rank correlations obtained by experts are incorporated and the 17×17 matrix $\mathbf{R}_r^{\text{WLe}}$. According to the methodology, this matrix should be transformed to the matrix \mathbf{R}^{WLe} that is used for the sampling of the normal copula.

However, both these actions, namely the incorporation of the experts data and the transformation of the matrix $\mathbf{R}_r^{\text{WLe}}$, yield non-positive semi-definite matrices. In order to proceed, we should repair these violations of positive semi-definiteness. Repairing such violations comes down to the transformation of the initial non-positive semi-definite matrix $\mathbf{R}_r^{\text{WLe}}$ to a positive semi-definite one, namely $\mathbf{R}^{\text{WLe-psd}}$, that is as close as possible to the original. This is the most delicate part in the modeling based on the JNT methodology. A thorough analysis on the problems and repairing methods concerning positive semi-definiteness follows in the next section 4.3.

After the application of the repairing methods, the positive semi-definite matrix $\mathbf{R}^{\text{WLe-psd}}$ is obtained and used for the system analysis. In Fig. 4.7, the deviation between the non-positive semi-definite rank correlation matrix $\mathbf{R}_r^{\text{WLe}}$ and the one obtained after the application of the repairing method ($\mathbf{R}^{\text{WLe-psd}}$) is presented. We can see that, in general, the new matrix presents only a small deviation from the original one.

In Fig. 4.8, the aggregate onshore and offshore wind power for the second stage in the project planning is presented. In Fig. 4.9, the aggregate wind power for the system is presented, together with the net load distribution. The installation of the new offshore wind park leads to a more uniform distribution of the aggregate wind power in the system. The lower tail in the system net load distribution starts from smaller values when compared to the net load distribution of the first part of the project planning (Fig. 4.4).

4.3 Positive semi-definiteness: Problems and repairing methods

As mentioned, a necessary condition for the application of the JNT methodology is that the matrix \mathbf{R} is *positive semi-definite*, in order to be able to carry out the Cholesky decomposition $\mathbf{R} = \mathbf{T} \cdot \mathbf{T}^T$ in the simulation procedure (algorithm 4.1).

According to the general definition presented in section 3.4.3, the matrix \mathbf{R} is positive semi-definite when for all $x \in \mathbb{R} \setminus \{0\}$, $x^T \mathbf{R} x > 0$, or equivalently, when all its eigenvalues are non-negative. A real symmetric matrix \mathbf{R} is positive semi-definite if and only if there exists a real nonsingular matrix \mathbf{T} such that $\mathbf{R} = \mathbf{T} \cdot \mathbf{T}^T$.

4.3.1 Problems with positive semi-definiteness

A sample correlation matrix, calculated directly from (adequate) data, is by definition positive semi-definite. However, there are cases that a non-positive semi-definite correlation matrix is constructed:

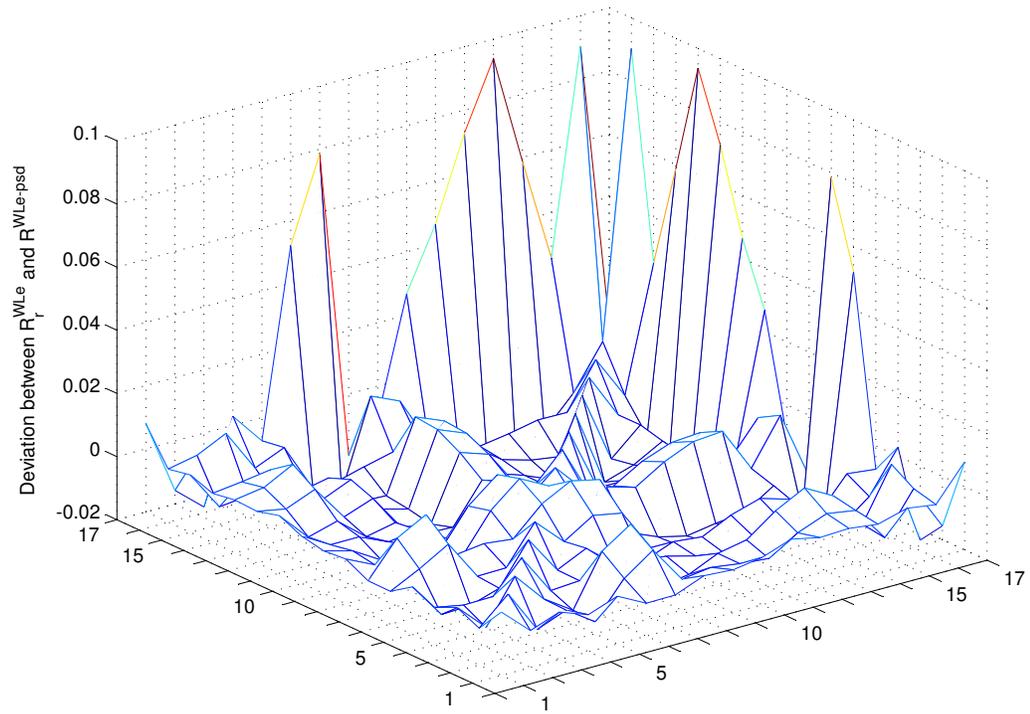


Figure 4.7: Deviation between the non-positive semi-definite rank correlation matrix $\mathbf{R}_r^{\text{WLe}}$ and the one obtained after the application of the repairing method, $\mathbf{R}^{\text{WLe-psd}}$.

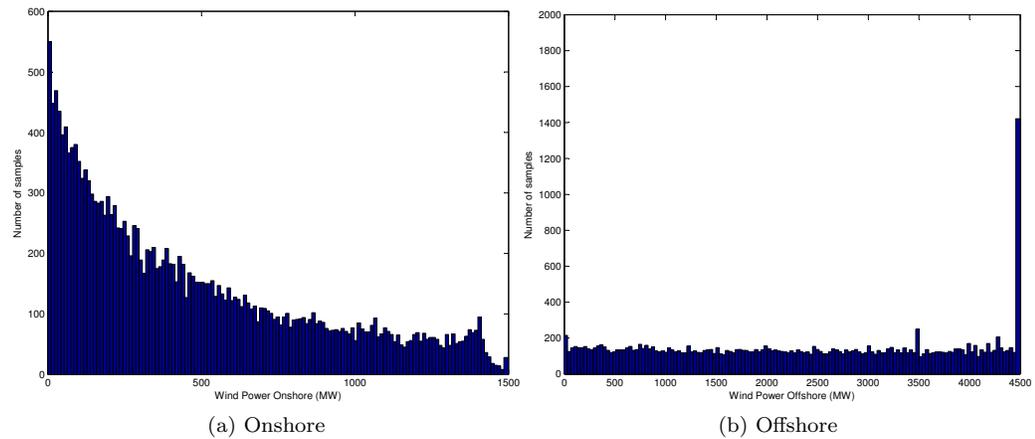


Figure 4.8: Onshore and offshore wind power distributions for the 6GW-integration scenario.

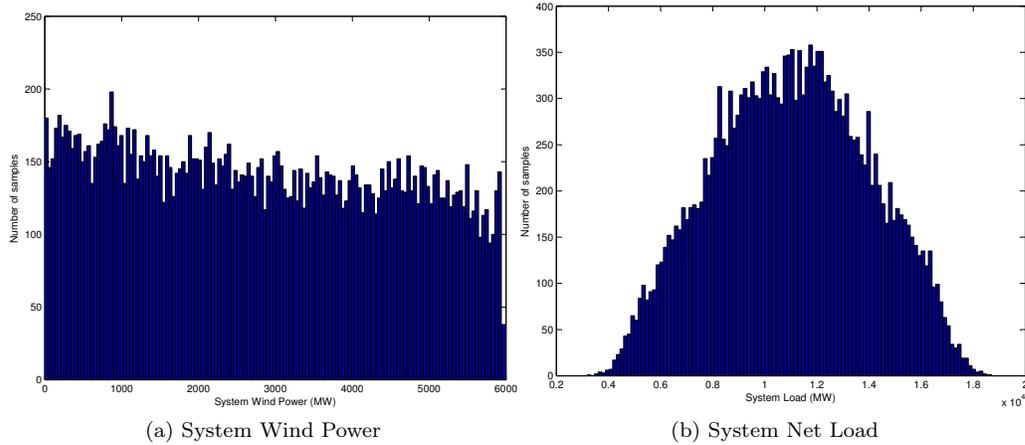


Figure 4.9: Total system wind power and system net load distributions for the 6GW-integration scenario.

1. *Data from different data systems*: the analysis presented above involved the extraction of a correlation matrix out of two different data sets. In practical cases, more data systems may be involved in the analysis, i.e. load measurements from different operators, different SG types etc. The non-uniformity of these data sets, due to different measurement frequencies, different measurement equipment and/or period of observations, may lead to a non-positive semi-definite correlation matrix.
2. *Transformation of rank correlation matrix*: as discussed, the matrix \mathbf{R} used in the simulation core of the JNT methodology is obtained by a transformation of the data rank correlation matrix \mathbf{R}_r , according to the relation between product moment and rank correlation for the normal distribution. This transformation however does not guarantee that the product matrix \mathbf{R} is positive semi-definite, even if the initial matrix \mathbf{R}_r is. In particular, the probability that a randomly chosen correlation matrix stays positive semi-definite after the transformation in equation 4.1 goes rapidly to zero with the matrix dimension [51].
3. *Data reliability*: it happens often that, due to measurement problems, data sets of reduced reliability are utilized. This low reliability may lead to a non-positive semi-definite correlation matrix.
4. *Experts data*: in cases when data are not available, we wish to rely on experts opinion for the definition of these missing values. In such cases, there is no guarantee that the resulting correlation matrix will be positive semi-definite.
5. *Paradigm shift*: statistical estimation techniques based on the analysis of historical data are intrinsically ill-suited to provide predictions of future quantities when a paradigm shift has taken place. Such a paradigm shift could be, for instance, the change in system load or weather conditions, especially in cases when our data

base consists of a relatively short time-period. In such cases, it is to some extent possible to make, based on the available data, forward-looking estimates for the respective r.v.; this procedure involves the definition of estimates of the volatility of the r.v. and also of the mutual dependencies. This procedure leads to a partial (or complete) alteration of the correlation matrix, which may result in non-positive semi-definiteness [75].

In all these cases, a non-positive semi-definite correlation matrix is obtained from the data analysis. In order to apply the JNT method, specific procedures should be employed in order to repair the violations of positive semi-definiteness.

4.3.2 Repairing methods

Repairing the violations of positive semi-definiteness comes down to the transformation of the initial non-positive semi-definite matrix to a positive semi-definite matrix, that is as close as possible to the original one. In particular, a method has to be used that:

1. is guaranteed to produce a positive-definite matrix;
2. does not require a pre-existing acceptable (positive-definite) matrix to begin with;
3. is fast to implement even for large matrices;
4. allows the determination of a feasible matrix that most closely approximates a target real symmetric (but non positive-definite) matrix in a well-defined and quantifiable sense.

In Appendix B, two main methods for repairing violations of positive semi-definiteness are presented, according to [75]; the *Hypersphere Decomposition*, which satisfies all above properties and a second, faster method, the *Spectral Decomposition*, which shares the first three properties but is not guaranteed to satisfy the fourth property. However, as mentioned in [75], empirical studies show that the results obtained using this second approach are extremely close, albeit not identical, to the ones obtained using the first technique.

4.4 Problems with JNT methodology

As presented in the previous section, violations of positive semi-definiteness can be repaired by the transformation of the initial matrix to another, positive semi-definite one. The deviation of the transformed matrix from the initial one depends on the divergence from positive definiteness. In the study case in section 4.2, this divergence is not high (one of the eigenvalues of the initial matrix equals to -0.0172). The deviation between the two matrices is presented in Fig. 4.7. We can see that just few of the entries of the initial matrix remain unchanged; the largest deviations are observed for the entries in the 17th row and column of the matrix, corresponding to the data that is obtained by experts. Hence, in order to proceed with the application of the method, we should change the values obtained by the elicitation of experts. This is however equivalent to distortion

of the expert opinion. In such cases, a method should be used that ensures that such distortions are minimal.

In analyses involving a large number of r.v. such problems are magnified, mainly due to the fact that the method requires the specification of all the entries of the correlation matrix. For example, in a problem involving 100 r.v., in total 4950 distinct correlations should be assessed. It is highly unlikely that in such problems the correlations obtained by different data systems (or experts) will form a positive semi-definite matrix. In other cases, we may want to leave some entries of the matrix unspecified, due to lack of data. Working with a partially-specified correlation matrix, leads to the so-called *matrix completion problem*: can a partially-specified matrix be extended to a positive semi-definite one? A basic matrix completion procedure is presented in [45], by taking the unspecified entries as close to zero as possible. This, however, corresponds to the assumption of independence between the respective inputs and may not lead to the best solution of this problem [51].

In general, when the uncertainty in the system model is increased, either due to high dimensionality or lack of adequate data, the deviations from positive semi-definiteness may be significant. In this case, we should use methods that permit the stochastic modeling of the system without requiring the specification of the correlation matrix. Two graphical methods are presented in the next two sections for the treatment of this problem: *dependence models trees and vines with copulas*. The main difference with the JNT methodology is that these methods do not require the full correlation matrix for the stochastic modeling of the system. Instead of using a multidimensional copula, the sampling routines derive directly from the bivariate dependencies. In this way, the system modeling can be performed even in cases of poor knowledge of the correlation matrix.

The dependence trees with copulas constitute a straightforward extension of the bi-dimensional dependence modeling using copulas. The limitations imposed by the concept of trees leads further to the definition of vines.

4.5 Dependence Trees with Copulas

4.5.1 General

In the previous chapter, the two-step algorithm 3.2 has been presented for the sampling of two correlated r.v. X_1 and X_2 . In particular, first the correlated uniform rank distributions U_1 and U_2 are simulated using the joining copula and then they are transformed into the desired marginals by the application of the inverse cdf transformation $F_{X_i}^{-1}(U_i)$. The main point here is that by sampling U_1 , we obtain not only the value u_1 , but also the conditional distribution of U_2 for the specific value u_1 , according to the functional relationship defined by the bivariate copula $C_{12|u_1}$. The conditional sampling of the second variable is then performed by the *independent* sampling of a uniform r.v. U_{r_2} and by the application of the transformation $C_{12|u_1}^{-1}(u_{r_2})$. Thus, the mechanism for the bivariate sampling is to sample the first r.v. and then propagate this value through the copula in order to sample the other r.v. conditionally. This idea may be directly extended to higher dimensions.

In Algorithm 4.2, the extension of this principle for the case of three dependent r.v. X_1, X_2, X_3 is presented, with marginals F_1, F_2 and F_3 , joined by the bivariate copulas C_{12} and C_{13} with rank correlations $\rho_{r_{12}}$ and $\rho_{r_{13}}$ respectively. As may be seen, starting from

Algorithm 4.2 Simulation algorithm 1 for the modeling of three correlated r.v.

1. Sample $U_{r_1}, U_{r_2}, U_{r_3}$. Denote realizations u_{r_1}, u_{r_2} and u_{r_3} .
 2. $u_1 = u_{r_1}$: the sampling of the rank distribution $U_1 = U_{r_1}$.
 3. Find $C_{12|u_1}$ and $C_{13|u_1}$. These are the conditional distributions of U_2, U_3 given u_1 .
 4. Calculate the rank distributions U_2 and U_3 : $u_2 = C_{12|u_1}^{-1}(u_{r_2})$ $u_3 = C_{13|u_1}^{-1}(u_{r_3})$.
 5. Transform the rank distributions according to the marginals: $x_1 = F_1^{-1}(u_1)$, $x_2 = F_2^{-1}(u_2)$, $x_3 = F_3^{-1}(u_3)$.
-

the independent sampling of three r.v. $U_{r_1}, U_{r_2}, U_{r_3}$, we obtain their respective correlated ranks U_1, U_2 and U_3 by sampling their joining copulas conditionally $u_2 = C_{12|u_1}^{-1}(u_{r_2})$ and $u_3 = C_{13|u_1}^{-1}(u_{r_3})$. Hence, for the application of this technique, the r.v. X_2 and X_3 should be *conditionally independent given X_1* . This is the basic restriction on the applicability of the method. Of course, the Algorithm 4.2 may also be applied as follows:

Algorithm 4.3 Simulation algorithm 2 for the modeling of three correlated r.v.

1. Sample $U_{r_1}, U_{r_2}, U_{r_3}$. Denote realizations u_{r_1}, u_{r_2} and u_{r_3} .
 2. $u_2 = u_{r_2}$: the sampling of the rank distribution $U_2 = U_{r_2}$.
 3. Find $C_{12|u_2}$.
 4. Calculate the rank distribution U_1 : $u_1 = C_{12|u_2}^{-1}(u_{r_1})$.
 5. Find $C_{13|u_1}$.
 6. Sample the rank distribution U_3 : $u_3 = C_{13|u_1}^{-1}(u_{r_3})$.
 7. Transform the rank distributions according to the marginals: $x_1 = F_1^{-1}(u_1)$, $x_2 = F_2^{-1}(u_2)$, $x_3 = F_3^{-1}(u_3)$.
-

The two realizations presented in Algorithms 4.2 and 4.3) show how the use of copulas may enable the propagation of random numbers through a specific dependence structure. This principle can be extended for the modeling of more than 3 r.v. This led to the method of *dependence trees with copulas* [24].

4.5.2 Definitions

According to this method, the dependence structure in the system is depicted as an undirected acyclic graph, i.e. a *tree*. The nodes of the tree are used to specify the marginal distributions, while the edges of the tree are used to specify bivariate dependencies. In particular, a univariate distribution F_i is assigned to each node of the tree and the bi-

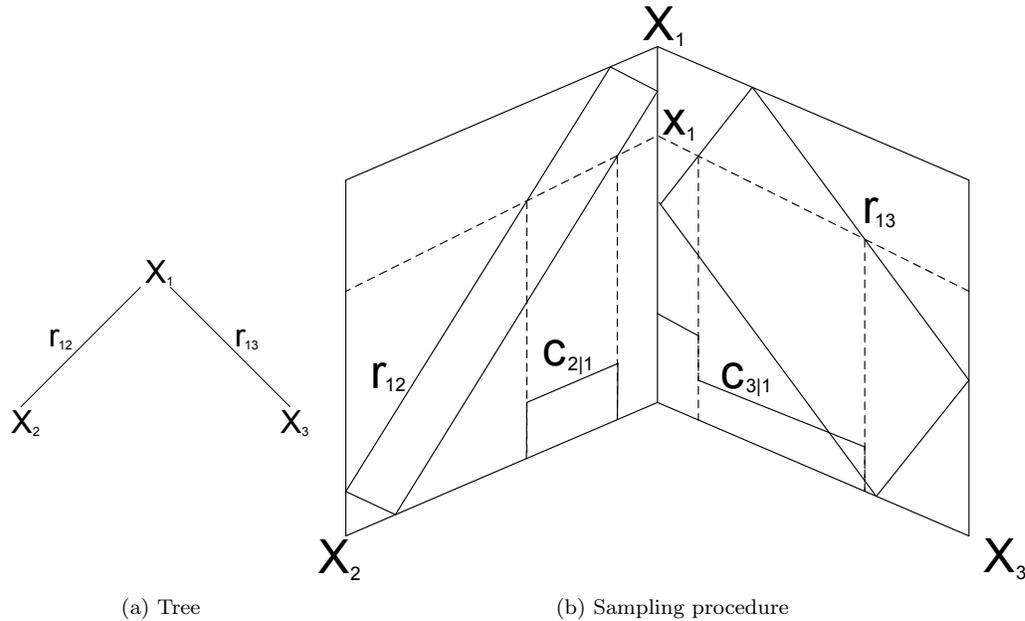


Figure 4.10: Sampling procedure for a tree of 3 r.v.

variate copula C_{ij} for X_i, X_j is assigned to each edge (i, j) . In Fig. 4.10, a tree on 3 r.v. is presented, together with the sampling procedure for the case that the r.v. are joined by a diagonal band copula. The sampling procedure corresponds to the realisation of Algorithm 4.2 For a more detailed definition of dependence trees with copulas, one may refer to [51].

By generalizing the algorithms presented above, the sampling of all types of trees is possible. Two basic restrictions should be fulfilled: no cycles should be present in the graph and all r.v. that share a common adjacent node should be conditionally independent (as discussed in the analysis of algorithm 4.2). Thus, non-adjacent variables on a path are conditionally independent given any set of variables separating them on the path. These restrictions correspond to the *Markov tree dependence* property of the trees. When this property is fulfilled, the joint probability density function of the r.v. can be obtained as a product of the one-dimensional marginal pdfs f_i and the bivariate copulas C_{ij} [51].

Summarizing, the copula-tree method represents high dimensional distributions by specifying the following elements:

1. continuous invertible marginal distributions assigned to the nodes of the tree;
2. a set of copulas assigned to the edges of the tree.

In Fig. 4.11 two graphs are presented. The graph on the left is a tree on 6 variables and the graph on the right is undirected graph with a cycle. For the Markov realization of this tree, variables 1 and 5 would be sampled independently. Variables 2 and 3 would then be sampled independently but conditional on the value of variable 1. Variable 6 would

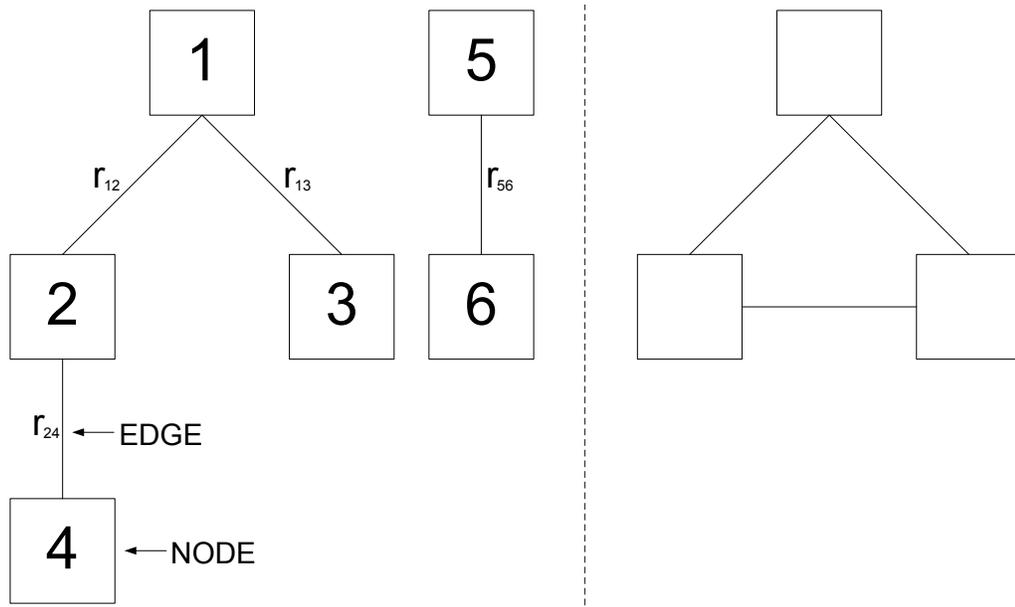


Figure 4.11: A tree with 6 nodes (left) and an undirected graph with a cycle (right). The notation r_{ij} refers to the rank correlation $\rho_{r_{ij}}$.

be sampled conditional on the value of variable 5. Finally, variable 4 would be sampled conditional on the value of variable 2. According to the Markov realization, variables 2 and 3 are conditionally independent given 1, variables 1 and 4 are conditionally independent given 2 and variables 3 and 4 are conditionally independent given 1,2.

These specifications of conditional independence in the copula-trees pose the main restrictions on the applicability of the method; in order to overcome these restrictions, the dependence vines concept should be employed.

4.6 Dependence Vines with Copulas

4.6.1 General

The development of dependence vines came from the restrictions of the Markov realization of the copula trees. Coming back to the tree in Fig. 4.11 with 6 r.v., we see that only the correlations $\rho_{r_{12}}$, $\rho_{r_{13}}$, $\rho_{r_{24}}$ and $\rho_{r_{56}}$ are specified. The other correlations are determined by the realization and the choice of the copula. The rank correlation matrix obtained for the 4-r.v. tree of the first 4 r.v. in Fig. 4.11, assuming the normal copula and the correlations $\rho_{r_{12}} = 0.7$, $\rho_{r_{13}} = 0.8$, $\rho_{r_{24}} = 0.6$, the specific tree realization is the

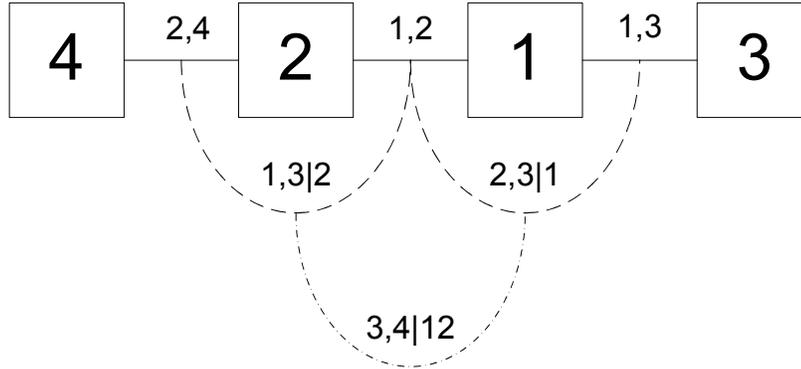


Figure 4.12: A D-vine on 4 variables.

following[‡]:

$$\mathbf{R}_r = \begin{pmatrix} r.v. & 4 & 2 & 1 & 3 \\ 4 & 1 & 0.6 & 0.4265 & 0.346 \\ 2 & 0.6 & 1 & 0.7 & 0.565 \\ 1 & 0.4265 & 0.7 & 1 & 0.8 \\ 3 & 0.346 & 0.565 & 0.8 & 1 \end{pmatrix}$$

Due to the tree realization, specific values for the correlations $\rho_{r14} = 0.4265$, $\rho_{r43} = 0.346$, $\rho_{r23} = 0.565$ are obtained. This is a basic restriction in the applicability of trees; in particular, in a tree on n variables, only $n - 1$ correlations out of $\binom{n}{2}$ can be specified. It would be possible however to *enrich* the copula tree with additional information on the conditional dependence between 2 and 3 given 1, between 1 and 4 given 2 and between 3 and 4 given 1 and 2. This is the basic idea behind the graphical models *vines* introduced in [24].

4.6.2 Definitions

A vine on n variables is a nested set of trees T_j where the edges of the j^{th} tree become the nodes of the $(j - 1)^{\text{st}}$ tree for $j = 1, \dots, n$. A *regular* vine on n variables is a vine in which two edges in tree j are joined by an edge in tree $j - 1$ only if these edges share a common node. A *D-vine* is a special case of a regular vine in which each node in T_1 has degree of at most 2[§], hence each node in the first tree has at most two neighbors. In Fig. 4.12, a D-vine on 4 variables is presented, corresponding to the tree on the four first variables in Fig. 4.11.

Each edge in the regular vine may be associated with a conditional rank correlation and a copula, and each node with a marginal distribution. All assignments of rank correlations to edges of a vine are consistent and each one of these correlations may be realized by a copula. Based on the bivariate and conditional bivariate distributions, the joint distribution may be constructed. Distributions specified by (conditional) rank

[‡]The sequence of the r.v. in the correlation matrix is 4-2-1-3.

[§]The degree of a node of a tree is the number of edges attached to the specific node.

correlations on a vine with copula can be sampled easily. Details on the construction of the joint distribution and the sampling algorithms may be found in [51]. These sampling algorithms provide a valuable alternative for multivariate stochastic modeling. For this, one may use the dedicated software package UNICORN [32].

When we revert to the example on the definition of the rank correlations for the 4-r.v. tree in Fig. 4.11, we may enrich the copula tree by specifying the conditional correlations $\rho_{r41|2} = 0.5$, $\rho_{r23|1} = 0.9$ and $\rho_{r43|12} = 0.4$. In this case, the following rank correlation matrix is obtained:

$$\mathbf{R}_r = \left(\begin{array}{c|cccc} r.v. & 4 & 2 & 1 & 3 \\ \hline 4 & 1 & 0.6 & 0.7102 & 0.7166 \\ 2 & 0.6 & 1 & 0.7 & 0.9467 \\ 1 & 0.7102 & 0.7 & 1 & 0.8 \\ 3 & 0.7166 & 0.9467 & 0.8 & 1 \end{array} \right)$$

The vine realization permits all sets of values for the correlations ρ_{r14} , ρ_{r43} , ρ_{r23} . All different conditional rank correlations $\rho_{r41|2}$, $\rho_{r23|1}$ and $\rho_{r43|12}$ can be assigned, without problems with the consistency of the model. In this way, always a positive definite matrix \mathbf{R}_r is obtained. Vines offer an alternative method for the definition of consistent correlation matrices from partial specifications and an alternative repairing method for problems with positive semi-definiteness [51].

These graphical methods offer alternative and always consistent algorithms for the system stochastic modeling. Instead of requiring the complete correlation matrix, as with the JNT methodology, the stochastic model is defined based on the specification of conditional correlations. However, obtaining these conditional correlations from data or experts is a cumbersome procedure. A way out of this problem is by assuming a normal copula; in this case, the conditional correlation equals the partial correlation, which is much easier to measure by data [51]. This assumption is actually the same as what is used in JNT methodology. For the elicitation procedures for obtaining these correlations from experts one may refer to [60].

4.7 Conclusions

The cornerstone in the multivariate uncertainty analysis is the splitting of the stochastic model in the modeling of the marginal distributions and the modeling of the dependence structure. Modeling the dependence structure is the most cumbersome part; a multidimensional dependence model has to be employed, based on the available information. When adequate information is available, one may use a multidimensional copula (JNT methodology) for the system modeling. In this case, the system model is build on the system correlation matrix. The main problems with this method derive from the requirement that this matrix should be positive semi-definite. Although by definition every correlation matrix fulfills this property, in most cases correlation matrices are obtained by noisy procedures that may lead to non-consistency. In this case, we should repair the problems with positive semi-definiteness and convert the matrix into a consistent one. When the deviation from the original matrix is small, this procedure may provide a good answer to the problem. When the dimension and the uncertainty of the model increases,

this deviation may be high. In this case, we need methods that may help us to model, based on the available information only. These methods are the graphical models trees and vines. In this case, the system model is based on the specification of conditional correlations and the obtained model is always consistent. Obtaining these conditional correlations however may prove a cumbersome procedure.

An alternative approach for the modeling of high-dimensional distributions is presented in the next chapter. In particular, we try to simplify the multidimensional dependence model by a specific risk-averse model-reduction technique. In this case, instead of trying to build the dependence model between all the r.v., we divide the set of r.v. in groups and define a reduced multidimensional dependence model of these groups. These groups are called *stochastic plants*.

Chapter 5

Stochastic Plants

In the previous chapter, the basic methodologies for the sampling of the joint distribution of a vector of correlated r.v. were presented. The sampling routines are based on the splitting of the modeling of marginal distributions from the dependence modeling and the representation of the dependence structure by a multidimensional dependence model using multivariate copulas (JNT methodology) or appropriate graphical models (trees/vines). For the construction of these models, overall or partial information on the bivariate dependencies is needed. For models involving a large number of r.v., the difficulty in representing uncertainty is increased, either due to a lack of data or model complexity. In this chapter we discuss alternative methods for dealing with the problem of high-dimensionality. In particular, model reduction techniques are proposed, i.e. simplifying approximations of parts of the stochastic model. We present the theory for the derivation of such approximations that are consistent with two prerequisites: they are *simpler* and they *bound the uncertainty of the SSM*. These model approximations are called *stochastic plants (SP's)*.

In section 5.1, we present a basic observation that is used for the formulation of these model approximations, namely the existence of clusters of system inputs that co-fluctuate in time. Perfect positive dependence may be used for the modeling of this type of stochasticity. In section 5.2, the theory and properties of such models are presented (*Stochastic Bounds Methodology*). It is shown that these stochastic models correspond to the Fréchet-Hoeffding bounds C_L and C_U and that they offer the natural bounds to the aggregate stochasticity of the cluster. Perfect positive dependence (comonotonicity - upper bound C_U) corresponds to the highest-risk case, while perfect negative dependence (countermonotonicity - lower bound C_L), or under some assumptions independence, corresponds to the lowest-risk case. It is concluded that comonotonicity offers a prudent but yet accurate approach for the modeling of such clusters of r.v. This basic idea is used in section 5.3 for the definition of the stochastic plant concept. In section 5.4, the properties of this concept for the case of a wind stochastic plant are investigated, by looking at the sensitivity of the aggregate power output to the stochastic dependence structure and the marginal wind speed distributions. In section 5.5, the conclusions for the use of stochastic plants in power system stochastic modeling are presented.

5.1 The co-fluctuation of stochastic inputs

In Fig. 5.1, measurement data of the power output from a windpark in The Netherlands are presented*. The WTGs are grid-connected by means of four feeders and a single distribution transformer at the point of common coupling. The measurements consist of 5-min average active power values, supplied by each feeder to the transformer. We can see that the power output at the feeders co-fluctuate in time, due to the uniform wind activity in the area of the wind park. In Fig. 5.1b, the scatter diagram for the measurements in two of the feeders is presented. The co-fluctuation corresponds to a strong positive dependence.

This co-fluctuation is explicitly observed within clusters of stochastic inputs in the power system. Two types of such clusters may be identified:

- *Load clusters* comprising loads of the same type. They are positively correlated due to their subjection to a common customer behavior.
- *Stochastic generation clusters* comprising SG units of the same type, situated in a relatively small geographic area. They are positively correlated due to mutual coupling to a uniform prime-energy mover.

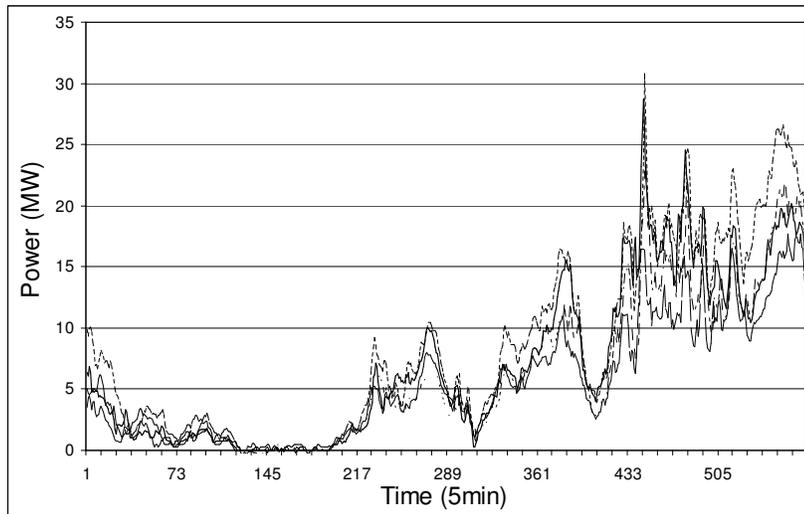
Hence, the r.v. that belong to such a cluster are strongly positively dependent. Therefore, models of perfect positive dependence may be used to approximate the dependence structure within the cluster. In the following, we present the theory concerning the derivation of these models and investigate their properties: they offer the natural bounds to the stochasticity of the cluster, corresponding to the worst-case scenario for the cluster's aggregate power output, i.e. the case of maximum variability.

5.2 Stochastic Bounds Methodology (SBM)

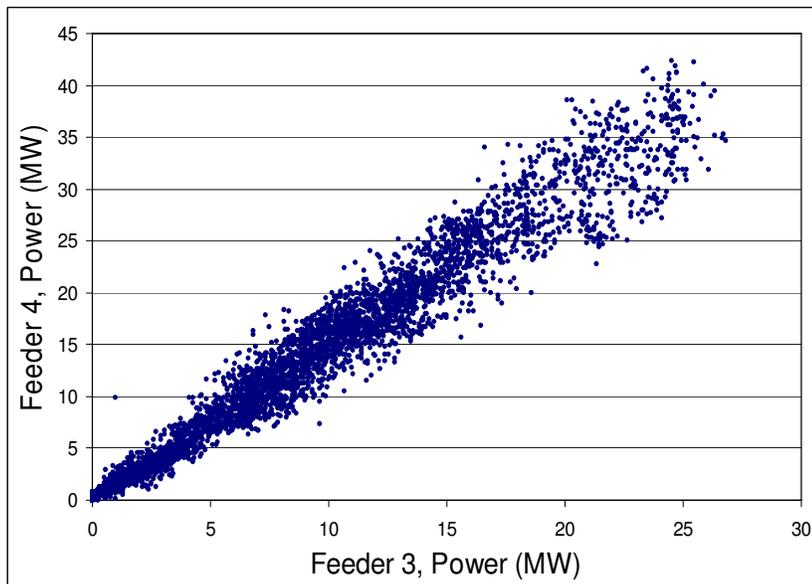
In section 3.7.2, the basic bounds inequality for copula functions has been presented (equation 3.8), which states that the Fréchet-Hoeffding bounds C_L and C_U offer the natural bounds to all possible copulas. According to proposition 3.7.2, these two copulas correspond to the cases of perfect negative and positive dependence respectively. Thus, the copula inequality (3.8) may be interpreted in the following way: "all possible dependence structures between two r.v. are bounded between the cases of perfect negative (lower bound) and perfect positive dependence (upper bound)". An alternative terminology for the copulas C_U and C_L , broadly used in risk management literature, is *comonotonicity* and *countermonotonicity* respectively [38, 35, 36].

In order to obtain an insight to this inequality, one can refer to the example in section 3.4, where the case of correlated normals was presented. Different distributions of the sum of r.v. around the same central point are obtained by varying the correlation between -1 and 1 (Fig. 3.5). For maximum negative correlation ($\rho = -1$ - countermonotonicity), the variance of the sum is minimal, while for maximum positive correlation ($\rho = 1$ - comonotonicity) the respective variance is maximal. In the following, we present the generalization of the above results for all types of r.v.

*The windpark is located in Zeewolde, The Netherlands. The data are provided by NUON.



(a) Power output time-series of four feeders - 24 hours



(b) Scatter diagram - 2 weeks

Figure 5.1: Time-series and scatter diagram of 5-minute measurements of active power output of a windpark.

5.2.1 Upper bound: comonotonicity

Comonotonicity is a dependence concept that represents a form of extreme positive dependence between r.v. Comonotonic r.v. always vary in the same way, meaning that if one increases, then all the others increase too. Such dependence can occur if all variables are increasing functions of the same underlying random factor. If the r.v. Y_1, Y_2, \dots, Y_M are comonotonic, they can be expressed as [36]:

$$\text{Comonotonicity} : Y_m = F_{Y_m}^{-1}(U), \quad m = 1 \dots M, \quad (5.1)$$

where U is a r.v. that is uniformly distributed on the unit interval. This functional relationship may be observed for the bi-dimensional case in Fig. 3.9, where the copula C_U is presented. In this case, the same random generator U is used for the system modeling. Comonotonicity is a more general concept than the perfect (linear) correlation, as it includes all cases of perfect positive non-linear dependence.

By $(Y_1^C, Y_2^C, \dots, Y_M^C)$ we denote the comonotonic version of the random vector (Y_1, Y_2, \dots, Y_M) , i.e. a random vector with the same marginal distributions, but with a comonotonic dependence structure between the Y_m^C 's. The sums $\sum_m Y_m$ and $\sum_m Y_m^C$ have the same expected value, but different overall probability distributions. As shown in [36], the comonotonic case provides always the extreme distribution for this sum.

A way to characterize the effect of dependence is via the concept of *stochastic orders*. It can be shown that for all convex functions g , the following relation holds [62]:

$$E \left[g \left(\sum_{m=1}^M Y_m \right) \right] \leq E \left[g \left(\sum_{m=1}^M Y_m^C \right) \right] \quad (5.2)$$

where $E[\cdot]$ denotes the expectation operator. This is equivalent to saying that 'the sum $\sum_m Y_m$ is smaller than $\sum_m Y_m^C$, with respect to the convex order on sets of r.v.'. Stochastic orderings, such as the one implied by equation (5.2), provide sophisticated means of comparing random quantities in terms of their variability. To see this, consider the function $g(x) = x^2$. Given that the sums have the same expectations, equation (5.2) immediately results into an ordering of the respective variances:

$$\text{Var} \left(\sum_{m=1}^M Y_m \right) \leq \text{Var} \left(\sum_{m=1}^M Y_m^C \right) \quad (5.3)$$

Thus, in terms of risk assessment, comonotonicity provides the higher risk case corresponding to a worst-case scenario (extreme distribution of the sum of the r.v.). By allowing any convex function in equation (5.2), a more comprehensive comparison of the distributions is achieved, than by just considering the variance as in equation (5.3). Just as comonotonicity provides a more general dependence concept than perfect correlation, the resulting convex order of equation (5.2) provides a more general comparison of risk than the variance.

5.2.2 Lower bounds: countermonotonicity - independence

Although comonotonicity always corresponds of the upper bound of the system stochasticity, the lower bound is more difficult to obtain, since the Fréchet-Hoeffding lower bound

C_L may be applied only in the case of two r.v. However, under the assumption of a weak positive dependence between a number of r.v., a lower bound can be obtained for the multidimensional case as well.

In the case of two r.v., the lower bound corresponds to the case of perfect negative correlation between them, i.e. the case of countermonotonicity [38]. Counter-monotonic pairs of r.v. constitute an extreme form of negative dependence, whereby the one variable is a decreasing function of the other. In the case that the r.v. Y_1 and Y_2 are counter-monotonic, they can be expressed as:

$$\text{Countermonotonicity : } Y_1 = F_{Y_1}^{-1}(U), Y_2 = F_{Y_2}^{-1}(1 - U). \quad (5.4)$$

By (Y_1^{CM}, Y_2^{CM}) we denote the countermonotonic version of the bidimensional random vector (Y_1, Y_2) . It can be shown that for all convex functions g , the following convex order holds [38]:

$$E [g (Y_1^{CM} + Y_2^{CM})] \leq E [g (Y_1 + Y_2)]. \quad (5.5)$$

Hence, countermonotonicity offers always the lower stochastic bound to the aggregate stochasticity of two r.v. However, in the case of more than two r.v., the application of this concept is not possible, since by definition more than two r.v. cannot be mutually perfectly negatively dependent[†].

However, in most cases, stochastic generators present non-negative dependence. Characteristic case is the correlations obtained for the wind speed measurements for The Netherlands, presented in section 4.2. As discussed, the minimum value in the correlation matrix is 0.47. In the case that it is known that a weak positive dependence exists between a number of r.v., the lower bound shifts to the case of independence. In particular, for a vector of r.v. that are *positive cumulative dependent*[‡], the case of independence offers the lower stochastic bound to the aggregate stochasticity of the vector. When the r.v. Y_1, Y_2, \dots, Y_M are independent, they can be expressed as:

$$\text{Independence : } Y_m = F_{Y_m}^{-1}(U_m), m = 1 \dots M. \quad (5.6)$$

As may be seen, different random generators U_m are used for each r.v. In this case it can be shown that for all convex functions g , the following convex order holds [38]:

$$E \left[g \left(\sum_{m=1}^M Y_m^I \right) \right] \leq E \left[g \left(\sum_{m=1}^M Y_m \right) \right] \quad (5.7)$$

where $(Y_1^I, Y_2^I, \dots, Y_M^I)$ is the independent version of the random vector (Y_1, Y_2, \dots, Y_M) . Thus, in terms of risk assessment, in the presence of a weak positive dependence between the r.v. belonging to a vector of r.v., the case of independence provides the lower-risk case, corresponding to a best-case scenario (distribution of the sum of the r.v. with minimum variability).

[†]Let Y_1 and Y_2 two perfectly negatively correlated r.v. If Y_3 is perfectly negatively correlated to Y_1 , then by definition it is perfectly positively correlated to Y_2 , since Y_2 is also perfectly negatively correlated to Y_1 .

[‡]As defined by Denuit et al. [34], positive cumulative dependence is a weak form of positive dependence. The intuitive meaning is that, under this assumption, the probability that these r.v. assume 'large' values together is greater than if they were independent.

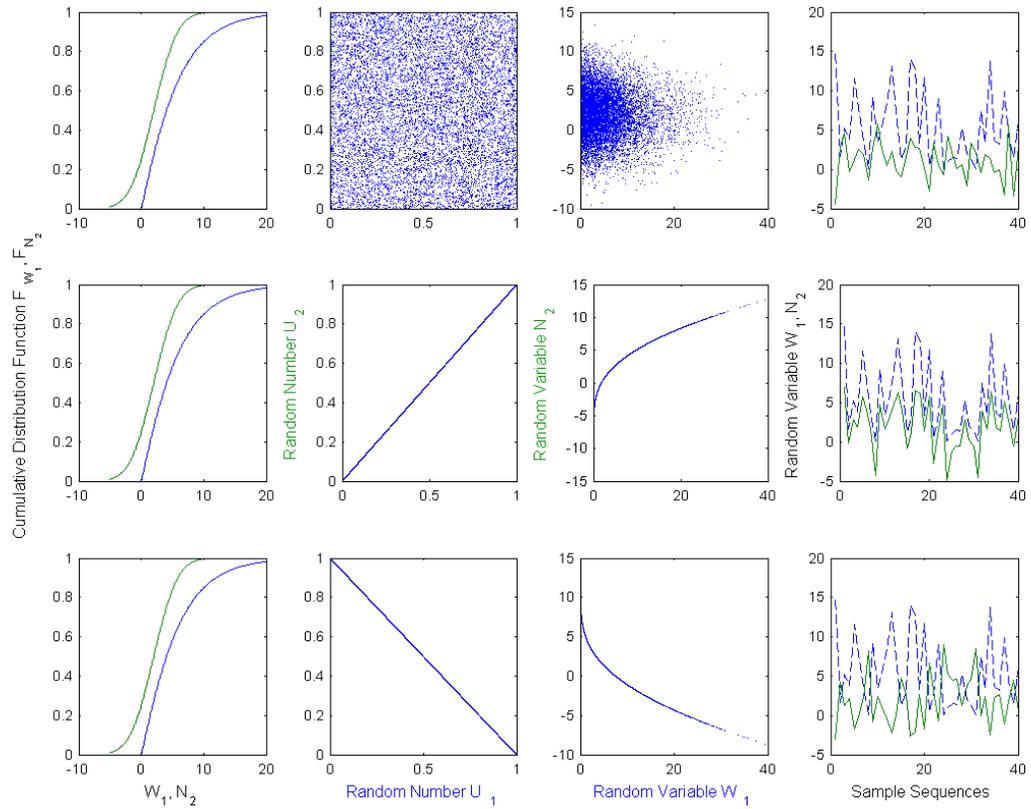


Figure 5.2: Independent/comonotonic/countermonotonic sampling and generated sequences.

In Fig. 5.2, the results of the modeling of a Weibull W_1 and a Normal N_2 r.v. according to these dependence concepts is presented. In the upper row, the case of independence is presented, in the second row the case of comonotonicity and in the third row the case of countermonotonicity. In the second column, the respective copulas i.e. the scatter diagrams for the random generators U_1 and U_2 are presented, while in the third column, the scatter diagrams for the r.v. W_1 and N_2 are shown, generated according to the equations (5.6), (5.1) and (5.4). In the fourth column, the generated sample sequences are presented. We can see that the comonotonic sequences follow the same variations, while the countermonotonic ones vary in an opposite fashion.

For power system modeling, different dependence scenarios correspond to different operating conditions. Comonotonicity corresponds to operation where the stochastic power injections co-fluctuate; independence corresponds to operation where the stochastic power injections are combined randomly, while in the case of countermonotonicity, the injections fluctuate in an opposite fashion. As presented in the example in section 2.10, each dependence scenario corresponds to a different ‘matching’ of the system inputs, leading to a unique distribution of their sum. By considering different dependence structures between the system inputs, we obtain different distributions of their sum around the same mean. The extreme dependence concepts correspond to the extreme forms of this probability distribution.

As comonotonicity produces sums of r.v. with the highest risk, it is concluded that in a prudent approach, when r.v. are positively dependent but information about the dependence structure is unavailable, comonotonicity should be used rather than independence, as the latter concept would yield a substantial understatement of the aggregate volatility.

In the following section, we show how the stochastic bounds may be used to reduce the high-dimensionality of power system stochastic modeling. The main idea is to approximate clusters of system inputs by their upper stochastic bounds and use these bounds in the stochastic model.

5.3 Stochastic plants

As discussed in section 5.1, clusters of positively dependent inputs are present in the power system stochastic modeling. The dependence structure within a cluster can be approximated by comonotonicity; according to the SBM, this will correspond to the highest-risk case for the aggregate stochasticity of the cluster. This leads to a model reduction, since the system dependence structure is reduced to the definition of the interdependencies between the respective clusters, instead of the complete dependence structure between all variables. In the following, we present how this model reduction is performed.

5.3.1 Stochastic model reduction

In the general case, the system stochastic model can be subdivided into K SG and M load clusters of positively correlated inputs. Each SG cluster k and load cluster m contains a total of g_k SG units and l_m loads respectively, as shown in Fig. 5.3. Each load r.v. is denoted as L_{ml} and each SG r.v. as G_{kg} . Thus, according to equation (5.1), the

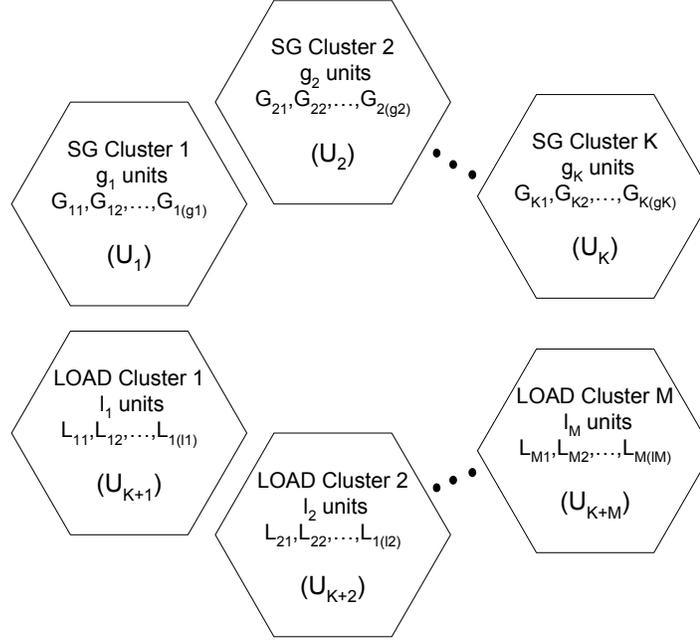


Figure 5.3: Model reduction by the application of the stochastic plants concept. A single random generator U is used for the modeling of each stochastic plant.

dependence structure in each cluster is modeled as:

$$SG \text{ cluster } k : G_{kg} = F_{G_{kg}}^{-1}(U_k), \quad g = 1, \dots, g_k \quad (5.8)$$

$$Load \text{ cluster } m : L_{ml} = F_{L_{ml}}^{-1}(U_{K+m}), \quad l = 1, \dots, l_m \quad (5.9)$$

for $k = 1 \dots K$ and $m = 1 \dots M$.

This comonotonic approximation of the dependence structure of a cluster is defined as a *Stochastic Plant (SP)*. According to the SBM, the obtained aggregate distributions $\Sigma_g G_{gk}(U_k)$ and $\Sigma_m L_{ml}(U_m)$ correspond to the extreme power output distribution of the cluster.

This is a worst-case approach and offers a tradeoff between modeling simplicity and accuracy. In particular, the closer the cluster dependence structure is to perfect positive dependence, the less the aggregate distribution deviates from reality. The choice of clusters is a decision of the designer; however, when the dependence structure is far from perfect positive correlation, such an approximation will lead to conservative modeling results. The general applicability of the method permits the definition of clusters that can comprise different types of generators (loads), i.e. different types of marginal distributions. Thus, the stochastic plant concept provides a pragmatic solution for the system modeling when no information concerning the dependence structure is available, or when it is too cumbersome to work with.

The application of the stochastic plants leads to a model reduction. In particular, instead of defining the dependence structure between all the system r.v., the system

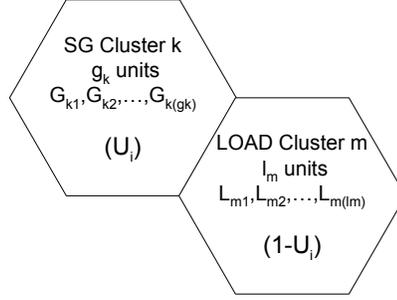


Figure 5.4: Clustering of stochastic plants comprising stochastic generators and loads.

stochastic model is decomposed into the definition of the dependence structure between the comonotonic clusters. Hence, a problem involving a total number of $\sum_{k=1}^K (g_k) + \sum_{m=1}^M (l_m)$ r.v. (Fig. 5.3) is decomposed into the modeling of the dependencies between $K + M$ stochastic plants. Since for each stochastic plant one random generator U is used, the techniques for the multidimensional dependence modeling presented in the previous chapter may be applied directly.

5.3.2 Stochastic plants comprising SG units and loads

A special case is when a stochastic plant comprises both stochastic generators as well as loads. A worst-case approach in this case corresponds to a dependence scenario with a combination of high (low) load and low (high) stochastic generation. This is equivalent to a re-clustering (grouping) of generation and load clusters. Note that the power output of the combined cluster corresponds to the difference (rather than the sum) of generation and load. Thus, if the upper bound corresponds to comonotonicity between generation and the negative of the load, then generation and load must be countermonotonic. Hence, the sampling of the upper bound for the group is obtained by equation 5.4:

$$SG \text{ component } k : G_{kg} = F_{G_{kg}}^{-1}(U_i), \quad g = 1, \dots, g_k \quad (5.10)$$

$$Load \text{ component } m : L_{ml} = F_{L_{ml}}^{-1}(1 - U_i), \quad l = 1, \dots, l_m \quad (5.11)$$

In Fig. 5.4, this clustering is presented. A single random generator U_i is used for the modeling of the cluster.

Thus, the stochastic plants are model approximations that permit the reduction of the stochastic system model to the definition of the mutual stochasticity between clusters of system inputs. As mentioned, the closer the cluster dependence structure is to perfect positive dependence, the better these approximations represent reality. In the following section we investigate the accuracy of the approximations for the case of a stochastic plant comprising wind turbine generators (WTGs).

5.4 Wind stochastic plant: example

We consider a stochastic plant comprising 50 WTGs situated in distinct sites in the system (*wind stochastic plant - WSP*). Given the set of Weibull marginal wind speed distributions, we investigate the impact of the stochastic dependence structure on the power output of the WSP.

The wind speed distributions describing the wind behavior at the location of each WTG are generated randomly based on the general characteristics of the wind regime in The Netherlands. This is achieved by the generation of the Weibull parameters β and η for the distributions at the 50 generation sites as random numbers drawn in a respective interval: $\beta \in [1.6, 2.8]$ and $\eta \in [6, 10]$. These intervals are obtained from the wind speed data analysis presented in the European Wind Atlas [92], and characterize the wind regime in The Netherlands. Using this method, 50 different Weibull distributions are generated randomly, corresponding to the general wind characteristics in the greater geographic area. A pitch-controlled WTG of 1MW nominal power is connected to each site, with cut-in, nominal and cut-out wind speeds of 3.5, 14 and 25 m/s respectively.

We choose these random wind speed marginals in order to investigate the impact of the dependence structure on the aggregate power output of the WSP. We divide the analysis in two parts:

1. *Part I*: we investigate the properties of the aggregate power output distributions of the WSP for different dependence structures, varying from independence to comonotonicity, for the same set of randomly generated wind speed marginal distributions. This analysis shows the impact of the different dependence structures on the type of the aggregate power output distribution of the WSP.
2. *Part II*: We repeat the simulation of part I 10000 times, each time for a different set of randomly generated Weibull marginal distributions. We investigate the impact of the different marginal distributions on the properties of the aggregate power output distribution of the WSP. This analysis shows the sensitivity of the aggregate power output distribution of the WSP to the marginal distributions.

5.4.1 Part I: Impact of the dependence structure

For this study case, the aggregate power output distribution of the WSP is investigated for different stochastic dependence scenarios. For each scenario, all mutual correlations between the r.v are considered to be equal to each other. In total 12 scenarios were obtained by increasing the correlations from zero to 1, with a step of 0.1, including the case of $\rho = 0.95$. In this way, consecutive dependence structures are obtained, from the lower stochastic bound (independence) to the upper one (comonotonicity).

In Appendix C.2, the Weibull parameters β and η for the wind speed distributions at the 50 generation sites are presented, generated by the random process presented above. The JNT methodology has been used for the simulation of the WSP and 10000 samples were obtained.

In Fig. 5.5, the distributions of the aggregate power output from the WSP for the different dependence scenarios are presented. In Table 5.1, the respective mean values and standard deviations are shown. We can see that, for all dependence scenarios, the power

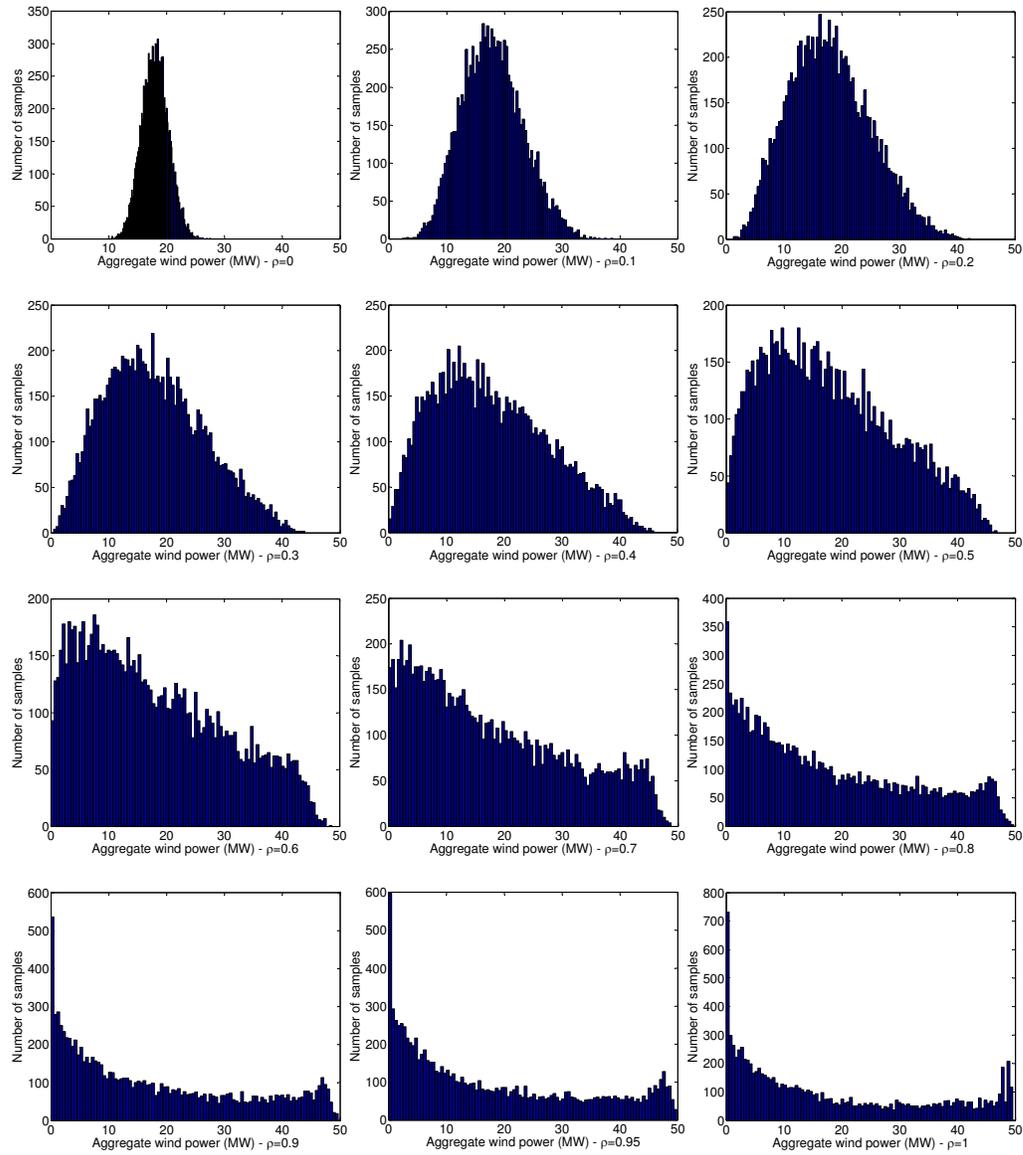


Figure 5.5: Wind stochastic plant aggregate power output distributions.

Table 5.1: Wind stochastic plant aggregate power output mean values, standard deviations and 5%-95% quantiles.

ρ	0	0.1	0.2	0.3	0.4	0.5
Mean (MW)	18.48	18.42	18.42	18.36	18.47	18.53
St.Dev. (MW)	2.44	5.30	7.14	8.61	9.88	11.10
q^5	14.58	10.10	7.41	5.36	4.09	2.85
q^{95}	22.54	27.54	30.91	33.44	36.61	38.52
ρ	0.6	0.7	0.8	0.9	0.95	1
Mean (MW)	18.29	18.77	18.42	18.38	18.35	18.33
St.Dev. (MW)	12.06	13.31	14.26	15.22	15.54	16.18
q^5	1.90	1.33	0.78	0.43	0.27	0.18
q^{95}	40.29	42.54	44.20	45.79	46.80	47.56

output distributions have the same mean. This is an expected result since, as discussed in section 3.1, the mean value of the sum is determined by the marginal distributions and is not affected by the dependence structure. Thus, the different dependence structures yield different distributions around the same central point.

In particular, the transition from independence to comonotonicity results in more extreme forms of the aggregate power distributions. The standard deviation of these distributions is increasing with ρ , ranging from a minimum of 2.44MW in the case of independence ($\rho = 0$) to a maximum of 16.18MW in the case of comonotonicity ($\rho = 1$), as may be seen in Table 5.1. The 5% and 95% quantiles (q^5 and q^{95} respectively) are also presented in the same table; as may be seen, q^5 shifts from 14.58MW in the case of independence to 0.18MW for comonotonicity, while q^{95} shifts from 22.54MW to 47.56MW respectively. Hence, the more the stochastic dependence structure approximates perfect correlation, the more extreme the aggregate distribution becomes.

The aggregate power output in the case of independence approximates a normal distribution; this is due to the applicability of the central limit theorem as discussed in section 3.2.2. In this case, the variability of the power output is minimal, in accordance to the SBM theory: independence yields the lower stochastic bound, i.e. the lowest risk case, the case of minimum variability of the aggregate power output. Hence, independence corresponds to minimum variability of the stochastic power injections to the system and a 'best-case scenario' for the system modeling, leading to a severe underestimation of the system risk.

On the other hand, comonotonicity ($\rho = 1$) corresponds to the highest risk case, i.e. the case when the system stochastic power injections present the highest variability. As may be seen from the distributions in Fig. 5.5, when the stochastic inputs are highly correlated, the assumption of comonotonicity actually yields a good approximation, ensuring however safety, since it corresponds to the worst-case scenario. As may be seen from Fig. 5.5, starting from $\rho = 0.8$, the approximation offers adequate accuracy.

5.4.2 Part II: Impact of the marginal distributions

The results of part I correspond to one specific set of marginals. In order to investigate the sensitivity of the obtained results to the choice of the marginal distributions,

Table 5.2: Mean values and standard deviations of the distributions in Fig. 5.6a.

ρ	0	0.1	0.2	0.3	0.4	0.5
Mean (MW)	18.33	18.33	18.34	18.34	18.34	18.34
St.Dev. (MW)	0.03	0.08	0.11	0.12	0.14	0.16
ρ	0.6	0.7	0.8	0.9	1	
Mean (MW)	18.34	18.34	18.34	18.33	18.34	
St.Dev. (MW)	0.17	0.19	0.21	0.22	0.24	

Table 5.3: Mean values and standard deviations of the distributions in Fig. 5.6b.

ρ	0	0.1	0.2	0.3	0.4	0.5
Mean (MW)	2.44	5.42	7.29	8.78	10.08	11.27
St.Dev. (MW)	0.02	0.05	0.06	0.07	0.08	0.09
ρ	0.6	0.7	0.8	0.9	1	
Mean (MW)	12.37	13.42	14.44	15.46	16.47	
St.Dev. (MW)	0.09	0.10	0.10	0.11	0.11	

the experiment is repeated for 5000 cases. In each case, a different set of 50 randomly generated Weibull distributions is used for the calculation of the aggregate power output distribution for 11 dependence scenarios corresponding to the transition from independence to comonotonicity. For each scenario, 5000 samples were obtained using the JNT methodology, yielding a total of $5000 \text{ cases} \times 11 \text{ dependence scenarios} \times 5000 \text{ samples}$ for the system analysis.

For each of the 5000 cases and the 11 dependence scenarios, the mean value and standard deviation of the WSP power output is computed from the 5000 samples. This yields a total of 5000 values for each of the two statistics for each dependence scenario. In Fig. 5.6, the distributions of the two statistics over the 5000 values are presented, for each dependence scenario.

In Fig. 5.6a, 11 distributions for the mean value are presented, each for a different dependence scenario. In Table 5.2, the mean values and standard deviations of the distributions in Fig. 5.6a are presented. The distributions in Fig. 5.6a have the same mean of 18.33MW and a low standard deviation that increases during the transition from independence to comonotonicity, from 0.03MW to 0.24MW. Although the mean value of the aggregate power output of the WSP is defined by the marginal distributions, we can see that its sensitivity on the specific wind regime is low. Hence, for the wind regime in The Netherlands, the mean value of the output of the WSP may be expected to be in a small interval around 18.33MW. This is mainly due to the fact that the power output of the WSP is obtained as the sum of the power output of the 50 WTGs; the uncertainty is in a way shared between each set of wind speed distributions for the 50 sites and therefore the mean falls in the same region. The increase in the standard deviation of the mean distribution while shifting from independence to comonotonicity is an indication of the increase in system variability: the higher the dependence between the members of the WSP, the higher the uncertainty of the mean value of their aggregate distribution.

Similarly, in Fig. 5.6b, 11 distributions of the standard deviation for each dependence scenario are presented. These distributions indicate that the standard deviation of the

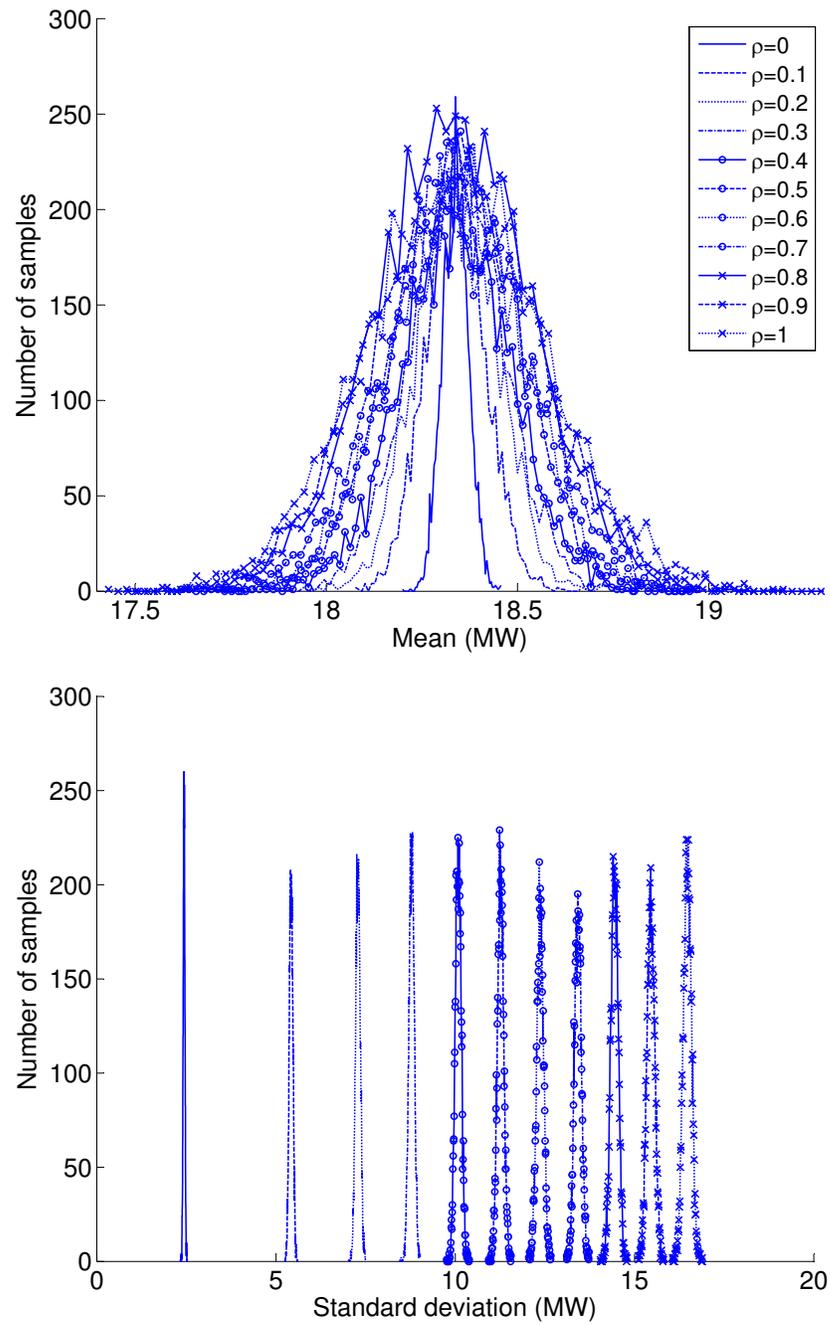


Figure 5.6: Mean value and standard deviation distributions for the WSP aggregate power output distribution for the 5000 cases of different marginals.

power output of the WSP shows a clear transition to higher variability while shifting from independence to comonotonicity. We can see from Table 5.3 that the mean values of the distributions in Fig. 5.6b increase from 2.44MW in the case of independence to 16.47MW for comonotonicity. The standard deviation of the distributions in Fig. 5.6b increases only slightly during this transition and remains low. We may therefore infer that the sensitivity of the variability of the aggregate output power of the WSP to the marginal distributions is low, and is mainly determined by the choice of the dependence structure.

Therefore, a basic result may be obtained for a WSP in a certain wind regime: the aggregate power output is mainly determined by the correct representation of the stochastic dependence structure. Thus, even if the accuracy in the definition of the wind speed marginal distributions is low, the accuracy in the definition of the aggregate power output depends mainly on the dependence assumption. The wide-spread use of the assumption of independence leads to a best-case scenario for the aggregate power output distribution, i.e. a case of minimum variability. The case of comonotonicity offers a prudent and, in many cases, adequately accurate modeling approach.

5.5 Conclusions

An alternative approach for the treatment of the high-dimensionality of the stochastic system model has been presented in this chapter. In particular, the stochastic plants concept has been introduced as model entities that bound the stochasticity of a cluster of r.v. Comonotonicity is proposed for the modeling of the dependence structure in each stochastic plant. In this case, the multidimensional dependence model is reduced to the definition of the interdependencies between the stochastic plants. The results for a wind stochastic plant show that this model offers a prudent yet accurate approximation of reality. The sensitivity of the aggregate power output of the wind stochastic plant to the marginal distributions is low. Yet, the sensitivity to the dependence structure is high.

In the following chapter, we present the application of this theory to the power system stochastic modeling. A distribution system may be reduced to a single stochastic plant, while the model reduction technique for the case of a bulk power system leads to the minimization of the dimension of the stochastic model.

Chapter 6

Applications

In this chapter, the impact of stochastic generation in power systems is investigated. Two basic planning problems are considered: the integration of SG in a distribution system and the integration of SG in a bulk power system.

1. *Distribution System*: the impact of the large-scale implementation of wind power in a typical distribution system in The Netherlands is presented. We discuss the appropriateness of different stochastic models and investigate the transition from the traditional passive to an active network structure by increasing the SG penetration level. A separate section is devoted to the voltage control problem, i.e. the analysis of the participation of SG units in the reactive power support of the network.
2. *Bulk Power System*: the transition from the vertical to a horizontal power system structure is investigated, by increasing the penetration level of wind power in a bulk power system.

6.1 Integration of SG in Distribution Systems

6.1.1 Simulation data

The study case involves the implementation of 17 WTGs in a typical 10kV radial distribution network in The Netherlands*. The system comprises a total of 34 buses and 38 lines. The system data are presented in Appendix C.3. The single-line diagram for the system is given in Fig. 6.1. The load data presented in Table C.1 correspond to the mean value of the high-load state of the respective load. Node 2 is the slack bus for the system load-flow analysis.

The mean aggregate system load for the high-load state is 3.486MW/1.851MVar. Two loading states are investigated:

1. *Current state*: the current state of the system is considered, namely 3.484MW/1.845MVar high-load state aggregate load.

*The distribution test system is located in Lelystad, The Netherlands. The data are provided by NUON.

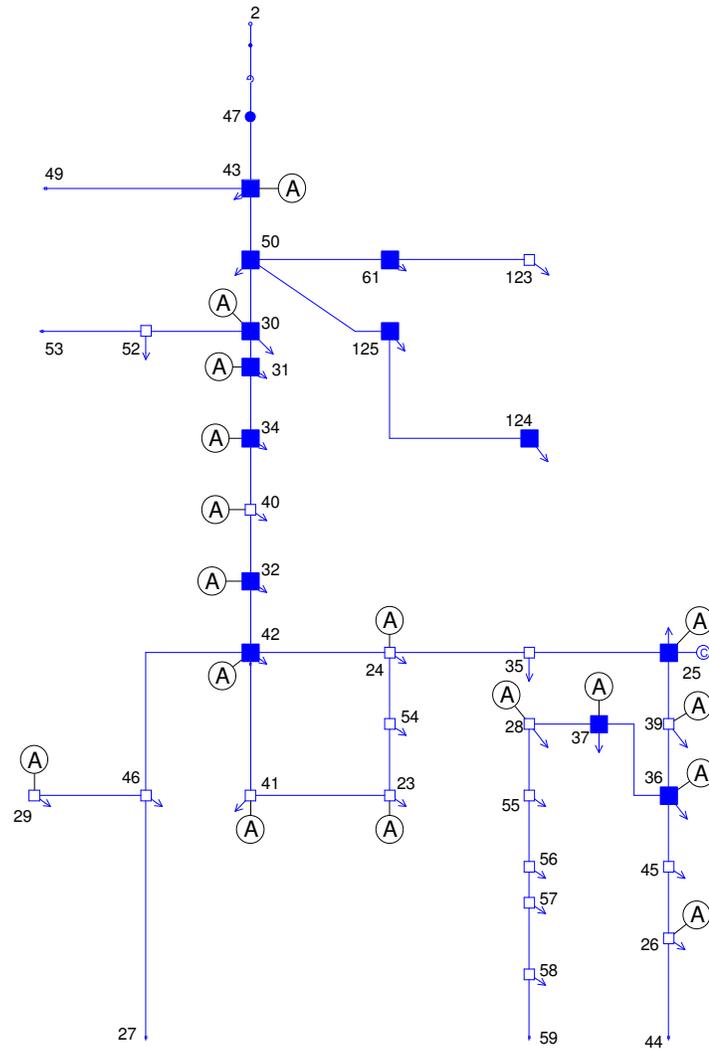


Figure 6.1: Single-line diagram for the 34-bus/38-line radial test distribution system.

2. *Future state*: a load growth of 100% is considered as a future state of the system; the high-load state aggregate load in this case is 6.972MW/3.702MVar.

Three wind power penetration levels are investigated:

1. *No penetration*: this scenario corresponds to the system operation without the incorporation of wind power.
2. *Medium penetration*: this scenario corresponds to the implementation of 17 WTGs, each of 0.2MW nominal power. In this case, the nominal wind power capacity in the system equals 3.4MW, reaching the low loading state of the system.
3. *High penetration*: the implementation of 17 WTGs, each of 0.4MW nominal power is considered. The nominal wind power capacity reaches 6.8MW, corresponding to the loading in the future state of the system.

The study case is consistent to the work presented in [70] and [68].

6.1.2 Marginal distributions

Modeling of loads

The system loads are the aggregate loads at the MV/LV transformers in the system. The methodology presented in section 2.9.2 is used for the modeling of the load active power: the active power distribution of each load is obtained as a mixture of normals corresponding to different time-frames (TF) of operation. A 3-TF segmentation is chosen. The values specified in the Table C.1 correspond to the mean of the high-load TF, while the standard deviation and the relative duration of each TF are consistent with the values presented for the 3-TF segmentation in Table 2.1. The load reactive power samples are obtained from the active power samples under the assumption of a constant load power factor.

As presented in section 2.9.2, the mixture of normals can be sampled by the use of an independent uniform r.v. U_{TF} as TF-indicator. In particular, based on the relative duration of each TF, each sample drawn from U_{TF} is matched to a specific TF and a sample is drawn from the normal distribution belonging to the respective TF. Since the U_{TF} corresponds to the time period of system operation, the time-dependence of all loads is modeled using the same indicator; according to the choice of the TF, the respective distribution is chosen for all system loads. For example, when the TF-indicator shows that the system is in the high-load state, the normals corresponding to the high-load state are chosen for all loads. Further they can be sampled based on a defined inter-TF dependence structure, corresponding to the correlation of the random component of the system load in each TF.

Modeling of wind power generation

The output of each WTG is modeled based on the principles presented in section 2.9.3. The generation sites of the WTGs are indicated in Fig. 6.1, by the notation 'A'. The Weibull parameters for the wind speed distributions in hub height at those sites are presented in Table C.1. The reactive power output of the WTGs is depends on control

strategies. The role of stochastic generators with regard to the reactive power support of the system is discussed in section 6.2.

According to the approach presented in this thesis, for the stochastic modeling of the system, the dependence structure between the system stochastic inputs has to be modeled. In the following section we investigate the appropriateness of different dependence models for the stochastic modeling of distribution systems.

6.1.3 Appropriate dependence model

For system design purposes, it is important to know the system behavior under extreme stochastic generation/consumption patterns. In most modeling cases, appropriate data are not available for the modeling of the dependence structure in the system, so a worst-case scenario approximation should be used for a safe system design.

These extreme stochastic generation/consumption patterns may be obtained by the use of the *stochastic plants* theory presented in chapter 5. Since the distribution system spans a limited geographic area, comonotonicity provides a good approximation of the dependence structure between the wind speed r.v. The same holds for the loads in each TF, since the same type of loads are present in the distribution system. Therefore, all SG units in the distribution network can be considered as one generation stochastic plant and all time-conditioned loads as one load stochastic plant.

Based on the extreme dependence concepts presented in chapter 5, three extreme cases for the modeling of the dependence structure between the two stochastic plants (intra-SP dependence structure) can be considered:

1. *Comonotonicity*: the intra-SP dependence structure is modeled based on comonotonicity. Since the aggregate power output of the SPs corresponds to their difference, this case corresponds to the lower stochastic bound of the intra-SP dependence structure and therefore the lower extreme case for the stochastic modeling of the system.
2. *Independence*: independence is used for the modeling of the intra-SP dependence structure. A moderate extreme for the system modeling is obtained in this case, yet applicable in many cases when the SG presents no correlation to the system load.
3. *Countermonotonicity*: this case corresponds to the upper stochastic bound of the intra-SP dependence structure and therefore the upper extreme case for the stochastic modeling of the system (maximum variability)[†].

The system was simulated for the above three extreme dependence scenarios. In addition, the system lower stochastic bound was simulated, i.e. the case of total independence between the system inputs (loads and WTGs). In Fig. 6.2, the results of a 5000-sample MCS for the voltage distributions at a typical system node for the four different dependence scenarios and two wind power penetration levels (no penetration/high penetration) are presented. In all cases, the same marginals are used and only the dependence structure is changed. We can see that the case of total independence (labeled '*lower bound*') yields a minimum variability of the voltage distribution. The use of this assumption leads to fallacy, since the minimum variability for the node voltage is obtained.

[†]This case corresponds to a stochastic plant comprising both SG units and loads, as presented in section 5.3.2.

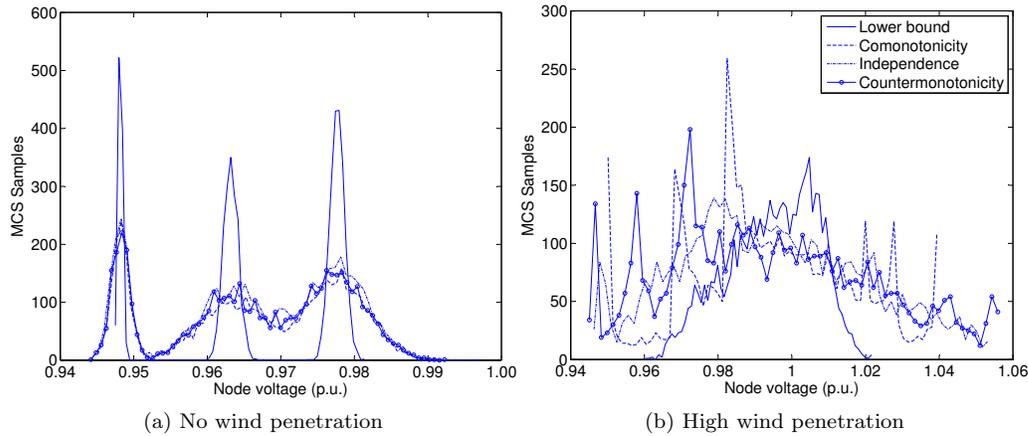


Figure 6.2: Voltage distributions at node 42 under different extreme dependence scenarios.

We can also see that the implementation of SG leads to an increase in the variability of the node voltage. The system upper stochastic bounds in this case yield a higher voltage deviation. Fig. 6.2b shows that the assumption of total independence results in a voltage that doesn't exceed the limits of 0.95p.u. and 1.05p.u. However, the upper stochastic bounds show that this is a fallacy. The positive dependence between the system inputs leads to an increase in the variability of the node voltage and a severe risk of exceeding the voltage limits.

In Fig. 6.3, the power flow distributions in the 47-43 are presented, i.e. the power flow between the distribution network and the transmission system. It is evident that the same results are obtained: the lower stochastic bound yields the minimum variability for the power flow distributions. Fig. 6.3b shows that the wind power integration in the system leads to a severe increase in the variability of the power flows. The assumption of total independence leads to a severe underestimation of this variability and a faulty system design. In particular, the lower stochastic bound shows that the link should be dimensioned for about 2MW of power, while the upper stochastic bounds show that an inverse power flow of 6MW can occur.

In Table 6.1, the mean values and standard deviations for the voltage and power flow distributions in the different dependence concepts are presented. We can see that for each penetration level, the distributions present the same central points (mean), while the standard deviations are maximal for the upper stochastic bounds. The integration of wind power in the network causes a severe increase of the variability of the voltage distributions, which is maximal in the case of intra-SP countermonotonicity. This case yields the worst-case scenario for the system.

The upper stochastic bound corresponds to the maximum variability of the power injections into the system, i.e. maximum system stochastic stress, maximum power flows and maximum voltage deviation. As this case offers a prudent system design in terms of safety and generality, the upper stochastic bound will be further used for the system analysis.

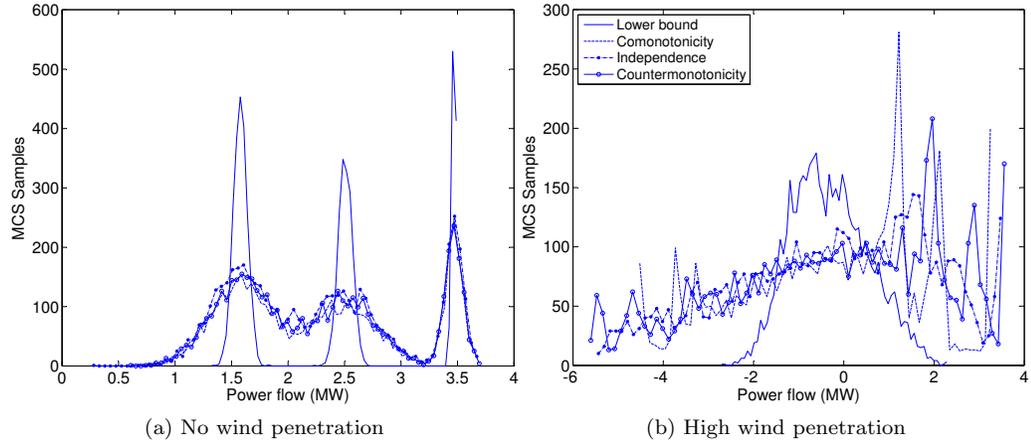


Figure 6.3: Power flow distributions in branch 47-43 under different extreme dependence scenarios.

Table 6.1: Mean values and standard deviations for the voltage and power flow distributions in the different extreme dependence concepts.

	Voltage node 42 (p.u.)				Power flow line 47-43 (MW)			
	Mean		St.D.		Mean		St.D.	
	No	High	No	High	No	High	No	High
Wind Penetration:								
Lower bound	0.97	0.99	0.0113	0.0119	2.30	-0.30	0.72	0.85
Comonotonicity	0.97	0.99	0.0120	0.0218	2.30	-0.27	0.77	2.00
Independence	0.97	0.99	0.0119	0.0350	2.29	-0.30	0.76	2.22
Countermonotonicity	0.97	0.99	0.0119	0.0377	2.29	-0.27	0.76	2.40

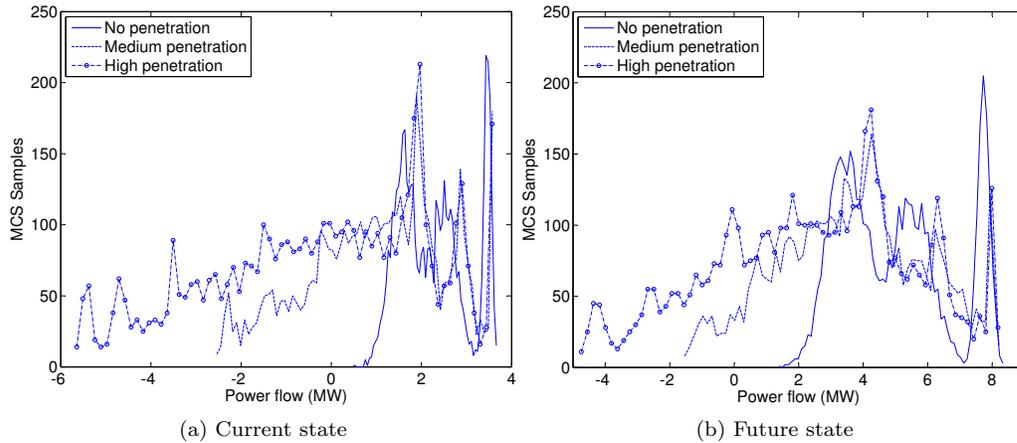


Figure 6.4: Power flow distributions in the branch 47-43, in two system states under different wind power penetration levels.

Table 6.2: Mean values/standard deviations of the power flow distributions in line 47-43.

Penetration level:	Mean value (MW)			Standard deviation (MW)		
	No	Medium	High	No	Medium	High
Current state	2.29	0.99	-0.25	0.76	1.51	2.39
Future state	5.02	3.68	2.39	1.71	2.29	3.09

6.1.4 Active distribution network

In a horizontally-operated power system, the distribution systems become *active* networks; besides distributed loads, they also contain generation. The power flow between the transmission system and the active distribution networks is no longer uni-directional (down) but can be bi-directional. In this section, we investigate this transition towards an active distribution network due to the incorporation of wind power in the distribution system of Fig. 6.1.

In particular, the power flow distribution in the link between the HV and the MV system (branch 47-43) is computed. In Fig. 6.4, the results of a 5000-sample MCS are presented for the three wind power penetration levels and the two system states defined in section 6.1.1. In the case that there is no wind power penetration, the distribution system is a passive network and the power flow distribution is always positive (power flow from the transmission system towards the distribution network). The increasing wind power penetration induces negative power flows, i.e. power flows from the distribution to the transmission system.

The distributions in Fig. 6.4 show the transition from a passive to an active network structure due to the integration of wind power. The power flows become bi-directional and the distributions span the negative axis.

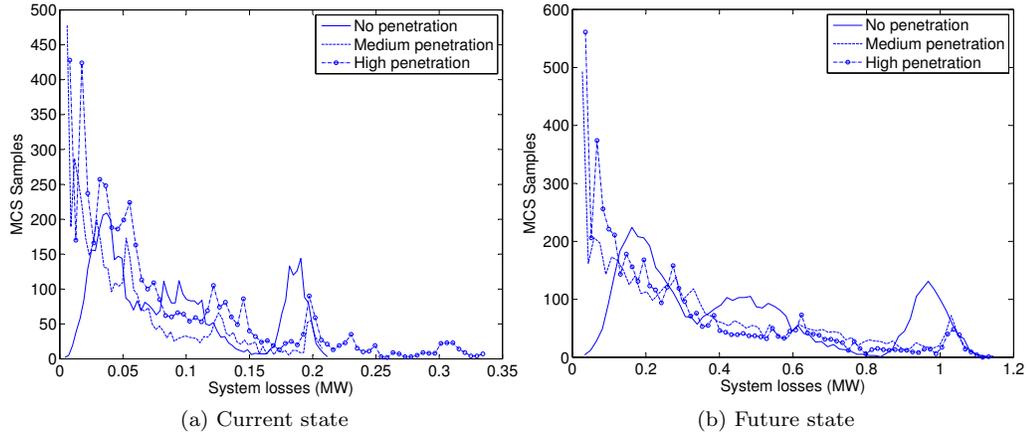


Figure 6.5: System power loss distributions for two system states and three wind power penetration levels.

Table 6.3: Mean and annual system losses.

Penetration level:	Mean system losses (MW)			Annual system losses (GWh)		
	No	Medium	High	No	Medium	High
Current state	0.089	0.056	0.078	0.78	0.49	0.69
Future state	0.45	0.34	0.28	3.94	2.94	2.48

6.1.5 System losses

In Fig. 6.5, the system losses are presented for the two loading states of the system and for the different penetration levels of wind power. In Table 6.3 the mean values for the distributions of the system losses of Fig. 6.5 are presented.

The integration of wind power has a diverse impact on the system losses, depending on both the penetration level and the system state. In the current system state, the medium penetration level leads to a minimization of the system losses, due to the reduction of the system power flows, while in the high penetration level the system losses increase again, due to the presence of high reverse power flows. For the future system state, the increase in wind power penetration leads to a subsequent reduction of the system losses.

Thus, the anticipated loss reduction due to the incorporation of SG in the distribution systems depends on the loading level of the system. It should be noted however that by choosing the upper stochastic bound for the system dependence modeling, the resulting distributions correspond to the maximum system losses since the dependence model corresponds to the maximum stochastic power flows in the system.

6.2 Voltage control by SG units

Besides the supply of active power to the system, the DG units can perform additional tasks that are mainly associated with system stability and the quality of power supply, the so-called ancillary services. Those ancillary services are achieved by connecting the DG units to the system via 'smart' *power-electronic converters (PECs)*-interfaces. The primary task of the PECs is to control the operation and optimize the real power output of the DG unit, while their secondary task involves the following functions (*ancillary services*):

- reactive power support of the system (voltage control),
- improvement of power quality (voltage sag compensation, harmonics filtering),
- fault behavior, frequency control.

For an extensive analysis on these issues, one should refer to [61]. Here we investigate the role of SG units in a distribution system with regard to the first issue, i.e. the reactive power support of the system.

6.2.1 Voltage control: theory

The objective of voltage control is to maintain the RMS value of the voltage within specified limits, independent of the generation and consumption [57]. Conventional voltage control in the HV transmission network is mainly performed by the large power plants. In distribution networks, the voltage control is done by tap changers in distribution transformers. This control is relatively slow and compensates for the voltage drop along the line, based on the assumption that only loads are connected to it. As discussed above, the introduction of DG units changes the power flows in the distribution system. The changes in magnitude and direction of the power result in changing voltages due to the current-dependent voltage drop along the line, which has a relatively high impedance in the LV and MV networks [61]. DG units connected to the grid by PECs can contribute to the voltage control, by regulating their reactive power output. Due to their geographical dispersion, distributed/local voltage control can be performed.

DG reactive power capacity

The reactive power output of the DG unit i can take all values in an interval ranging between maximum inductive ($-Q_{i_{max}}$) and maximum capacitive ($+Q_{i_{max}}$) power, that is limited by the maximum apparent power output of the DG unit S_i , which in its turn is limited by the maximum current rating of the PEC. If $P_i(t)$ is the active power production of DG unit i at time t , the maximum reactive power that can be delivered by the unit is:

$$Q_{i_{max}}(t) = \pm \sqrt{S_i^2 - P_i^2(t)} \quad (6.1)$$

Thus, the reactive power capacity of each unit $Q_{i_{max}}$ varies in time, depending on the active power generation $P_i(t)$. As such, the voltage control capability of the unit varies in time and depends on the active power output of the unit. In the case of SG, $P_i(t)$

is a random variable; therefore, the reactive power capacity of each unit will also vary stochastically.

This necessitates a stochastic approach to analyze the contribution of stochastic generation in the reactive power support of the system. By sampling the vector \mathbf{P} , we obtain the reactive power capacity of each unit $Q_{i_{max}}$ and the problem is decomposed from the time-domain. Knowing the vector \mathbf{Q}_{max} for the different snapshots of operation allows us to proceed with the analysis of different voltage control strategies for the system operation.

Voltage control strategies

The voltage V at the connection node defines the general rule that is applied for the regulation of the reactive power of the unit. In general, two types of voltage control algorithms can be recognized:

1. *Threshold control*: an interval $[V_{min}^T, V_{max}^T]$ is defined where the voltage is considered to be normal and no action is necessary. The thresholds V_{min}^T, V_{max}^T are defined for a specific voltage deviation ΔV_T from the nominal voltage V_{nom} : $V_{min}^T = V_{nom} - \Delta V_T$ and $V_{max}^T = V_{nom} + \Delta V_T$. Typical values for V_{nom} and ΔV_T are: $V_{nom} = 1p.u.$ and $\Delta V_T = 0.05p.u.$ When V exceeds these thresholds, the reactive power output is set to a predefined value Q_{thr} :
 - IF $V \leq V_{min}$ THEN inject Q_{thr} to the system (capacitive behavior)
 - IF $V \geq V_{max}$ THEN absorb Q_{thr} from the system (inductive behavior).

Different setpoints can be considered for the reactive power Q_{thr} :

- *Constant power factor*: the reactive power is supplied by a predefined power factor: $Q_{i_{thr}} = \pm \tan \phi \cdot P_i$. A typical value for the power factor is $\cos \phi = 0.9^\ddagger$.
 - *Maximum reactive power*: maximum reactive power is supplied: $Q_{i_{thr}} = Q_{i_{max}} = \pm \sqrt{S_i^2 - P_i^2}$
2. *Droop control*: a droop characteristic is used for the regulation of the reactive power output of the unit in the interval $[V_{min}^D, V_{max}^D]$, where $V_{min}^D = V_{nom} - \Delta V_D$ and $V_{max}^D = V_{nom} + \Delta V_D$. Typical values for V_{nom} and ΔV_D are: $V_{nom} = 1p.u.$ and $\Delta V_D = 0.1p.u.$ For $V \in [V_{min}^D, V_{max}^D]$, the reactive power output of the unit i is proportional to the voltage deviation $\Delta V_i = V_{nom} - V_i$:

$$Q_i = \frac{Q_{i_{max}}}{\Delta V_D} \cdot \Delta V_i = c_i \cdot \Delta V_i \quad (6.2)$$

Outside the interval, maximum reactive power is supplied: $Q_i = Q_{i_{max}} = \pm \sqrt{S_i^2 - P_i^2}$

In Fig. 6.6, the $V - Q$ characteristics of the *maximum reactive power threshold control* and the *droop control* are presented. The gradient of the droop control strategy is defined based on $Q_{i_{max}}(t)$. Since $Q_{i_{max}}(t)$ varies stochastically in time, the $V - Q$ characteristic will also vary in time. An example of this variation is depicted in Fig. 6.7, where the respective characteristics for different points in time are presented. Thus, for each sample, a different voltage control characteristic is obtained for the operation of the unit.

[‡]This control strategy is used for the reactive power regulation of CHP plants.

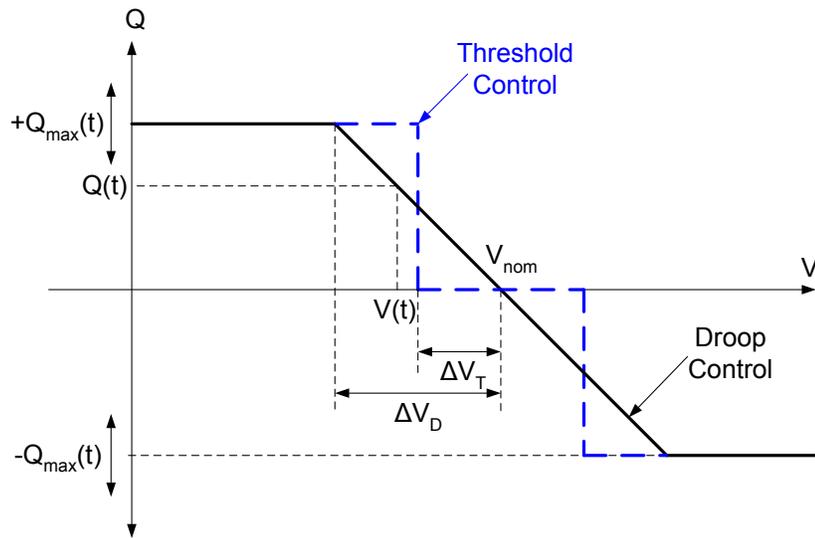


Figure 6.6: Voltage control characteristics of the *maximum reactive power threshold control* and the *droop control*.

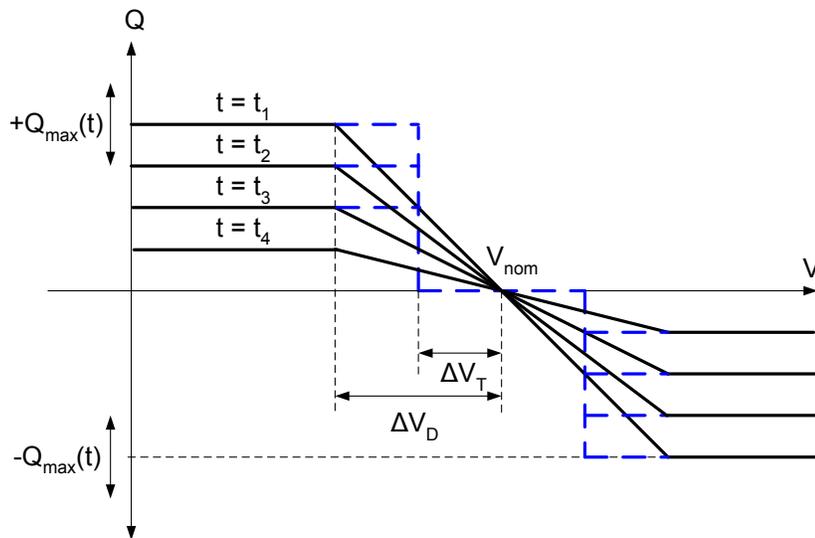


Figure 6.7: Voltage control characteristics for different points in time.

Equilibrium point

By engaging all units in the voltage control, a distributed control is achieved: the reactive power support of the system is distributed among the units, based on the voltage control at the connection nodes. Each unit is assigned a local task (reactive power injection), which induces a global system change (new voltage profile). This change in voltage can initiate a new (distributed) control action and this procedure continues until the system reaches an equilibrium point. This equilibrium point depends on the voltage control strategy and the active power injections in the system.

Thus, although the voltage control is a local action, it influences a global system variable, the system voltage. Therefore, the voltage control strategies should be evaluated based on their impact on the system operation.

Solution algorithm

The voltage control analysis of the system should be performed in two steps:

1. *Stochastic time-decomposition*: the system inputs are sampled and respective snapshots of the operation are obtained. For each snapshot the maximum reactive power capacity of each SG unit $Q_{i,max}(t)$ is calculated and subsequently the voltage control characteristics corresponding to the specific time instant are obtained. The system upper stochastic bound is used for the sampling of the system inputs. This leads to the worst-case scenario for the voltage distributions in the system.
2. *Iterative procedure*: an iterative procedure is applied in order to reach the equilibrium point for each snapshot of operation obtained by the stochastic time decomposition. The first iteration is calculated as the system steady-state with zero reactive power output from the DG units. Based on the obtained voltage profile, a control action is applied at the DG node where the maximum voltage deviation is observed (in the case of threshold control algorithms the control action takes place only if the voltage deviation exceeds ΔV_T)[§]. The system steady-state after the incorporation of this change is re-calculated and a new voltage profile is obtained. This procedure continues until the equilibrium point is reached, i.e. when the voltage profile between consecutive runs of the algorithm converges, or when there is no other available action that can be performed by the DG units. In Fig. 6.8, the iterative algorithms for the modeling of the voltage control in the system are presented.

The main assumption for the application of this method is that the system reaches this equilibrium point in each instant of time. In practice, the system operation is a continuous process, where the active power output of the SG units is constantly changing. According to this change, the reactive power output is also changing due to the voltage control. Since the response of the PECs to each control action is very fast, the validity of this assumption is not endangered.

[§]An alternative implementation of the algorithm is to apply the control action in all DG nodes in each iteration. This implementation showed better performance for the droop control algorithm. However, in the case of threshold control algorithms it may lead to a solution where more units are set in voltage control than necessary. By applying one control action on each iteration we employ the minimum amount of units for voltage control.

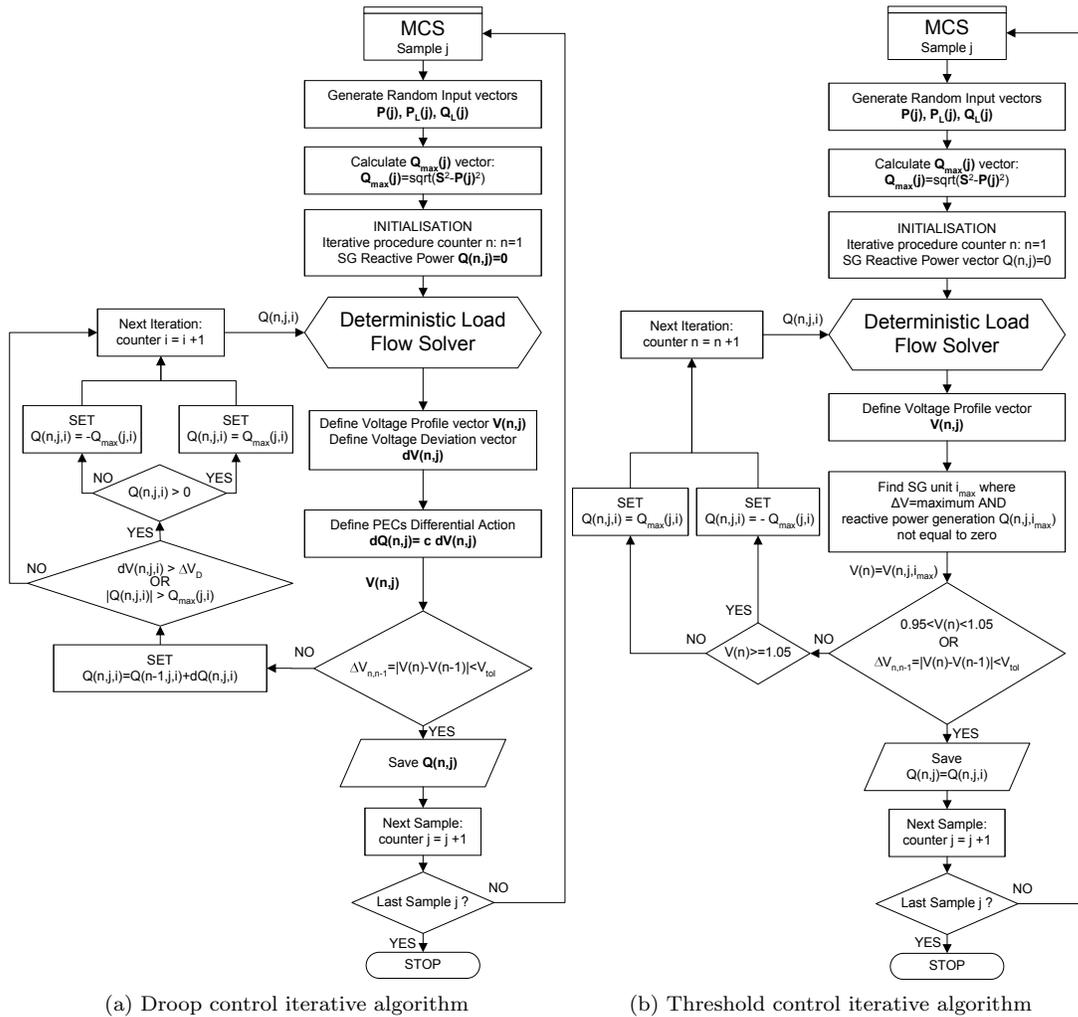


Figure 6.8: Voltage control iterative algorithms.

Table 6.4: Mean system reactive power production for the SG units [MVar].

Penetration level:	Current state			Future state		
	No	Medium	High	No	Medium	High
No voltage control	0	0	0	0	0	0
Threshold control	0	0.0003	-0.0001	0	0.0115	0.0129
Droop control	0	0.1036	0.0732	0	0.1572	0.2421

6.2.2 Voltage control: results for the study case

The capability of WTGs to contribute to the voltage control in a distribution system is investigated by the analysis of the distribution test system in Fig. 6.1. In particular, the system voltage profile is simulated for the three wind power penetration levels and the two system states, under three different voltage control strategies:

1. *No voltage control*: the reactive power output of the WTGs is zero, i.e. the DG units provide no ancillary services.
2. *Threshold control*: the constant power factor threshold control strategy is applied. The power factor is $\cos \phi = 0.9$ and the $\Delta V_T = 0.05p.u.$ This case is consistent with the analysis presented in [66].
3. *Droop control*: the droop control strategy is implemented with $\Delta V_D = 0.1p.u.$

In this analysis, the voltage distributions at all the system nodes in the two system states with the three wind power penetration levels are computed. In Fig. 6.9, these distributions at node 42 in the system are presented. In Table 6.4, the mean reactive power production by the SG units in the network for all the investigated scenarios is presented. The results show that the droop control strategy improves the voltage quality in the system, while the threshold control strategy has a low impact on the improvement of the system voltage.

This low performance of the constant power factor threshold control strategy is mainly due to the fact that for this strategy the reactive power output of the SG unit is proportional to the active power. As presented in Fig. 2.11, the WTG power output distribution presents a concentration of probability mass in the zero and nominal values. Thus, when P is zero, the reactive power output of the unit is also set to zero. Therefore, the probability mass at the zero output in the WTG provides zero reactive power output. The same is valid for the probability mass at the nominal power output: when the power output is nominal, the reactive power is also set to zero due to the current rating limit (equation 6.1). Therefore, the only cases when voltage control is applied is when the active power output is neither zero nor maximum and the voltage is outside the interval $[0.95 \ 1.05]$. This leads to a very low reactive power yield, as we can see in Table 6.4. The constant power factor threshold control strategy is inappropriate for controlling WTGs; due to the the active power output distribution of a WTG, this control strategy leads to a low utilization of the reactive power resources, defined by equation 6.1. The use of the maximum reactive power threshold control strategy is more appropriate for the control of WTGs, since the power output is defined based on the available reactive power resources.

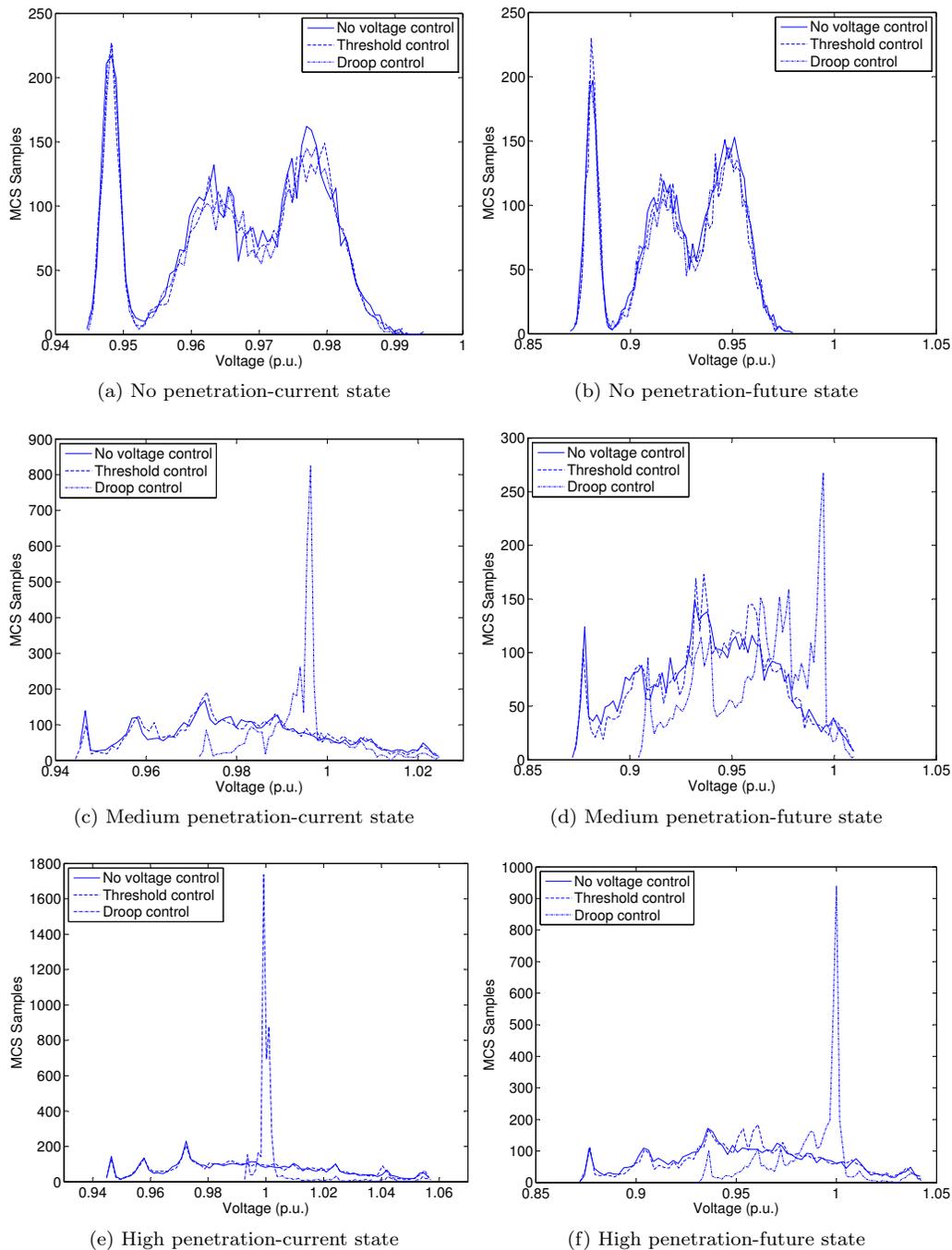


Figure 6.9: Voltage distributions at node 42 for the different voltage control strategies (5000-sample MCS).

Table 6.5: Characteristics of the New England test system.

System characteristic	value
Busses	39
Central generators	10
Loads	19
Transmission lines	46
Voltage level	100kV
Total generation	6140.7MW / 1264.3MVar
Total load	6097.1MW / 1408.7MVar

The droop control strategy allows a good control of the voltage in the system. In Fig. 6.10, the box-plots of the voltage distributions at the system nodes without voltage control and with the droop voltage control are presented[¶]. The four-digit node numbers in the box-plot '10-' correspond to open ends of system lines (not connected to the node). We can see that the introduction of WTGs leads to the improvement of the voltage profile in the system, i.e. at all system nodes the variability of the voltage is decreased and the voltage is kept within the operating limits of $[0.9p.u \ 1.1p.u]$.

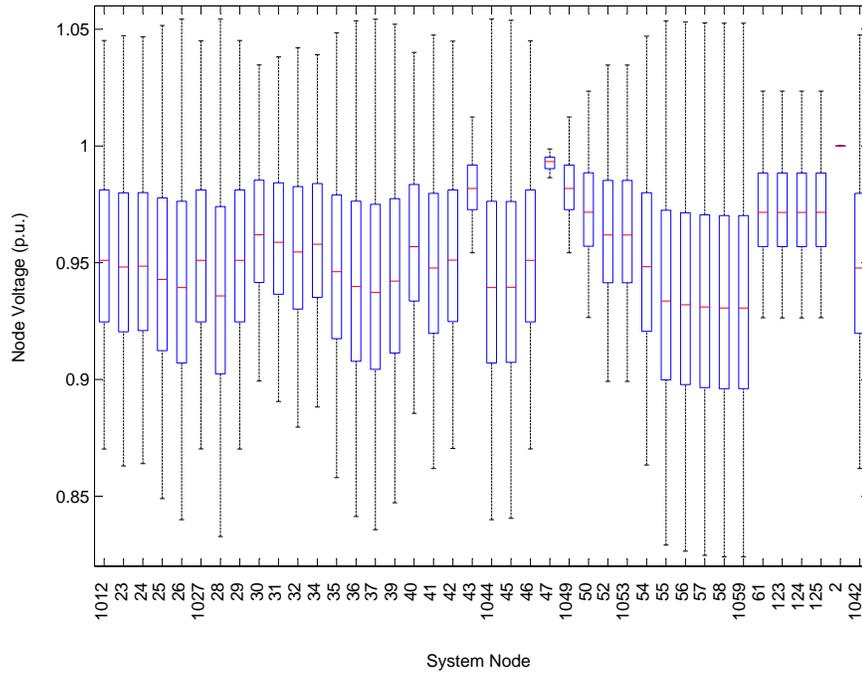
These results are based on the upper system stochastic bound, which is the worst-case scenario for the system stochasticity. Since the droop control strategy keeps the voltages in the system inside the operational limits for the worst-case scenario, we can be sure that no voltage problems will arise in the system. In fact, the incorporation of SG units actually leads to an *improvement* of the voltage quality in the network, through the incorporation of the voltage control.

6.3 Integration of SG in a Bulk Power System

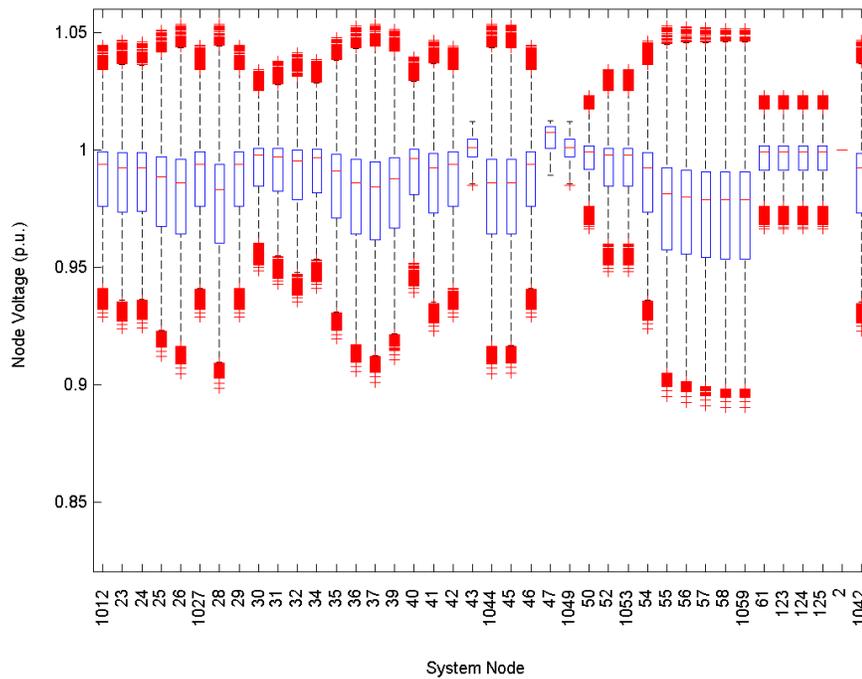
6.3.1 Simulation data

The study case involves the large-scale integration of wind power in a bulk power system. The IEEE 39-bus New England test system is used as a system model. The system comprises 39 buses, 10 CG units and 46 transmission lines (100kV). The single-line diagram of the system is given in Fig. 6.11. The basic characteristics of the test system are presented in Table 6.5, while in Appendix C.4 the system data are presented. The wind generation is geographically distributed throughout the system and connected to 15 system nodes as presented in Fig. 6.11, indicated by W .

[¶]In a box-plot, the box has lines at the lower quartile, median, and upper quartile values. Lines extending from each end of the box to show the extent of the rest of the data (whiskers). Outliers are data with values beyond the ends of the whiskers. Each of these data is represented by the marker '+' and corresponds to the tails of the distributions.



(a) No voltage control



(b) Droop control

Figure 6.10: Box-plots of the nodal voltage distributions for the highest wind penetration level in the future system state (5000-sample MCS).

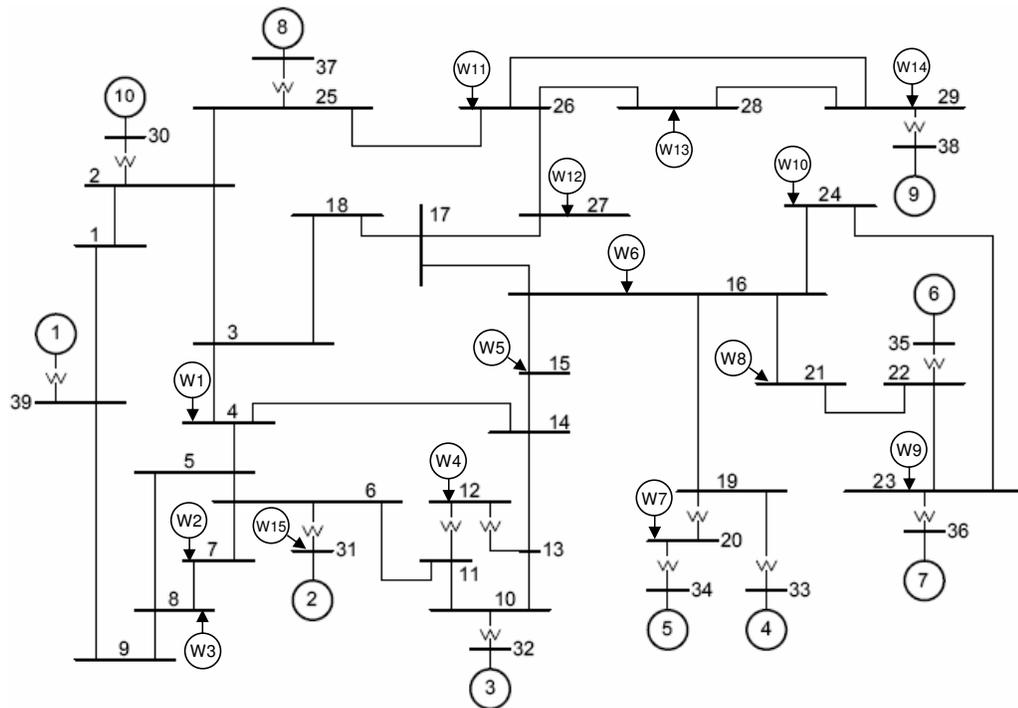


Figure 6.11: Single-line diagram of the 39-bus New England test system [65].

Table 6.6: TF settings for a 4-TF load modeling of the New England test system.

	Time Ratio	Mean Load (% high-load-TF mean)	St. Deviation (% mean load)
<i>TF1</i>	0.2	0.5	0.06
<i>TF2</i>	0.3	0.65	0.1
<i>TF3</i>	0.3	0.85	0.1
<i>TF4</i>	0.2	1	0.03

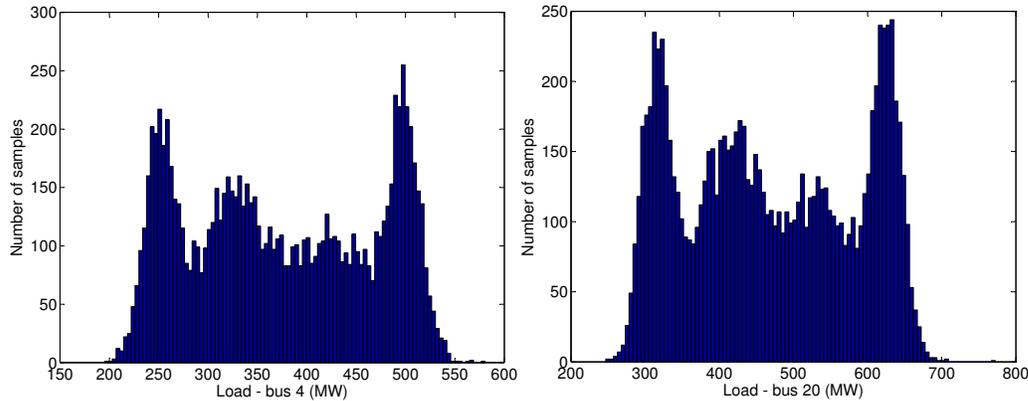


Figure 6.12: 4-TF load modeling for the New England test system (10000-sample MCS).

6.3.2 System loads

Marginals

A 4-TF segmentation is chosen for the modeling of the system loads, according to the analysis presented in section 2.9.2. Accordingly, a normal distribution is used for the modeling of the time-conditioned load in each TF and the resulting distribution is obtained by an aggregation procedure as a mixture of these normals. The settings for the 4-TF modeling are presented in Table 6.6. The load data presented in Appendix C.4 correspond to the mean value of the high-load-TF of the respective load. In Fig. 6.12, the resulting distributions for a 10000-sample MCS for the loads in buses 4 and 20 are presented.

Dependence structure

Two different approaches can be followed for the modeling of the dependence structure between the time-conditioned loads:

1. *Intra-TF dependence*: the dependence structure is defined for the aggregate distribution. So, first the marginal distribution is generated for all TFs and then the JNT methodology is used for the modeling of the correlated r.v. In Fig. 6.13a, the scatter diagram for the modeling of an intra-TF rank correlation $\rho_r = 0.7$ between the

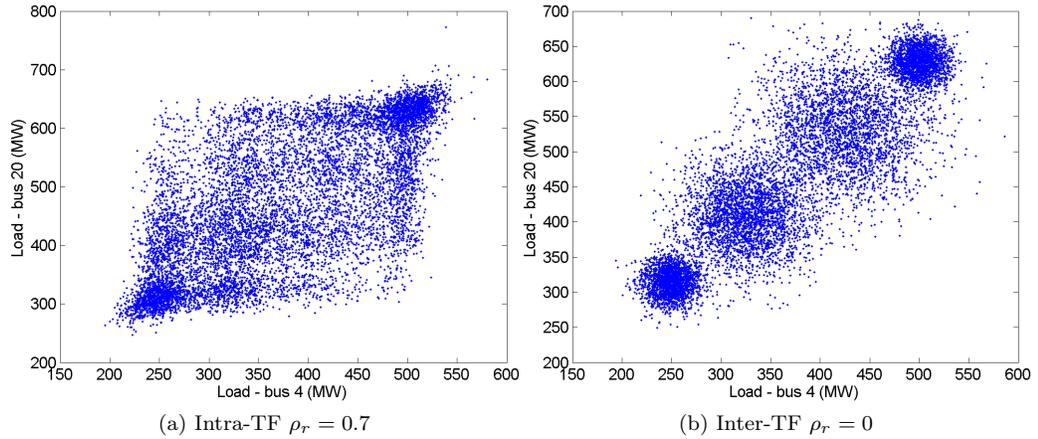


Figure 6.13: Scatter diagrams for the load modeling based on specified intra- and inter-TF rank correlation (10000-sample MCS).

loads in buses 4 and 20 is presented for a 10000-sample MCS. The rank correlation obtained at the output samples is 0.69.

2. *Inter-TF dependence*: the dependence structure is defined inside each time-frame. In this case, the time-dependence is superimposed on the inter-TF dependence and the correlation between the aggregate distributions is different. For this, first the TF of operation is chosen and then the JNT methodology is used for the generation of the correlated normals in the respective TF. Since the TF analysis corresponds to time-conditioning, all system loads are considered to be in the same TF. This imposes a high dependence between the loads. In Fig. 6.13b, the scatter diagram for the modeling of an inter-TF rank correlation $\rho_r = 0$ between the loads in buses 4 and 20 is presented for a 10000-sample MCS. Although in each TF the r.v. are independent, the resulting distributions are highly correlated. In particular, the rank correlation obtained at the output samples is 0.89.

The modeling approach can be decided depending on the specific problem. When we are not interested in modeling the time-dependence and we need to model based on specific correlations between the aggregate distributions, the first approach is applied. When the dependence structure in each TF should be modeled, the second approach should be used. For the present study case we choose the second approach for the modeling of time-dependent loads in the system. All loads are considered to be in the same TF for each sample and independence is used for the modeling of the time-conditioned normals in each TF. Based on this modeling, the obtained mutual rank correlation between the aggregate load distributions is 0.89.

Table 6.7: Capacity factor for the WSPs in the New England test system.

WSP	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8
η_W	39.2	39.3	38.2	37.9	35	40.4	40.2	40.7
WSP	W_9	W_{10}	W_{11}	W_{12}	W_{13}	W_{14}	W_{15}	
η_W	40.3	40.5	39.5	38.5	39.1	39.8	38.9	

Reactive power

The load reactive power samples are obtained from the active power samples based on a constant load power factor. The power factor is computed for each load by the mean values given in Table C.2 (Appendix C.4).

6.3.3 System wind power

Marginals

The wind generation indicated with W in the single-line diagram in Fig. 6.11 corresponds to clusters of WTGs connected to the underlying distribution network. The wind power generation of a cluster is modeled based on the wind stochastic plant (WSP) theory presented in chapter 5. For each WSP, data corresponding to 10 sites in the underlying system are considered. Therefore, the analysis comprises a total of $15 \times 10 = 150$ wind speed r.v. The Weibull parameters β and η for the wind speed distributions at hub height for the 150 generation sites are generated as random numbers drawn in the interval: $\beta \in [1.9, 2.6]$ and $\eta \in [8, 10]$.

Four wind power penetration levels are considered for the system planning:

1. *No penetration*: this case corresponds to the system operation with no wind power (vertically-operated power system).
2. *100MW capacity for each WSP*: the nominal wind power capacity in the system is in this case $15 \times 100MW = 1500MW$, corresponding to a penetration level of 25%.
3. *200MW capacity for each WSP*: the nominal wind power capacity in the system is in this case $15 \times 200MW = 3000MW$, corresponding to a penetration level of 50%.
4. *300MW capacity for each WSP*: the nominal wind power capacity in the system is in this case $15 \times 300MW = 4500MW$, corresponding to a penetration level of 75%.

It should be mentioned here that increasing the penetration level up to 100% leads to convergence problems. In particular, for 112 samples out of a total of 10000 samples, the power flow algorithm could not converge.

In Fig. 6.14, the power output of the WSPs connected at the buses 4 and 20 for the penetration level 2 are presented. The capacity factors for the 15 WSPs in the system range between 35% and 40.7% (Table 6.7).

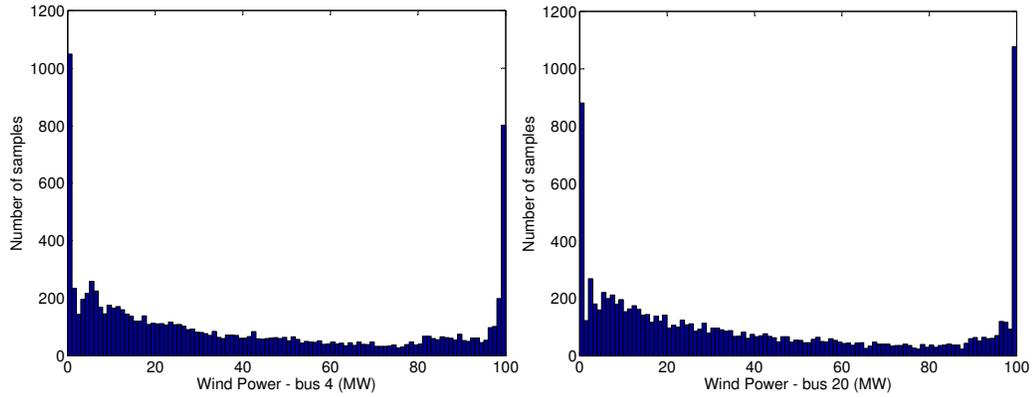


Figure 6.14: WSP power output in the New England test system (10000-sample MCS).

Dependence structure

The wind resources between the 15 different sites in the system are considered to be correlated. The mutual correlations between the different stochastic plants are considered to be equal to 0.7. The JNT methodology is used for the modeling of the correlated Weibull wind speed distributions based on the specific correlations. In Fig. 6.15a, the scatter diagram for the wind speed r.v. at WTG 3 of the WSP at bus 4 and WTG 3 of the WSP at bus 20 is presented for a 10000-sample MCS. The rank correlation measured at the output wind speed samples equals $\rho_r^{W-W} = 0.7021$ for all sites belonging to different WSPs and equals 1 for sites belonging to the same WSP.

In Fig. 6.15b, the scatter diagram for the wind power output of the WSPs connected at bus 4 and 20 is presented. We can see that the impact of the non-linear wind speed/power characteristic and the sum of the different wind power distributions lead to a transformed scatter diagram. The rank correlation for the output wind power samples yields values in the interval $\rho_r^{P_W-P_W} \in [0.6903 \ 0.7173]$. This change in correlation is due to the impact of the non-increasing wind speed/power characteristic of the WTG. In particular, for the nominal and cut-out wind speeds, the transformation from wind to power is non-monotonic. The rank correlation is thus not invariant during this transformation. However, as may be seen, the wind power correlation values are close to the wind speed ones because the wind speed/power characteristic is for most of the wind speed values monotonic.

Reactive power

The wind parks are considered to be not responsible for the voltage support of the system. Therefore, the reactive power generation of the wind generators is zero.

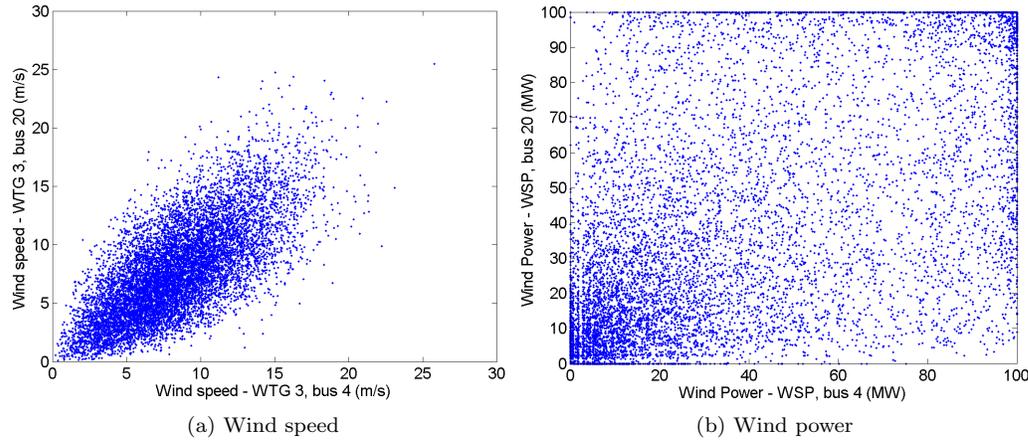


Figure 6.15: Wind speed and wind power scatter diagrams (10000-sample MCS).

Table 6.8: Mean value and standard deviation for the power injections at bus 4 for the 4 wind power penetration levels.

Wind Penetration:	0 MW	1500 MW	3000 MW	4500MW
Mean [MW]	373.91	335.77	296.85	259.18
St.D.[MW]	93.23	99.40	117.75	139.77

6.3.4 System power injections

The system loads are considered to be independent from the wind power injections in the system^{||}. In Fig. 6.16, the power injection distribution (net load distribution) at bus 4 of the test system for the 4 wind power penetration levels is presented. We can see that the connection of wind power at the bus leads to an increase in the variability of the power injections; the lower tail of the distribution is shifted towards the negative axis. This may be seen in Table 6.8 where the mean and the standard deviation of the power injections are presented. In particular, the increase in wind penetration leads to a decrease of the mean value of the distribution and a subsequent increase in the standard deviation. At bus 4, the connection of a wind park of 300MW nominal power (penetration level 4) leads to bidirectional power injections into the system.

6.3.5 System operation: CG units

The dispatch of the CG units in the system should be modeled for each system state, defined by the sampling of the system loads and wind power. In Table C.2 (Appendix C.4), the capacity C_c^M of the CG units is presented. CG unit 2 connected at bus 31 is the slack bus of the system.

In the specific study case, the CG units in the system are considered to be thermal

^{||}This assumption is in accordance with the results presented in section 4.2 for the Dutch power system.

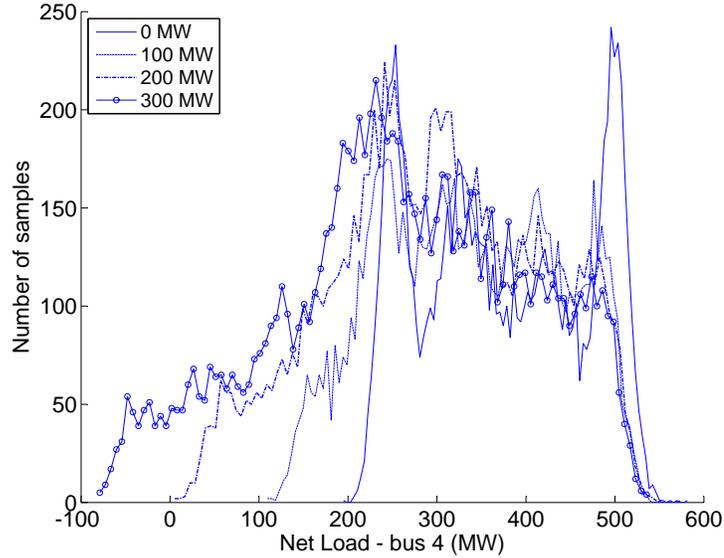


Figure 6.16: Power injection at bus 4 for the 4 wind power penetration levels (1000-sample MCS).

units of the same type. For each sample, the power production of the CG units (except for the slack bus) is pro-rata (see algorithm 6.1). For each CG unit, the minimum power output is 10% of the unit capacity, due to restrictions with shutting down thermal units. For a system comprising

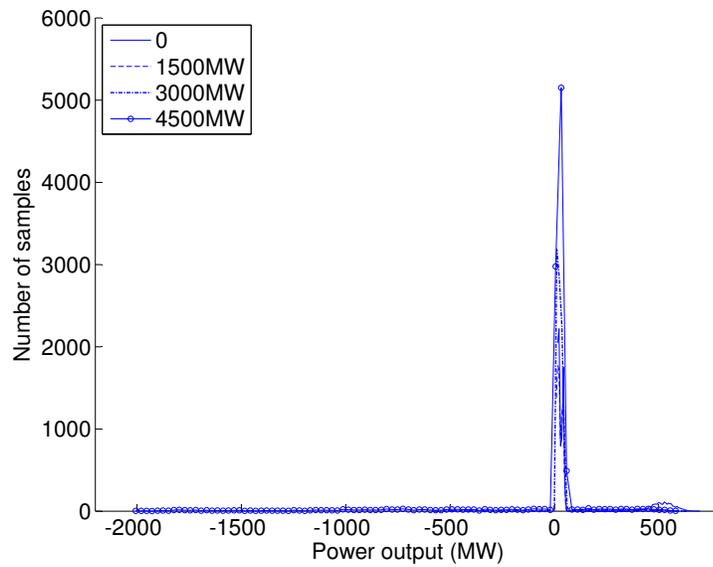
- c_T CG units C_1, C_2, \dots, C_{c_T} ,
- g_T SG units G_1, G_2, \dots, G_{g_T} ,
- l_T loads L_1, L_2, \dots, L_{l_T} ,

the CG unit dispatch algorithm is presented in algorithm 6.1.

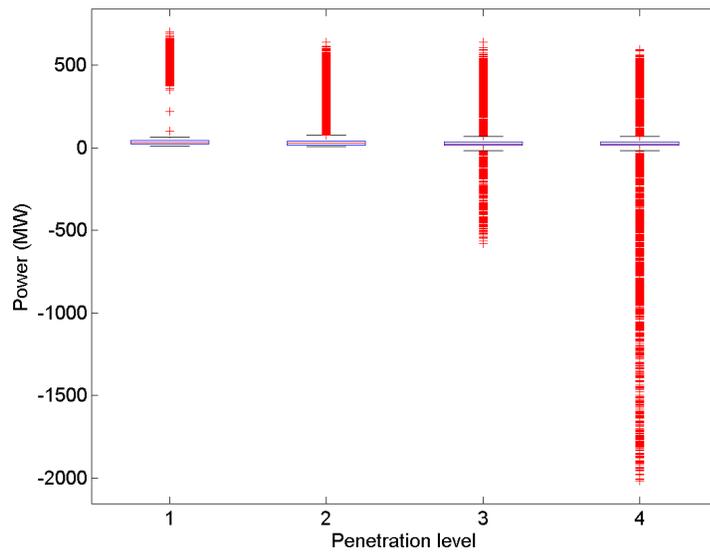
In this case, the system slack bus covers the system losses and also consumes the excess of wind power when the CG units are at their minimum production.

6.3.6 System slack bus

In Fig. 6.17, the power distributions and box-plots for the power injection by the slack bus are presented for the different wind power penetration levels. In Table 6.9, the respective mean values and standard deviations for the slack bus power injection distributions are presented. We can see that the increase in wind power penetration leads to a radical increase in the variability of the power flows from/to the slack bus. In particular, for the penetration level of 4500MW of wind power, the slack bus has to absorb an excess power up to 2000MW. This corresponds to the case when all thermal CG units operate at their minimum power output (10% of the unit capacity). In reality, this will correspond to power exports to neighboring systems.



(a) Distributions



(b) Box-plots

Figure 6.17: Slack bus power injection distributions and boxplots (10000-sample MCS).

Algorithm 6.1 CG unit dispatch algorithm.

For all samples i :

1. Calculate system net load $L_N(i) = \sum_{l=1}^{l_T} L_l(i) - \sum_{g=1}^{g_T} G_g(i)$
2. Calculate aggregate CG capacity without slack generator $C_T = \sum_{c=1}^{c_T} C_c^M$
3. Calculate percentage $w(i) = \frac{L_N(i)}{C_T}$
 - if $w(i) < 0.1$: $C_c(i) = 0.1 \cdot C_c^M$
 - elseif $w(i) > 1$: $C_c(i) = C_c^M$
 - else $C_c(i) = w(i) \cdot C_c^M$

Table 6.9: Mean value and standard deviation for the slack bus power injection distributions.

Wind Penetration:	0 MW	1500 MW	3000 MW	4500MW
Mean [MW]	126.73	55.97	35.06	-42.39
St.D. [MW]	198.12	97.80	88.61	307.06

6.3.7 System power flows

In the figures 6.18, 6.19, 6.20 and 6.21, the box-plots for the power flow distributions in the system lines are presented. In Fig. 6.22, the respective power flow distributions for some characteristic lines in the system are presented. These plots actually reveal the transition from a vertical to a horizontally-operated power system. The increase in wind power in the system leads to an increase in the variability of the system power flows; most of them become bi-directional (the power flow distributions extend to both the positive and negative axis).

In many cases the incorporation of wind power leads to higher reverse power flows than the direct ones. Characteristic cases are for example line 10-13 (Fig. 6.22b), line 6-11 (Fig. 6.22d), line 4-5 (Fig. 6.22e) and line 5-6 (Fig. 6.22f). In other cases the increase in wind power integration leads to an increase in the power flows in the line, as is the cases for line 14-15 (Fig. 6.22a) and line 6-11 (Fig. 6.22d).

Thus, the integration of stochastic generation in the power system leads to an increase in the variability of the power flows. Either a strong power system to facilitate a large-scale integration of SG units or system reinforcements are necessary. In order to assess these reinforcements, a stochastic methodology is necessary, since counterintuitive results may be obtained due to the meshed system structure and the correlation between the stochastic resources.

6.3.8 System losses

In Fig. 6.23, the distributions of the system losses in the 4 wind power penetration scenarios are presented, while in Table 6.10, the respective means and standard deviations

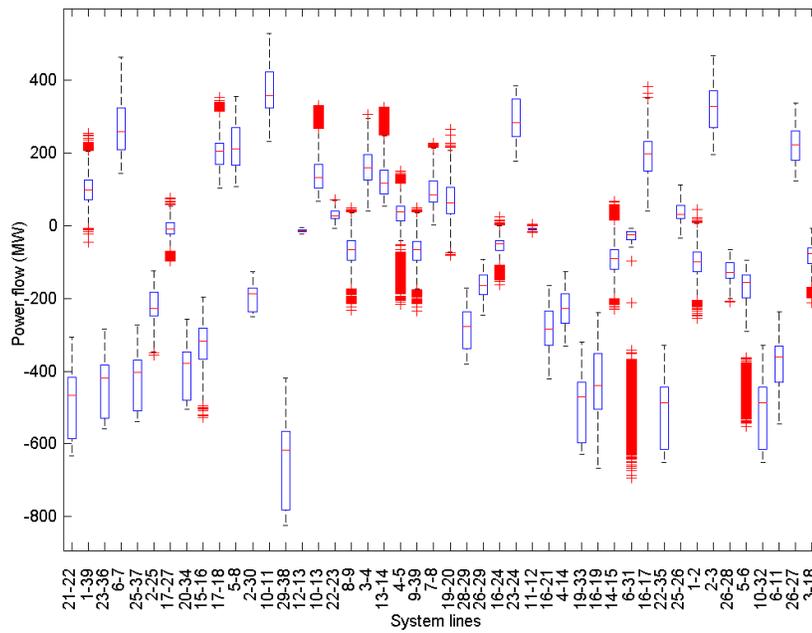


Figure 6.18: Box-plot for the power flows in the system lines in case of no wind power penetration (10000-sample MCS).

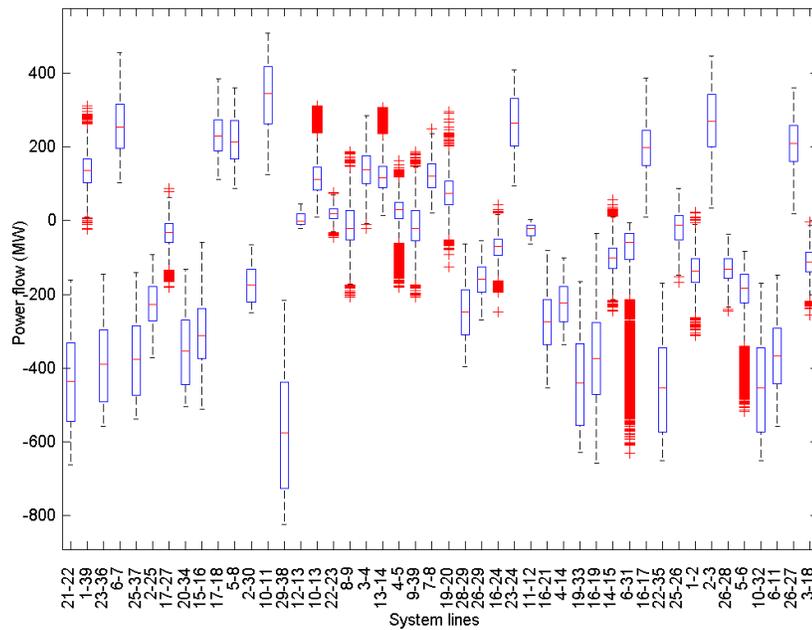


Figure 6.19: Box-plot for the power flows in the system lines in case of 25% wind power penetration (10000-sample MCS).

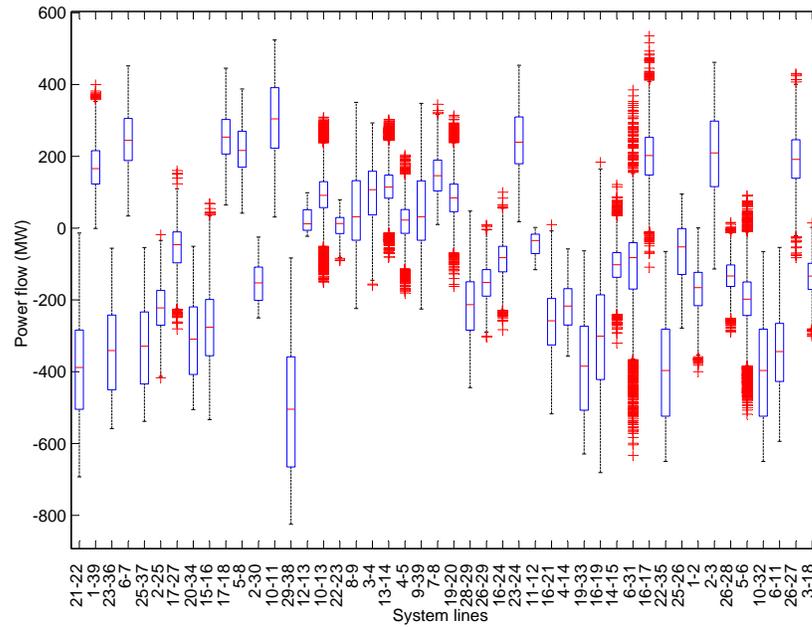


Figure 6.20: Box-plot for the power flows in the system lines in case of 50% wind power penetration (10000-sample MCS).

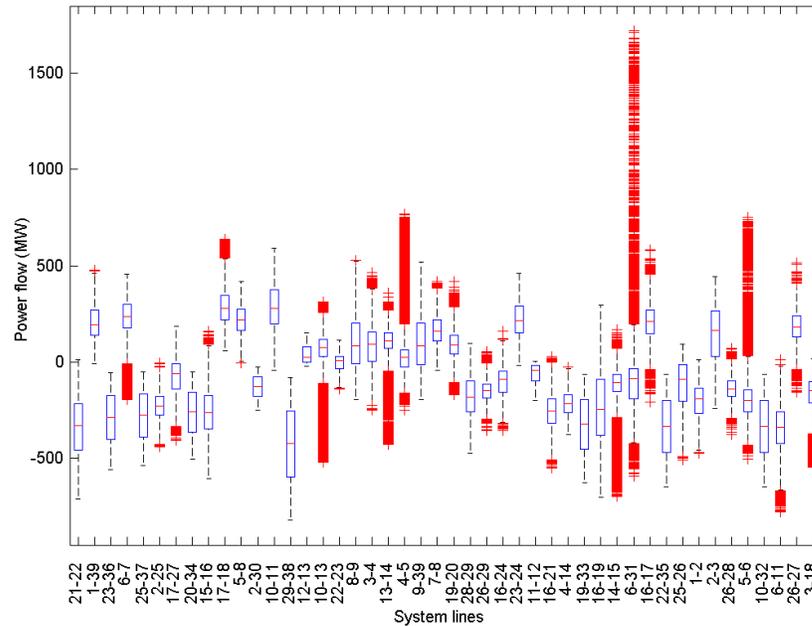


Figure 6.21: Box-plot for the power flows in the system lines in case of 75% wind power penetration (10000-sample MCS).

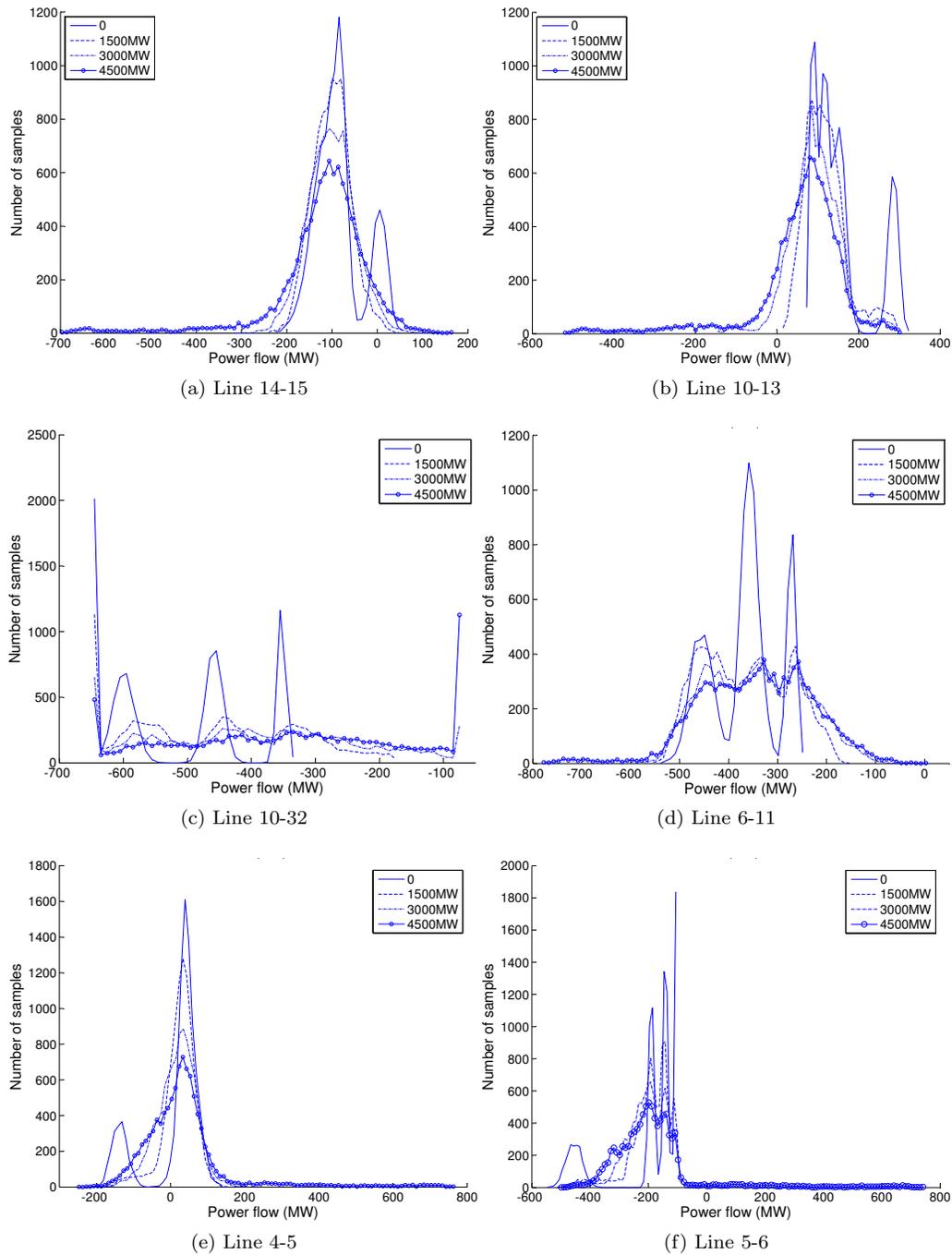


Figure 6.22: Some specific power flow distributions (10000-sample MCS).

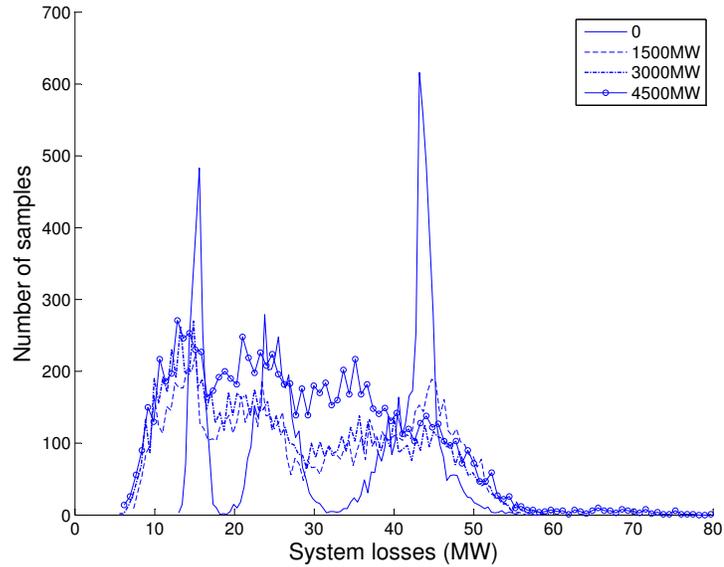


Figure 6.23: Distributions of system losses (10000-sample MCS).

Table 6.10: Mean value and standard deviation for the distributions of the system losses.

Wind Penetration:	0MW	1500MW	3000MW	4500MW
Mean [MW]	31.98	29.02	26.75	27.85
St.D.[MW]	11.68	13.10	12.72	12.72

are specified. We can see that the incorporation of wind power in the system leads to a change in the system losses. In a vertical power system, the system losses follow the time-dependent system loads, i.e. the system power flows correspond to the TF of system operation, presenting a concentration of probability at certain values (Fig. 6.23). In a horizontally-operated power system, these peaks are shaved due to the impact of the SG. A high SG penetration however, leads to longer tails in the distributions, due to the increase in the system power flows.

We can see from Table 6.10 that up to a penetration level of 50%, wind power integration leads to a decrease in the system losses. With higher penetration levels of wind power the system losses start to rise again, due to the increase in the system power flows.

6.4 Conclusions

In this chapter, the impact of stochastic generation in power systems has been investigated for both a distribution system and a bulk power system.

The incorporation of stochastic generation transforms the distribution system to an active network cluster, i.e. the power flows between the distribution and transmission system become bidirectional. Simulation results show that underestimating the impact

of dependence between the SG units and between the loads may lead to severe design fallacies. The upper stochastic bound has been proposed for the stochastic modeling of the system, as a prudent model that enables the inference of general results for the system design. The system analysis shows that the incorporation of stochastic generation leads to an increase of the variability in the system. The SG units can however be used to reduce the variability of the node voltages by the regulation of their reactive power output. The results show that the droop voltage control leads to a considerable improvement of the voltage quality of the system.

In the case of a bulk power system, the integration of stochastic generation leads to a horizontal operation of the system. The power flows become bidirectional and at high stochastic generation penetration levels, reverse power flows may exceed the direct ones. The methodology presented here offers a unifying approach for the modeling of the time-dependent stochasticity of the system loads and the non-time-dependent uncertainty of the stochastic generation. The results of such an analysis can be used for the quantification of the risk of overloading system lines and the assessment of necessary reinforcements.

Chapter 7

Conclusions and recommendations

This chapter recapitulates the main results obtained in the framework of this Ph.D. project by providing general conclusions and discussions on the key findings, followed by suggestions for possible extensions of the work reported in the dissertation.

7.1 Conclusions

Currently, concerns on environmental issues lead to an increasing implementation of environmentally-friendly power generation in power systems. Three main characteristics distinguish this type of generation from the traditional conventional generation (CG): the power plants are relatively small so that most units are connected to the distribution systems, the utilization of non-dispatchable (stochastic) prime energy movers, and the utilization of new technologies for power generation instead of the synchronous generator paradigm. This ubiquitous uncontrollable power infeed (*stochastic generation - SG*) leads to a radical change of the vertical structure of the system. In particular, the power does not only flow 'vertically', i.e. from the higher to the lower voltage levels of the system, but also 'horizontally', i.e. from one distribution system to another or from a generator to a load within the same distribution system. The investigation of the power flows in such a horizontally-operated power system is a key issue for the system operational planning and design. For this, a new modeling approach that permits the incorporation of generation uncertainty in the system analysis is necessary. The main goals set in this Ph.D. project are the development of new tools for the uncertainty analysis of the power system and the investigation of the impact of the horizontal operation of the power system due to the large-scale integration of stochastic generation.

The methods presented in literature so far for the modeling of stochasticity in power system studies focus mainly on the modeling of the system load uncertainty. This uncertainty corresponds to a time-dependent stochasticity, i.e. by conditionalizing the analysis in time, the load uncertainty is limited in a small region around the time-conditional mean. The assumptions of independence and normality, used for the modeling of time-

conditioned loads, lead to fallacies when used for the modeling of SG, due to its non time-dependent stochasticity. In this thesis, a new approach is introduced (the *power system steady-state uncertainty analysis*), that is not limited by the type of distribution and can represent complex stochastic interdependencies between the system inputs. The problem definition leads to a multivariate uncertainty analysis problem and a Monte-Carlo (MCS) approach is proposed for the solution, based on the splitting of the modeling procedure in two basic components, i.e. the modeling of the one-dimensional marginal distributions and the modeling of the multi-dimensional stochastic dependence structure. The marginals represent the output spectrum of each stochastic input, while the stochastic dependence structure determines the mutual interaction between the stochastic inputs. While the modeling of stochastic dependence is underestimated in the related literature by assuming independence, it turns out to be the main modeling issue for SG since it determines the variability of the aggregate output of the SG units.

We devote the 3rd chapter to the discussion of models of stochastic dependence and we clarify the problems of modeling dependent resources by comparing examples of system loads (normal distributions) and wind turbine generator (WTG) units. We show that by assuming independence the impact of stochasticity in the system is underestimated, since the aggregate is approaching a normal distribution due to the applicability of the *central limit theorem*. Thus, by assuming independence one predefines the outcome of the analysis to normality. Appropriate measures and models of stochastic dependence is a source of confusion in the related literature. Although linear (product moment) correlation is the canonical measure in the world of multivariate normal and more generally for spherical and elliptical distributions, when we leave the world of normals, it becomes just a measure of linear dependence, that may be misleading for the representation of dependence. We show that *rank correlation* (ρ_r), i.e. the product moment correlation of the ranks of the r.v., is a more adequate measure of dependence since it captures non-linear, monotonic relationships between r.v. The transformation of the marginal distributions into ranks is performed by the cdf (cumulative density function) transformation. This transformation is the cornerstone of the proposed stochastic modeling, offering a decoupling between the marginal distributions and the stochastic dependence structure, so that the dependence structure can be defined on these ranks by specific functions, the *copula functions*. By using these functions, one can simulate two r.v. that are correlated according to rank correlation ρ_r , by first simulating a copula and then transforming the obtained ranks into the respective marginals.

The 4th chapter presents the techniques for multidimensional dependence modeling. The *Joint Normal Transform Methodology* is presented as the main technique for the modeling of correlated inputs when related measurements are available. In particular, one can calculate the product moment/rank correlation matrix from the time-series data describing the stochastic inputs in the system and use a multidimensional normal copula for the dependence modeling. The main restriction to the applicability of the methodology is that the correlation matrix has to be *positive semi-definite*. We show that in many situations this is not the case and therefore present methods to repair the violations of positive semi-definiteness. In problems that involve high uncertainty, and where it is not possible to define all mutual correlations between the system stochastic inputs, we need sampling routines that can deal with incomplete specifications and give *exactly* what we specify, up to sampling error, with minimal additional information. These are the

Copula-Trees and *Copula-Vines* methods.

The high dimensionality of power system studies may however render the applicability of the proposed techniques, mainly due to lack of data or model complexity. Therefore, in chapter 5, we propose model reduction techniques by simplifying parts of the stochastic model. These approximations are called *stochastic plants* and are defined as the comonotonic versions of vectors of stochastic inputs. It is shown that these simple models bound the uncertainty of the stochastic model, i.e. they provide the worst-case scenario for the aggregate stochasticity of the vector. By applying these approximations, a problem involving the definition of the dependence structure between a large number of r.v. is decomposed into the modeling of the interdependencies between a manageable number of stochastic plants. The investigation of the properties of a cluster of WTGs shows that the sensitivity of the aggregate power output to the dependence structure is high in contrast with the sensitivity to the choice of marginal wind speed distributions. Thus, by using comonotonicity a good approximation of reality is obtained, even in cases where the marginals are not modeled in detail.

The developed methodologies are applied in chapter 6. We investigate the transformation of a distribution system into an active network cluster and the vertical to horizontal transformation of a bulk power system due to the integration of SG. By the application of the proposed techniques, we show how the time-dependent stochasticity of the system loads and the non time-dependent stochasticity of SG can be incorporated under a unified approach for the system analysis. The results of this study are:

- *Distribution system*: the upper stochastic bound is proposed as an appropriate dependence model of a distribution system, that offers prudent system design in terms of safety and generality. We analyze the transition from a passive to an active distribution network structure due to the integration of wind power and show that the anticipated reduction of system losses depends on the penetration level, i.e. when the SG penetration level exceeds a limit, the system losses start to increase again due to reverse power flows.
- *Voltage control*: we analyze the role of SG units in the reactive power support of the system by performing voltage control. It is shown that the reactive power capacity of each unit varies in time, in an opposite fashion with the stochastic active power generation. In this way, a distributed voltage control is achieved. The voltage control analysis of the system is performed by the incorporation of an iterative procedure in a MCS. Different voltage control algorithms were tested and the results show that the droop control strategy is the most suited for the voltage control by WTGs.
- *Bulk power system*: the transition from a vertical to a horizontally-operated power system due to a large-scale integration of wind power is analyzed. We show how the rise in wind power penetration leads to an increase in the variability of the system power flows. The results of the analysis quantify the risk of overloading the system lines and can be used for the assessment of system reinforcements.

7.2 Recommendations for future work

The work presented in this thesis sets the theoretical background for the analysis of the impact of stochastic generation in power systems. The recommendations for future work involve possible applications and extensions of this work.

Wind power integration in the interconnected European power system

A main application is the study of the wind power integration in a large system. As is known, the penetration of wind power in the European system is steadily increasing, with many countries being engaged in large-scale projects. The correlation between the wind resources in Europe necessitates the use of the techniques developed in this thesis for the planning and risk analysis of the interconnected European power system. Such a large-scale project should involve analysis of load and wind data from different countries and quantification of the risk of overloading the system lines.

Forecasting

In chapter 6, the integration of wind power in a bulk power system was investigated. As mentioned, in the presented approach the system dispatch has been set to follow the system net load, which corresponds to perfect forecast, i.e. zero forecast error. When the forecast in the system is not perfect, an increase in the variability of the system power flows should be expected, due to the engagement of units possibly other than expected to support the imbalances. Based on real data and using the techniques presented in the thesis, the forecast error may be sampled and incorporated in the analysis. This will provide a more realistic modeling of the system, where the impact of the uncertainty of the market actors is translated into the power system analysis.

Impact of large-scale storage

In this thesis, a memory-less power system is considered, i.e. the system state on each moment does not depend on the previous state of the system. This corresponds to a power system without large-scale storage. In order to investigate the impact of storage, a time-series approach should be employed, that is consistent to the principles presented in this thesis. In particular, specific algorithms should be developed that permit the generation of spatially-correlated time-series that are consistent to given marginal distributions and present specific autocorrelation characteristics. The transformations presented in this thesis concerning the decoupling of marginals and dependence structure offer the basis for the treatment of this problem.

Reactive power pricing

As presented in chapter 6, the voltage control operation of SG units may be analyzed by the incorporation of an iterative algorithm in a stochastic simulation. This work can be extended to the analysis of different voltage control strategies, such as exponential control. Since this approach gives insight in the reactive power output of each SG unit, the approach can form the basis for reactive power pricing.

Stochastic stability studies

The focus of this thesis is on the steady-state operation of the system. A direct extension of the work is the application of the obtained results for voltage/transient stability analysis of systems with a high penetration of SG.

Appendix A

Steady-State System Model

The steady-state or *load flow* analysis is the fundamental study for power systems analysis. The exact formulation of this problem concerns the determination of real and reactive power flows in each line (branch) of the power system [46]. The data used are the active and reactive power consumed by each load, as well as the real power generation and voltage magnitudes at the CG buses. The general steady-state model formulation can be divided into 2 parts:

1. *Determination of the State Vector of the system*: the magnitude and phase angle of the unknown bus voltages in the system are computed.
2. *Determination of the power flows on each line of the system*: the state vector is used in this step for the determination of the active and reactive power flows in every branch.

The system is modeled by a set of N buses (nodes), which are interconnected by transmission lines (branches). The transmission links are represented by their nominal π -equivalent circuits. Thus, for each line, numerical values for the series impedance Z and the total line-charging admittance Y are necessary for the determination of the elements in the $N \times N$ bus admittance matrix whose typical element is:

$$Y_{ij} = |Y_{ij}| \angle \theta_{ij} = |Y_{ij}| \cos \theta_{ij} + j |Y_{ij}| \sin \theta_{ij} = G_{ij} + jB_{ij} \quad (\text{A.1})$$

The admittance matrix has the following well-known properties:

- It is symmetric,
- Y_{ii} , the ii -th element (diagonal element), is equal to the sum of all admittances connected to the i -th node,
- Y_{ij} , the ij -th element of Y (non-diagonal element) is equal to the negative of the admittance in between nodes i and j .

The voltage at a typical bus i is given by:

$$V_i = |V_i| \angle \delta_i = |V_i| (\cos \delta_i + j \sin \delta_i) = G_{ij} + jB_{ij} \quad (\text{A.2})$$

Thus, the net current injected into the network at bus i is given by:

$$I_i = Y_{i1}V_1 + Y_{i2}V_2 + \cdots + Y_{iN}V_N = \sum_{n=1}^N Y_{in}V_n \quad (\text{A.3})$$

Let P_i and Q_i denote the net real and reactive power entering the network at bus i . Then:

$$P_i - jQ_i = V_i^* \sum_{n=1}^N Y_{in}V_n = \sum_{n=1}^N |Y_{in}V_iV_n| \angle(\theta_{in} - \delta_{in}) \quad (\text{A.4})$$

So,

$$\begin{aligned} P_i &= \sum_{n=1}^N |Y_{in}V_iV_n| \cos(\theta_{in} - \delta_{in}) \\ Q_i &= -\sum_{n=1}^N |Y_{in}V_iV_n| \sin(\theta_{in} - \delta_{in}) \end{aligned} \quad (\text{A.5})$$

or,

$$\begin{aligned} P_i &= |V_i| \sum_{n=1}^N |V_n| [G_{in} \cos \delta_{in} + B_{in} \sin \delta_{in}] \\ Q_i &= |V_i| \sum_{n=1}^N |V_n| [G_{in} \sin \delta_{in} - B_{in} \cos \delta_{in}] \end{aligned} \quad (\text{A.6})$$

where $\delta_{in} = \delta_i - \delta_n$.

The above equations constitute the polar form of the ac load flow equations. It can be seen, that there are four potentially unknown quantities associated with each bus i : P_i , Q_i , δ_i and $|V_i|$. For the solution of this problem, the general formulation is to identify three types of buses in the system. At each bus, two of the four quantities are specified whereas the remaining two are calculated. The formulation is the following:

1. *Load buses ($P - Q$ nodes)*: in these buses loads are connected, so P_i and Q_i are known from historical record, load forecast or measurements; δ_i and $|V_i|$ are the unknown quantities.
2. *Voltage controlled buses (generator buses, $P - |V|$ nodes)*: at each bus where a CG is connected, the active power generation can be controlled by adjusting the prime mover, and the voltage magnitude can be controlled by adjusting the generator excitation. Therefore, P_i and $|V_i|$ are known and δ_i and Q_i are the unknown quantities.
3. *Slack bus*: the power injection at this bus is determined by the power balance equation of the system. In the formulation of the power flow problem, the power injection in this bus is not pre-specified or scheduled. After the power flow problem has been solved, the difference between the total power which is injected into the system at all other buses and the total output plus the losses, are assigned to the slack bus. The voltage angle of this bus serves as a reference for the angles of all other bus voltages. The other known quantity for this bus is the voltage magnitude $|V_i|$. Obviously, there is no requirement to include the power flow equations for the slack bus in the power flow problem.

With this formulation, the problem is transformed into a system of $(2N - N_g - 2)$ non-linear equations with $(2N - N_g - 2)$ state variables to be calculated, where N_g is the number of voltage-controlled buses in the system. Due to the non-linearity of these

equations, power-flow calculations usually employ iterative techniques such as *Gauss-Seidel* and *Newton-Raphson* procedures.

When solving large-scale power transmission systems, an alternative strategy for improving computational efficiency is the *Decoupled Power-Flow Method*. This approach is based on the following two observations:

1. Change in the voltage angle δ at a bus primarily affects the flow of real power P and leaves the flow of reactive power Q relatively unchanged.
2. Change in the voltage magnitude $|V|$ at a bus basically affects the flow of reactive power Q and leaves the flow of real power P relatively unchanged.

Finally, from the solution of the system of equations, the power flows in line in can be calculated as follows:

$$\begin{aligned} P_{in} &= G_{in}|V_i|^2 - |V_i||V_n|[G_{in} \cos \delta_{in} + B_{in} \sin \delta_{in}] \\ Q_{in} &= -B_{in}|V_i|^2 + |V_i||V_n|[B_{in} \cos \delta_{in} - G_{in} \sin \delta_{in}] \end{aligned} \quad (\text{A.7})$$

Appendix B

Repairing violations of positive semi-definiteness

B.1 Hypersphere Decomposition

Given a non-positive semi-definite correlation matrix \mathbf{C} , the hypersphere decomposition method proposes the construction of a positive semi-definite correlation matrix $\hat{\mathbf{C}}$:

$$\hat{\mathbf{C}} = \mathbf{B} \cdot \mathbf{B}^T \quad (\text{B.1})$$

that best matches the target matrix \mathbf{C} .

Therefore, we have to define the matrix \mathbf{B} . The best method for doing this is to view the elements of the row vectors of \mathbf{B} as coordinates lying on a unit hypersphere. For this, the key is to obtain the $n \times n$ coordinates b_{ij} from $n \times (n - 1)$ angular coordinates θ_{ij} , according to:

$$b_{ij} = \cos \theta_{ij} \cdot \prod_{k=1}^{j-1} \sin \theta_{ik} \text{ for } j = 1, \dots, n - 1 \text{ and } b_{ij} = \prod_{k=1}^{j-1} \sin \theta_{ik} \text{ for } j = n \quad (\text{B.2})$$

For an arbitrary set of angles $\{\theta_{ij}\}$, a matrix $\hat{\mathbf{C}}$ formed from \mathbf{B} as in equation B.1, satisfies all given constraints required of a correlation matrix by construction. In particular, thanks to the trigonometric relationship B.2 and to the requirement that the radius of the unit hypersphere should be equal to one, the main diagonal elements are guaranteed to be unity.

In general, matrix $\hat{\mathbf{C}}$ will bear no resemblance to the target matrix \mathbf{C} . However, after using the above transformation and after defining a suitable error measurement ϵ in the resulting approximate correlation matrix $\hat{\mathbf{C}}$,

$$\epsilon = \|\mathbf{C} - \hat{\mathbf{C}}\| \quad (\text{B.3})$$

one can use an optimization procedure over the angles θ_{ij} to find the best possible fit given the chosen error measure. Different choices of the measure ϵ may be used, according to

the portions of the target correlation matrix \mathbf{C} that have to be recovered with particularly high accuracy. The fundamental benefit of this method is that when the underlying space over which the optimization is carried out is expressed in terms of angle vectors describing coordinates in unit hypersphere, no further constraints have to be satisfied, which can be of substantial benefit to the numerical fitting procedure [75].

B.2 Spectral Decomposition

A matrix is positive semi-definite if all its eigenvalues are positive. According to the spectral decomposition method, the positive semi-definite matrix $\hat{\mathbf{C}}$ is constructed by replacing all the negative eigenvalues of the non positive semi-definite matrix \mathbf{C} by zero. Given the right-hand-side eigensystem \mathbf{S} of \mathbf{C} and its associated set of eigenvalues $\{\lambda_i\}$ such that:

$$\mathbf{C} \cdot \mathbf{S} = \mathbf{\Lambda} \cdot \mathbf{S}, \text{ where } \mathbf{\Lambda} = \text{diag}(\lambda_i), \quad (\text{B.4})$$

we may define $\mathbf{\Lambda}'$ as a version of $\mathbf{\Lambda}$ where all negative elements are set to zero. We may also define the non-zero elements of a diagonal scaling matrix \mathbf{T} with respect to the eigensystem \mathbf{S} by:

$$\mathbf{T} : t_i = \left[\sum_m s_{im}^2 \lambda'_m \right]^{-1} \quad (\text{B.5})$$

In this case, the following matrix \mathbf{B} can be constructed:

$$\mathbf{B}' := \mathbf{S} \cdot \sqrt{\mathbf{\Lambda}'} \quad (\text{B.6})$$

and

$$\mathbf{B} := \sqrt{\mathbf{T}} \cdot \mathbf{B}' = \sqrt{\mathbf{T}} \cdot \mathbf{S} \cdot \sqrt{\mathbf{\Lambda}'} \quad (\text{B.7})$$

By construction,

$$\hat{\mathbf{C}} := \mathbf{B} \cdot \mathbf{B}^T \quad (\text{B.8})$$

is both positive semi-definite and has unit diagonal elements.

Following this procedure, we obtain an acceptable correlation matrix that is intuitively 'similar' to the target one (the more so, the fewer eigenvalues which have to be set to zero). As mentioned in [75], it has been empirically observed that the results obtained by this method are very similar to the results obtained by the hypersphere decomposition. Anyway, the result of this method may be used either as an accurate approximation to the best solution or as the starting point for the optimization performed in the first method.

Appendix C

Simulation data

C.1 Correlation matrices for the example in section 4.2

The system rank correlation matrix \mathbf{R}_r^{WL} obtained by the respective measurements is the following:

$$\mathbf{R}_r^{\text{WL}} =$$

1.00	0.24	0.18	0.19	0.13	0.19	0.18	0.19	0.19	0.09	0.01	0.03	0.00	0.05	0.03	0.00
0.24	1.00	0.82	0.85	0.74	0.78	0.77	0.88	0.89	0.76	0.58	0.57	0.54	0.76	0.68	0.52
0.18	0.82	1.00	0.83	0.74	0.82	0.79	0.79	0.80	0.69	0.54	0.56	0.59	0.69	0.64	0.51
0.19	0.85	0.83	1.00	0.81	0.74	0.73	0.82	0.82	0.75	0.56	0.53	0.52	0.72	0.63	0.48
0.13	0.74	0.74	0.81	1.00	0.65	0.63	0.76	0.73	0.75	0.57	0.48	0.47	0.69	0.60	0.47
0.19	0.78	0.82	0.74	0.65	1.00	0.89	0.77	0.79	0.63	0.52	0.58	0.70	0.66	0.65	0.53
0.18	0.77	0.79	0.73	0.63	0.89	1.00	0.77	0.81	0.64	0.54	0.63	0.71	0.69	0.66	0.57
0.19	0.88	0.79	0.82	0.76	0.77	0.77	1.00	0.91	0.83	0.67	0.59	0.55	0.82	0.73	0.58
0.19	0.89	0.80	0.82	0.73	0.79	0.81	0.91	1.00	0.79	0.64	0.62	0.59	0.81	0.75	0.58
0.09	0.76	0.69	0.75	0.75	0.63	0.64	0.83	0.79	1.00	0.69	0.56	0.51	0.81	0.66	0.54
0.01	0.58	0.54	0.56	0.57	0.52	0.54	0.67	0.64	0.69	1.00	0.50	0.48	0.68	0.62	0.70
0.03	0.57	0.56	0.53	0.48	0.58	0.63	0.59	0.62	0.56	0.50	1.00	0.56	0.64	0.58	0.48
0.00	0.54	0.59	0.52	0.47	0.70	0.71	0.55	0.59	0.51	0.48	0.56	1.00	0.60	0.61	0.50
0.05	0.76	0.69	0.72	0.69	0.66	0.69	0.82	0.81	0.81	0.68	0.64	0.60	1.00	0.72	0.54
0.03	0.68	0.64	0.63	0.60	0.65	0.66	0.73	0.75	0.66	0.62	0.58	0.61	0.72	1.00	0.62
0.00	0.52	0.51	0.48	0.47	0.53	0.57	0.58	0.58	0.54	0.70	0.48	0.50	0.54	0.62	1.00

The first line and column of the matrix correspond to the system load, while the other lines/columns correspond to the respective wind generation sites, in increasing order.

The rank correlations obtained by experts are presented in the following table:

r.v.	L	W_1	W_2	W_3	W_4	W_5	W_6	W_7
ρ_r	0	0.4	0.4	0.4	0.35	0.7	0.8	0.45
r.v.	W_8	W_9	W_{10}	W_{11}	W_{12}	W_{13}	W_{14}	W_{15}
ρ_r	0.5	0.55	0.7	0.8	0.9	0.7	0.65	0.85

C.2 Weibull parameters for wind speed distributions in section 5.4

The Weibull parameters β and η , used for the simulation in section 5.4.1 are:

	1	2	3	4	5	6	7	8	9	10
β	1.79	2.53	2.33	1.74	2.14	2.16	2.00	2.32	2.51	1.71
η	7.40	9.54	7.54	9.64	7.07	9.03	7.70	6.63	7.18	9.83
	11	12	13	14	15	16	17	18	19	20
β	2.08	2.46	2.63	2.21	2.55	1.92	1.62	1.65	1.90	2.16
η	6.67	8.90	7.95	6.24	8.31	8.42	8.98	6.74	9.65	8.86
	21	22	23	24	25	26	27	28	29	30
β	2.49	2.62	2.24	2.23	2.30	2.30	2.66	1.98	2.07	1.81
η	7.32	8.47	8.32	7.86	6.63	7.41	6.98	6.50	6.21	9.04
	31	32	33	34	35	36	37	38	39	40
β	1.84	1.66	2.36	2.45	2.30	1.75	2.74	2.42	1.85	2.77
η	6.33	6.73	9.92	8.36	9.55	6.31	9.81	6.53	7.79	8.60
	41	42	43	44	45	46	47	48	49	50
β	2.18	2.38	2.30	1.62	2.62	2.14	2.16	1.89	2.53	2.46
η	6.06	9.47	9.39	7.61	7.30	9.51	6.86	9.13	8.57	7.33

These values are used for the derivation of the wind stochastic plant aggregate power output distributions presented in Fig. 5.5 and were obtained randomly from the intervals $\beta \in [1.6, 2.8]$ and $\eta \in [6, 10]$.

C.3 Distribution test system data for the study case in section 6.1

Table C.1: System data for the 34-bus/38-line radial test distribution system.

System Lines Data						Load Data			Wind Data		
Node1	Node2	Length (m)	R (Ohm/km)	X (Ohm/km)	C (F/km)	Node	P (MW)	Q(Mvar)	Node	Weibull Param. A	Weibull Param. k
30	31	535	0.167	0.074	0.5	23	0.098	0.074	23	8.3	2
45	36	253	0.419	0.085	0.34	24	0.259	0.126	24	7.9	1.92
32	40	400	0.167	0.074	0.5	25	0.351	0.218	25	8	1.89
42	46	260	0.167	0.074	0.5	26	0.135	0.065	26	8.3	1.94
35	25	479	0.419	0.085	0.34	28	0.142	0.069	28	8.4	1.95
47	43	1860	0.167	0.074	0.5	29	0.048	0.036	29	8.5	2
35	24	293	0.419	0.085	0.34	30	0.045	0.022	30	8.6	1.8
46	29	290	0.167	0.074	0.5	31	0.156	0.075	31	7.8	1.9
42	32	647	0.167	0.074	0.5	32	0.127	0.062	32	8	2
41	23	800	0.167	0.074	0.5	34	0.114	0.055	34	8.4	1.92
39	25	112	0.419	0.085	0.34	35	0.142	0.069			
31	34	150	0.167	0.074	0.5	36	0.135	0.065	36	8	1.93
37	36	897	0.265	0.078	0.42	37	0.157	0.076	37	9	2
28	37	210	0.796	0.091	0.28	39	0.142	0.069	39	7.9	2.1
34	40	180	0.167	0.074	0.5	40	0.072	0.054	40	8	2.1
36	39	420	0.419	0.085	0.34	41	0.284	0.138	41	8.4	2.06
24	42	529	0.167	0.074	0.5	42	0.187	0.091	42	8.4	2.08
45	26	370	0.419	0.085	0.34	43	0.000	0.000	43	8.8	1.98
30	50	1600	0.167	0.074	0.5	45	0.120	0.058			
43	50	1600	0.167	0.074	0.5	46	0.116	0.087			
30	52	500	0.167	0.074	0.5	50	0.016	0.008			
23	54	210	0.167	0.074	0.5	52	0.078	0.038			
24	54	300	0.167	0.074	0.5	54	0.000	0.000			
55	28	333	0.796	0.091	0.28	55	0.116	0.056			
55	56	318	0.796	0.091	0.28	56	0.112	0.054			
56	57	298	0.796	0.091	0.28	57	0.109	0.053			
57	58	191	0.796	0.091	0.28	58	0.161	0.100			
61	50	999	0.167	0.074	0.5	61	0.016	0.008			
123	61	999	0.167	0.074	0.5	123	0.016	0.008			
125	124	999	0.167	0.074	0.5	124	0.016	0.008			
50	125	999	0.167	0.074	0.5	125	0.016	0.008			

Continued on next page

Table C.1 – continued from previous page

System Lines Data						Load Data			Wind Data		
Node1	Node2	Length (m)	R (Ohm/km)	X (Ohm/km)	C (F/km)	Node	P (MW)	Q (Mvar)	Node	Weibull Param. A	Weibull Param. k
44*	26	252	0.419	0.085	0.34	Generator					
53*	52	710	0.167	0.074	0.5	25	0.2	0.124			
59*	58	411	0.167	0.074	0.5						
43	49*	10	0.167	0.074	0.5						
41	42*	635	0.167	0.074	0.5						
46	27*	320	0.167	0.074	0.5						
Reactance coil (Ohm)											
2	47	-	0.01	0.3005	-						

*: Disconnected node

C.4 Bulk power system data for the study case in section 6.3

Table C.2: Bus Data of the New England 39 Bus Test System [65]

Bus (nr.)	Volts (pu)	Load (MW)	Load (MVar)	Gen (MW)
1	-	0.0	0.0	-
2	-	0.0	0.0	-
3	-	322.0	2.4	-
4	-	500.0	184.0	-
5	-	0.0	0.0	-
6	-	0.0	0.0	-
7	-	233.8	84.0	-
8	-	522.0	176.0	-
9	-	0.0	0.0	-
10	-	0.0	0.0	-
11	-	0.0	0.0	-
12	-	7.5	88.0	-
13	-	0.0	0.0	-
14	-	0.0	0.0	-
15	-	320.0	153.0	-

Continued on next page

Table C.2 – continued from previous page

Bus (nr.)	Volts (pu)	Load (MW)	Load (MVar)	Gen (MW)
16	-	329.0	32.3	-
17	-	0.0	0.0	-
18	-	158.0	30.0	-
19	-	0.0	0.0	-
20	-	628.0	103.0	-
21	-	274.0	115.0	-
22	-	0.0	0.0	-
23	-	247.5	84.6	-
24	-	308.6	-92.2	-
25	-	224.0	47.2	-
26	-	139.0	17.0	-
27	-	281.0	75.5	-
28	-	206.0	27.6	-
29	-	283.5	26.9	-
30	1.0475	0.0	0.0	250
31	0.982	9.2	4.6	-
32	0.9831	0.0	0.0	650
33	0.9972	0.0	0.0	632
34	1.0123	0.0	0.0	508
35	1.0493	0.0	0.0	650
36	1.0635	0.0	0.0	560
37	1.0278	0.0	0.0	540
38	1.0265	0.0	0.0	830
39	1.03	1104.0	250.0	1000

Table C.3: Line Data of the New England 39 Bus Test System [65]

Line Bus	Data Bus	Resistance (p.u.)	Reactance (p.u.)	Susceptance (p.u.)	Transformer Magnitude	Tap Angle
1	2	0.0035	0.0411	0.6987	0	0
1	39	0.0010	0.0250	0.7500	0	0
2	3	0.0013	0.0151	0.2572	0	0
2	25	0.0070	0.0086	0.1460	0	0
3	4	0.0013	0.0213	0.2214	0	0
3	18	0.0011	0.0133	0.2138	0	0
4	5	0.0008	0.0128	0.1342	0	0
4	14	0.0008	0.0129	0.1382	0	0
5	6	0.0002	0.0026	0.0434	0	0
5	8	0.0008	0.0112	0.1476	0	0
6	7	0.0006	0.0092	0.1130	0	0

Continued on next page

Table C.3 – continued from previous page

Line Bus	Data Bus	Resistance	Reactance	Susceptance	Transformer Magnitude	Tap Angle
6	11	0.0007	0.0082	0.1389	0	0
7	8	0.0004	0.0046	0.0780	0	0
8	9	0.0023	0.0363	0.3804	0	0
9	39	0.0010	0.0250	1.2000	0	0
10	11	0.0004	0.0043	0.0729	0	0
10	13	0.0004	0.0043	0.0729	0	0
13	14	0.0009	0.0101	0.1723	0	0
14	15	0.0018	0.0217	0.3660	0	0
15	16	0.0009	0.0094	0.1710	0	0
16	17	0.0007	0.0089	0.1342	0	0
16	19	0.0016	0.0195	0.3040	0	0
16	21	0.0008	0.0135	0.2548	0	0
16	24	0.0003	0.0059	0.0680	0	0
17	18	0.0007	0.0082	0.1319	0	0
17	27	0.0013	0.0173	0.3216	0	0
21	22	0.0008	0.0140	0.2565	0	0
22	23	0.0006	0.0096	0.1846	0	0
23	24	0.0022	0.0350	0.3610	0	0
25	26	0.0032	0.0323	0.5130	0	0
26	27	0.0014	0.0147	0.2396	0	0
26	28	0.0043	0.0474	0.7802	0	0
26	29	0.0057	0.0625	1.0290	0	0
28	29	0.0014	0.0151	0.2490	0	0
12	11	0.0016	0.0435	0.0000	1.006	0
12	13	0.0016	0.0435	0.0000	1.006	0
6	31	0.0000	0.0250	0.0000	1.07	0
10	32	0.0000	0.0200	0.0000	1.07	0
19	33	0.0007	0.0142	0.0000	1.07	0
20	34	0.0009	0.0180	0.0000	1.009	0
22	35	0.0000	0.0143	0.0000	1.025	0
23	36	0.0005	0.0272	0.0000	1	0
25	37	0.0006	0.0232	0.0000	1.025	0
2	30	0.0000	0.0181	0.0000	1.025	0
29	38	0.0008	0.0156	0.0000	1.025	0
19	20	0.0007	0.0138	0.0000	1.06	0

Bibliography

- [1] T. ACKERMANN, G. ANDERSSON, AND L. SÖDER, *Distributed generation: a definition*, Electric Power Systems Research, 57 (2001), pp. 195–204.
- [2] R. N. ALLAN AND M. R. G. AL-SHAKARCHI, *Linear dependence between nodal powers in probabilistic a.c. load flow*, Proceedings IEE, 124 (1977), pp. 529–534.
- [3] ———, *Probabilistic techniques in a.c. load flow analysis*, Proceedings IEE, 124 (1977), pp. 154–160.
- [4] R. N. ALLAN AND M. R. G. AL-SHAKARCHI, *Probabilistic a.c. load flow*, Proceedings IEE, 123 (1975), pp. 531–536.
- [5] R. N. ALLAN, B. BORKOWSKA, AND C. H. GRIGG, *Probabilistic analysis of power flows*, Proceedings IEE, 121 (1974), pp. 1551–1556.
- [6] R. N. ALLAN AND A. M. L. DA SILVA, *Probabilistic load flow using multilinearisations*, Proceedings IEE, 128, Pt. C (1981), pp. 280–286.
- [7] R. N. ALLAN, A. M. L. DA SILVA, AND R. C. BURCHETT, *Evaluation methods and accuracy in probabilistic load flow solutions*, IEEE Transactions on Power Apparatus & Systems, PAS-100 (1981), pp. 2539–2546.
- [8] R. N. ALLAN, C. H. GRIGG, AND M. R. G. AL-SHAKARCHI, *Numerical techniques in probabilistic load flow problems*, International Journal for Numerical Methods in Engineering, 10 (1976), pp. 853–860.
- [9] R. N. ALLAN, C. H. GRIGG, D. A. NEWEY, AND R. F. SIMMONS, *Probabilistic power-flow techniques extended and applied to operational decision making*, Proceedings IEE, 123 (1976), pp. 1317–1324.
- [10] F. ALVARADO, Y. HU, AND R. ADAPA, *Uncertainty in power system modeling and computation*, in IEEE International Conference on Systems, Man and Cybernetics, vol. 1, 1992, pp. 745–760.
- [11] G. J. ANDERS, *Probability Concepts in Electric Power Systems*, Wiley Interscience, 1990.
- [12] D. BARRY AND P. SMITH, *Analysis of the integration of wind generation into the irish system*, in 2005 IEEE St. Petersburg Powertech, June 27–30 2005.

-
- [13] R. BILLINTON AND R. N. ALLAN, *Reliability Evaluation of Engineering Systems*, Plenum Press, New York, second ed., 1992.
- [14] J. A. BLOOM, *Probabilistic production costing with dependent generating sources*, IEEE Transactions on Power Apparatus & Systems, PAS-104 (1985), pp. 2064–2070.
- [15] A.-M. BORBELY AND J. F. KREIDER, *Distributed Generation: The Power Paradigm for the New Millennium*, CRC Press, 2001.
- [16] B. BORKOWSKA, *Probabilistic load flow*, IEEE Transactions on Power Apparatus & Systems, PAS-93 (1974), pp. 752–759.
- [17] R. E. BROWN, *Electric Power Distribution Reliability*, Marcel Dekker, Inc., New York, 2002.
- [18] M. C. CARAMANIS, R. D. TABORS, K. S. NOCHUR, AND F. C. SCHWEPPE, *The introduction of non-dispatchable technologies as decision variables in long-term generation expansion models*, IEEE Transactions on Power Apparatus & Systems, PAS-101 (1982), pp. 2658–2666.
- [19] M. C. CARIO AND B. L. NELSON, *Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix*, tech. report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois, 1997.
- [20] G. D. CASTRONUOVO AND J. A. PECAS-LOPES, *Bounding active power generation of a wind-hydro power plant*, in 8th International Conference on Probability Methods Applied to Power Systems (PMAPS), Iowa State University, Ames, Iowa, USA, September 12-16 2004.
- [21] CIGRE, *Impact of increasing contribution of dispersed generation on the power system*. CIGRE Study Committee No 37, Working Group 23, Final Report, September 1998.
- [22] R. COOKE, *The anatomy of the squizzel. the role of operational definitions in representing uncertainty*, Reliability Engineering and System Safety, 85 (2004), pp. 313–319.
- [23] R. M. COOKE, *Experts in Uncertainty: Opinion and Subjective Probability in Science*, Oxford University Press, Inc., New York, 1991.
- [24] ———, *Markov and entropy properties of tree- and vine-dependent variables*, in Proceedings ASA Section on Bayesian Statistical Science, American Statistics Association, Alexandria, VA., 1997, pp. 166–175.
- [25] A. M. L. DA SILVA, R. N. ALLAN, S. M. SOARES, AND V. L. ARIENTI, *Probabilistic load flow considering network outages*, Proceedings of the IEE, 132 (1985), pp. 139–145.
- [26] A. M. L. DA SILVA AND V. L. ARIENTI, *Probabilistic load flow by multilinear simulation algorithm*, Proceedings of the IEE, C (1990), pp. 276–282.

- [27] A. M. L. DA SILVA, V. L. ARIENTI, AND R. N. ALLAN, *Probabilistic load flow considering dependence between input nodal powers*, IEEE Transactions on Power Apparatus & Systems, PAS-103 (1984), pp. 1524–1530.
- [28] A. M. L. DA SILVA, S. M. P. RIBEIRO, V. L. ARIENTI, R. N. ALLAN, AND M. B. D. C. FILHO, *Probabilistic load flow techniques applied to power system expansion planning*, IEEE Transactions on Power Systems, 5 (1990), pp. 1047–1053.
- [29] DANISH WIND INDUSTRY ASSOCIATION, *Windpower website*. <http://www.windpower.org>, 2006.
- [30] M. W. DAVIS, *Microturbines - an economic and reliability evaluation for commercial, residential, and remote load applications*, IEEE Transactions on Power Systems, 14(4) (1999), pp. 1556–1562.
- [31] DELFT & EINDHOVEN UNIVERSITIES OF TECHNOLOGY, *Intelligent power systems project website*. <http://www.intelligentpowersystems.nl>, 2003.
- [32] DELFT INSTITUTE OF APPLIED MATHEMATICS, DELFT UNIVERSITY OF TECHNOLOGY, *Unicorn - 'uncertainty analysis with correlations' software tool*. <http://dutiosc.twi.tudelft.nl/risk/>, 2006.
- [33] S. DEMARTA AND A. J. MCNEIL, *The t copula and related copulas*, International Statistical Review, 73(1) (2005), pp. 111–129.
- [34] M. DENUIT, J. DHAENE, AND C. RIBAS, *Does positive dependence between individual risks increase stop-loss premiums?*, Insurance: Mathematics and Economics, 28(3) (2001), pp. 305–308.
- [35] M. DENUIT, C. GENEST, AND E. MARCEAU, *Stochastic bounds on sums of dependent risks*, Insurance: Mathematics and Economics, 25 (1999), pp. 85–104.
- [36] J. DHAENE, M. DENUIT, M. J. GOOVAERTS, R. KAAS, AND D. VYNCKE, *The concept of comonotonicity in actuarial science and finance: Theory*, Insurance: Mathematics and Economics, 31(1) (2002), pp. 3–33.
- [37] J. F. DOPAZO, O. A. KLITIN, AND A. M. SASSON, *Stochastic load flow*, IEEE Transactions on Power Apparatus & Systems, PAS-94 (1975), pp. 299–309.
- [38] P. EMBRECHTS, A. MCNEIL, AND D. STRAUMANN, *Correlation and dependence in risk management: Properties and pitfalls*, Risk Management: Value at Risk and Beyond, (2002), pp. 176–223.
- [39] J. H. R. ENSLIN, *Interconnection of distributed power to the distribution network*, in IEEE Young Researchers Symposium, Delft, the Netherlands, March 18-19 2004.
- [40] M. FRÉCHET, *Sur les tableaux de corrélation dont les marges sont données*, Ann. Univ. Lyon, (1951), pp. Sect. A 9, 53–77.
- [41] S. GHOSH AND S. HENDERSON, *Properties of the nostra method in higher dimensions*, in Proceedings of the 2002 Winter Simulation Conference, 2002, pp. 263–269.

- [42] G. GROSS, N. V. GARAPIC, AND B. MCNUTT, *The mixture of normals approximation technique for equivalent load duration curves*, IEEE Transactions on Power Systems, 3 (1988), pp. 368–374.
- [43] N. D. HATZIARGYRIOU, T. S. KARAKATSANIS, AND M. PAPADOPOULOS, *Probabilistic load flow in distribution systems containing dispersed wind power generation*, IEEE Transactions on Power Systems, 8 (1993), pp. 159–165.
- [44] W. HOEFFDING, *Masstabinvariante korrelationstheorie*, Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin, (1940), pp. 5:181–233.
- [45] R. IMAN, J. HELTON, AND J. CAMPBELL, *An approach to sensitivity analysis of computer models*, Journal of Quality Technology, (1981).
- [46] J. GRAINGER AND W. D. JR. STEVENSON, *Power Systems Analysis*, Electrical Engineering, McGraw-Hill International Editions, 1994.
- [47] N. JENKINS, R. ALLAN, P. CROSSLEY, D. KIRSCHEN, AND G. STRBAC, *Embedded Generation*, no. 31 in Power and Energy, IEE, 2000.
- [48] H. JOE, *Multivariate Models and Dependence Concepts*, Chapman & Hall, London, 1997.
- [49] B. KLÖCKL, *Stochastic Storage in Energy Systems*, PhD thesis, ETH Zürich, Zürich, Switzerland, 2006.
- [50] B. KLÖCKL AND G. PAPAETHYMIU, *An effort to overcome the chronological modeling methods for energy storage devices*, in International Conference on Future Power Systems (FPS2005), November 16-18 2005.
- [51] D. KUROWICKA AND R. M. COOKE, *Uncertainty Analysis with High Dimensional Dependence Modelling*, Wiley series in Probability and Statistics, 2006.
- [52] Z. LANDSMAN AND A. TSANAKAS, *Stochastic ordering of bivariate elliptical distributions*, Statistics & Probability Letters, 76 (2006), pp. 488–494.
- [53] D. D. MARIE AND S. KOTZ, *Correlation and Dependence*, Imperial College Press, London, 2001.
- [54] A. P. S. MELIPOULOS, A. G. BAKIRTZIS, AND R. KOVACS, *Power system reliability evaluation using stochastic load flows*, IEEE Transactions on Power Apparatus & Systems, PAS-102 (1983), pp. 1084–1091.
- [55] A. P. S. MELIPOULOS, G. J. COKKINIDES, AND X. Y. CHAO, *A new probabilistic power flow analysis method*, IEEE Transactions on Power Systems, 5 (1990), pp. 182–190.
- [56] A. P. S. MELIPOULOS, S. KANG, G. J. COKKINIDES, AND R. DOUGAL, *Animation and visualization of spotprices via quadratized power flow analysis*, in Proceedings of the 36th Hawaii International Conference on System Sciences, January 2003, pp. 6–9.

- [57] T. J. E. MILLER, *Reactive Power Control in Electric Systems*, John Wiley & Sons, New York, 1982.
- [58] V. MIRANDA AND M. A. C. C. MATOS, *Distribution system planning with fuzzy models and techniques*, in CIRED, Electricity Distribution, 10th International Conference on, May 1989, pp. 472–476.
- [59] V. MIRANDA AND J. T. SARAIVA, *Fuzzy modelling of power system optimal load flow*, IEEE Transactions on Power Systems, 7 (1992), pp. 843–849.
- [60] O. MORALES, D. KUROWICKA, AND A. ROELEN, *Elicitation procedures for conditional and unconditional correlations*, appearing in Reliability Engineering and System Safety, (2007).
- [61] J. MORREN, *Grid support by power electronic converters of Distributed Generation units*, PhD thesis, Delft University of Technology, Delft, the Netherlands, 2006.
- [62] A. MULLER AND D. STOYAN, *Comparison Methods for Stochastic Models and Risks*, Wiley series in Probability and Statistics, 2002.
- [63] R. B. NELSEN, *An Introduction to Copulas*, Springer, New York, 1999.
- [64] K. NIELSEN, *Wave energy plants - the way forward*, in IEE Colloquium on Wave Power: An Engineering and Commercial Perspective (Digest No:1997/098), March 13 1997, pp. 7/1–7/5.
- [65] M. A. PAI, *Energy Function Analysis for Power System Stability*, Kluwer Academic Publishers, Boston, 1989.
- [66] G. PAPAETHYMIU, J. MORREN, P. H. SCHAVEMAKER, W. L. KLING, AND L. VAN DER SLUIS, *The role of power electronic converters as intelligent interfaces in distributed power systems: Stochastic steady state analysis*, in CIGRE 2005 Symposium "Power Systems with Dispersed Generation", Athens, Greece, April 17-20 2005.
- [67] G. PAPAETHYMIU, P. H. SCHAVEMAKER, AND L. VAN DER SLUIS, *Survey on the modelling of the uncertainty in 'distributed' power systems*, in IEEE Young Researchers Symposium, Delft, the Netherlands, March 18-19 2004.
- [68] G. PAPAETHYMIU, A. TSANAKAS, M. REZA, P. H. SCHAVEMAKER, AND L. VAN DER SLUIS, *Reliability assessment of hv/mv transformer-links for distributed power systems planning*, in 3rd IEE International Conference on Reliability of Transmission and Distribution Networks (RTDN 2005), London, UK, February 15-17 2005.
- [69] G. PAPAETHYMIU, A. TSANAKAS, P. H. SCHAVEMAKER, AND L. VAN DER SLUIS, *Design of 'distributed' energy systems based on probabilistic analysis*, in 8th International Conference on Probability Methods Applied to Power Systems (PMAPS), Iowa State University, Ames, Iowa, USA, September 12-16 2004.

- [70] ———, *Design of wind energy distributed power systems: Investigation of stochastic bounds using monte carlo simulation*, in 4th IASTED International Conference on Power and Energy Systems, Rhodes, Greece, June 28-30 2004.
- [71] A. PAPOULIS AND S. U. PILLAI, *Probability, Random Variables and Stochastic Processes*, Electrical and Computer Engineering, McGraw-Hill, fourth ed., 2002.
- [72] M. R. PATEL, *Wind and Solar Power Systems*, CRC Press LLC, 1999.
- [73] W. PRESS, B. FLANNERY, S. TEUKOLSKY, AND W. VETTERLING, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, New York, USA, second ed., 1992.
- [74] H. B. PÜTTGEN, P. R. MACGREGOR, AND F. C. LAMBERT, *Distributed generation: Semantic hype or the dawn of a new era?*, IEEE Power and Energy Magazine, 1 (2003), pp. 22–29.
- [75] R. REBONATO AND P. JÄCKEL, *The most general methodology to create a valid correlation matrix for risk management and option pricing purposes*, The Journal of Risk, 2(2) (1999).
- [76] M. REZA, *Stability analysis of transmission systems with high penetration of distributed generation*, PhD thesis, Delft University of Technology, Delft, the Netherlands, 2006.
- [77] M. REZA, P. H. SCHAVEMAKER, L. VAN DER SLUIS, AND W. L. KLING, *A research program on intelligent power systems: self controlling and self adapting power system equipped to deal with the structural changes in the generation and the way of consumption*, in Proceedings of the 17th International Conference & Exhibition On Electricity Distribution, CIRED, Barcelona, Spain, no. 70, SESSION 4, May 2003.
- [78] ROYAL NETHERLANDS METEOROLOGICAL INSTITUTE (KNMI), *HYDRA Project: Wind climate assessment of The Netherlands*. <http://www.knmi.nl/samenw/hydra>, 2006.
- [79] R. RUBINSTEIN, *Simulation and the Monte Carlo Method*, John Wiley & Sons, New York, 1981.
- [80] J. T. SARAIVA, N. FONSECA, AND M. A. MATOS, *Fuzzy power flow-an ac model addressing correlated data*, in 8th International Conference on Probability Methods Applied to Power Systems (PMAPS), Iowa State University, Ames, Iowa, USA, September 12-16 2004, pp. 519 – 524.
- [81] P. W. SAUER AND G. T. HEYDT, *A generalized stochastic power flow algorithm*, in 1978 IEEE Power Engineering Society Summer Meeting, July 1978.
- [82] M. T. SCHILLING, A. M. L. DA SILVA, R. BILLINTON, AND M. A. EL-KADY, *Bibliography on power system probabilistic analysis (1962-1988)*, IEEE Transactions on Power Systems, 5 (1990), pp. 1–11.

- [83] F. SCHWARTZ, R. PEGALLAPATI, AND M. SHAHIDEHPOUR, *Small hydro as green power*, in 2005 IEEE Power Engineering Society General Meeting, June 12-16 2005, pp. 1883–1890.
- [84] SENTERNOVEM, *Netherlands offshore wind energy website*. http://www.senternovem.nl/Offshore_Wind_Energy/index.asp, 2006.
- [85] S. SHETH AND M. SHAHIDEHPOUR, *Geothermal energy in power systems*, in 2004 IEEE Power Engineering Society General Meeting, June 6-10 2004, pp. 1972–1977.
- [86] ———, *Tidal energy in electric power systems*, in 2005 IEEE Power Engineering Society General Meeting, vol. 1, June 12-16 2005, pp. 630–635.
- [87] A. SKLAR, *Fonctions de répartition à n dimensions et leurs marges*, Publications de l’Institut de Statistique de L’Université de Paris, (1959), pp. 8:229–231.
- [88] J. G. SLOOTWEG, *Wind Power: Modelling and Impact on Power Systems Dynamics*, PhD thesis, Delft University of Technology, Delft, the Netherlands, 2003.
- [89] N. SRIRAM AND M. SHAHIDEHPOUR, *Renewable biomass energy*, in 2005 IEEE Power Engineering Society General Meeting, June 12-16 2005, pp. 1910–1915.
- [90] G. K. STEFOPOULOS, A. MELIPOULOS, AND G. J. COKKINIDES, *Probabilistic power flow with non-conforming electric loads*, in 8th International Conference on Probability Methods Applied to Power Systems (PMAPS), Iowa State University, Ames, Iowa, USA, September 12-16 2004.
- [91] ———, *Advanced probabilistic power flow methodology*, in 15th Power Systems Computation Conference, Liege, Belgium, August 22-26 2005.
- [92] I. TROEN AND E. L. PETERSEN, *European Wind Atlas*, Published for the CEC by RisøNational Laboratory, Roskilde, Denmark, 1989.
- [93] A. M. TUCKEY, D. J. PATTERSON, AND J. SWENSON, *A kinetic energy tidal generator in the northern territory-results*, in 23rd International Conference on Industrial Electronics, Control and Instrumentation IECON 97, vol. 2, November 9-14 1997, pp. 937–942.
- [94] J. G. VLACHOGIANNIS, *Fuzzy logic application in load flow studies*, IEE Proceedings on Generation, Transmission and Distribution, 148(1) (2005), pp. 34–40.
- [95] WIKIPEDIA, *Fuzzy logic*. http://en.wikipedia.org/wiki/Fuzzy_logic, 2006.
- [96] N. WOHLGEMUTH AND F. MISSFELDT, *The kyoto mechanisms and the prospects for renewable energy technologies*, Solar Energy, 69 (2000), pp. 305–314.
- [97] L. A. ZADEH, *Fuzzy sets as a basis for a theory of possibility*, International Journal of Fuzzy Sets & Systems, 1 (1978), pp. 3–28.
- [98] P. ZHANG AND S. T. LEE, *Probabilistic load flow computation using the method of combined cumulants and gram-charlier expansion*, IEEE Transactions on Power Systems, 19 (2004), pp. 676–682.

List of Abbreviations

cdf	: cumulative distribution function
CG	: Conventional Generation
CHP	: combined heat and power
DG	: Distributed Generation
DSM	: Deterministic System Model
HOPS	: horizontally-operated power system
HV	: high-voltage
JNT	: joint normal transform
LV	: low-voltage
MCS	: Monte-Carlo simulation
MV	: medium-voltage
pdf	: probability density function
PECs	: power electronic converters
PLF	: probabilistic load flow
PV	: photovoltaic
RES	: Renewable Energy Sources
r.v.	: random variable
SBM	: Stochastic Bounds Methodology
SG	: Stochastic Generation
SSM	: Stochastic System Model
SP	: stochastic plant
TF	: time-frame
WSP	: wind stochastic plant
WTG	: wind turbine generator

Summary in English

Integration of Stochastic Generation in Power Systems

Concerns on environmental issues lead to an increasing implementation of environmentally-friendly energy sources in electrical power systems. The uncontrollability of this *stochastic generation* is its main difference from the conventional power generation. This transition leads to a transformation towards a new 'horizontal' structure of the system, where the transmission and distribution networks operate as energy buses that interconnect the ubiquitous stochastic power generation and the distributed system loads. The ability of the system to absorb the generated power, i.e. to transport it to the system loads derives as the main issue for the incorporation of these uncontrollable energy sources. Since the power output from power plants that make use of such energy sources cannot be controlled by the system dispatcher, a new modeling approach is necessary, that permits the incorporation of generation uncertainty to the system analysis. In this research project, new tools for the uncertainty analysis of the power system are developed and the impacts of the power system horizontal operation due to the large-scale integration of stochastic generation is investigated.

In the related literature, the modeling of stochasticity in power system studies is mainly focused on the modeling of the system load uncertainty, which corresponds to a time-dependent stochasticity, i.e. for specific points in time, the load uncertainty is limited to a small region around the time-conditional mean. The assumptions of independence and normality, used in this case, lead to fallacies when used for the modeling of stochastic generation. On the one hand, independence leads to a severe underestimation of the system risk, corresponding to the case of minimum variability of the aggregate stochastic generation. On the other hand, in the case of normals joined by the multivariate normal distribution, the well-known linear correlation provides a sufficient representation of dependence and based on the knowledge of the linear correlation matrix, the stochastic modeling of correlated normals can be performed. However, when we leave the world of normals, it becomes just a measure of linear dependence, that may be misleading as representation of dependence. Therefore, in the case of stochastic generation, other measures of dependence and new modeling algorithms are necessary.

A Monte-Carlo approach is proposed for the solution of this problem, based on the splitting of the modeling procedure in two basic components, i.e. the modeling of the one-dimensional marginal distributions and the modeling of the multi-dimensional stochastic dependence structure. The marginals represent the output spectrum of each stochastic input, while the stochastic dependence structure determines the mutual interaction be-

tween the stochastic inputs and has a direct impact on their aggregate. This approach is referred in the related literature as *copula theory* and proposes two steps for the generation of a vector of r.v. correlated according to a given correlation matrix:

1. *Generate correlated uniforms*: these uniforms correspond to the respective ranks and bear all the information concerning the dependence structure in the system.
2. *Transform each uniform into the respective marginal*: this is performed by the inverse cdf transformation $F^{-1}(\cdot)$.

The first step is the most cumbersome part in the modeling procedure, corresponding to the multidimensional dependence modeling in the system. Two techniques are proposed: the *Joint Normal Transform Methodology*, based on the theory of the multivariate normal distribution, used in cases when related measurements are available and the *Copula-Trees* and *Copula-Vines* methods, used in problems that involve high uncertainty.

The high dimensionality of the power system problems may however render the applicability of the proposed techniques, mainly due to lack of data or model complexity. The thesis also proposes model reduction techniques in terms of simplifying approximations of parts of the stochastic model. These approximations are called *stochastic plants* and are defined as the comonotonic versions of vectors of stochastic inputs. It is shown that these simple models bound the uncertainty of the stochastic model, i.e. provide the worst-case scenario for the aggregate stochasticity of the vector. By applying these approximations, a problem involving the definition of the dependence structure between a large number of r.v. is decomposed into the modeling of the interdependencies between a manageable number of stochastic plants. The investigation of the properties of a cluster of wind turbine generators shows that the sensitivity of the aggregate power output to the dependence structure is high and relatively low to the choice of marginal wind speed distributions. Thus, using comonotonicity a good approximation of reality is obtained, even in cases when the marginals are not modeled with high accuracy.

Using the developed methodologies, the horizontal transformation of the power system due to the integration of stochastic generation is investigated. Both the time-dependent stochasticity of the system loads and the non time-dependent stochasticity of stochastic generation are incorporated to the system analysis under a unified approach. For distribution systems, a general dependence model is proposed, offering a prudent system design. Based on this model, an iterative algorithm is proposed for the analysis of the reactive power support of the system by the use of stochastic generators. The results for the case of wind turbine generators show that the droop control strategy is most suited for the voltage control of such units. In general, the incorporation of stochastic generation leads to an increase in the variability in the system power flows. The proposed analysis offers the quantification of this variability and provides the basis for a system design based on the assessment of the risk of violation of the safety limits. The results show that, up to a limit, the incorporation of stochastic generation leads to a reduction of the system losses. When this limit is exceeded, the reverse power flows lead to an increase in the system losses.

The uncertainty analysis techniques presented in this thesis are necessary for the analysis of the operation of systems with high penetration of stochastic energy sources. This approach forms the basis for a redefinition of the basic system studies for the analysis

of the future power systems. Generation expansion, system operational planning and stability studies have to be redefined under the prism of uncertainty. The work presented here sets the foundations for these next steps.

Samenvatting in het Nederlands

Integratie van Stochastische Opwekking in Elektriciteitssystemen

Zorgen over milieukwesties leiden tot een toenemende invoering van milieuvriendelijk energiebronnen in elektriciteitsvoorzieningsystemen. De onbeheersbaarheid van deze *stochastische opwekking* is het belangrijkste verschil in vergelijking met conventionele opwekking. Deze transitie leidt tot een transformatie naar een nieuwe "horizontale" structuur van het systeem, waar de transmissie- en distributienetten functioneren als energiepaden die de alomtegenwoordige stochastische opwekking en de verspreide systeem-belastingen verbinden. De mate waarin het systeem in staat is om het gegenereerde vermogen te accommoderen, d.w.z. om het naar de afnemers te transporteren, blijkt het belangrijkste probleem te zijn bij het integreren van deze oncontroleerbare energiebronnen. Aangezien de vermogensoutput van elektriciteitscentrales die gebruik maken van dergelijke energiebronnen niet kan worden gecontroleerd door de netwerk beheerder, is een nieuwe aanpak op het vlak van modelleren nodig, die toelaat om de onzekerheid van de opwekking te integreren in de analyse van het systeem. In dit onderzoeksproject worden nieuwe hulpmiddelen ontwikkeld voor de onzekerheidsanalyse in het elektriciteitsstelsel, en wordt de impact van een horizontale bedrijfsvoering met grootschalige integratie van stochastische opwekkers onderzocht.

In de verwante literatuur is het modelleren van stochasticiteit in systeemstudies hoofdzakelijk gericht op het modelleren van de onzekerheid van de systeembelasting, wat overeenkomt met een tijdsafhankelijke stochasticiteit. Dit betekent dat voor specifieke momenten in de tijd de onzekerheid van de belasting beperkt blijft tot een klein gebied rond het tijdsconditionele gemiddelde. De aannames van onafhankelijkheid en normaliteit, gebruikt in dit geval, leiden tot misinterpretaties wanneer ze worden toegepast op het modelleren van stochastische opwekking. Enerzijds leidt onafhankelijkheid tot een sterke onderschatting van het systeemrisico, corresponderend met het geval van minimum variabiliteit van de totale stochastische opwekking. Anderzijds, in het geval van normalen verbonden door de multivariabele normale verdeling, levert de welbekende lineaire correlatie een voldoende representatie van afhankelijkheid. Gebaseerd op de kennis van de lineaire correlatiematrix kan het stochastisch modelleren van gecorreleerde normalen uitgevoerd worden. Echter, als we de wereld van de normalen verlaten, wordt dit slechts

een maat voor lineaire afhankelijkheid, welke misleidend kan zijn als voorstelling van afhankelijkheid. Bijgevolg zijn nieuwe afhankelijkheidsindicatoren en modelleringsalgoritmes nodig als het gaat om stochastische opwekking.

Als oplossing voor dit probleem wordt een Monte-Carlo benadering voorgesteld, gebaseerd op het splitsen van het probleem in twee componenten, namelijk het modelleren van de eendimensionale marginale verdelingen enerzijds, en anderzijds het opstellen van de multidimensionale stochastische afhankelijkheidsstructuur. De marginale distributies geven het outputspectrum van elke stochastische output weer, terwijl de stochastische afhankelijkheidsstructuur de wederzijdse interactie tussen de stochastische inputs bepaalt en een direct impact op hun som heeft. Deze aanpak, die in de verwante literatuur *copula theorie* wordt genoemd, stelt twee stappen voor voor het genereren van een vector van random variabelen die gecorreleerd zijn volgens een gegeven correlatiematrix:

1. *Genereer de gecorreleerde uniforme verdelingen:* deze uniforme verdelingen komen overeen met de respectievelijke rangen en bevatten alle informatie over de afhankelijkheidsstructuur in het systeem.
2. *Transformeer elke uniforme verdeling naar de respectievelijke marginale verdeling:* dit wordt bekomen door de inverse cdf transformatie $F^{-1}(\cdot)$.

De eerste stap is de moeilijkste in de modelleringsprocedure, aangezien dit overeenkomt met het modelleren van de multidimensionale afhankelijkheid in het systeem. Twee technieken worden voorgesteld: de *Joint Normal Transform Methodologie*, gebaseerd op de theorie van de multivariabele normale verdeling, gebruikt in gevallen waar verwante metingen beschikbaar zijn en de *copula-trees* en *copula-vines* methodes, gebruikt bij problemen met een hoge onzekerheid.

De hoge dimensionaliteit van de problemen in het elektriciteitsvoorzieningsstelsel kan echter de toepassing van de genoemde technieken bemoeilijken, hoofdzakelijk door een gebrek aan gegevens of door de complexiteit van het model. De thesis stelt ook modelreductietechnieken voor als vereenvoudigde benaderingen van delen van het stochastische model. Deze benaderingen worden *stochastic plants* genoemd en worden gedefinieerd als de comonotone versies van stochastische inputs. Er wordt aangetoond dat deze vereenvoudigde modellen de onzekerheid van het stochastische model beperken, d.w.z. dat zij resulteren in het worst-case scenario voor de geaggregeerde stochastische output van de vector. Door het toepassen van deze benaderingen wordt een probleem aangaande het opstellen van de afhankelijkheidsstructuur tussen een groot aantal random variabelen gereduceerd tot het modelleren van de onderlinge afhankelijkheid tussen een hanteerbaar aantal stochastische plants. Onderzoek naar de eigenschappen van een groep windturbines toont aan dat de gevoeligheid van de geaggregeerde vermogensoutput hoog is met betrekking tot de afhankelijkheidsstructuur, maar relatief laag met betrekking tot de keuze van de marginale windsnelheidsverdelingen. Bijgevolg wordt een goede benadering van de werkelijkheid verkregen door gebruik te maken van comonotoniciteit, zelfs in gevallen waar de marginale verdelingen niet met hoge precisie worden gemodelleerd.

Gebruik makend van de ontwikkelde methodes wordt de horizontale transformatie van het elektriciteitsvoorzieningsstelsel tengevolge van de integratie van stochastische opwekking onderzocht. Zowel de tijdsafhankelijke stochastische output van de systeembelastingen als de tijdsafhankelijke stochastische output van stochastische opwekking worden geïntegreerd

in een globale aanpak voor systeemanalyse. Een algemeen afhankelijkheidsmodel wordt voorgesteld voor distributiesystemen, waarmee een voorzichtig systeemontwerp kan worden uitgevoerd. Gebaseerd op dit model wordt een iteratief algoritme voorgesteld voor de analyse van blindvermogenondersteuning door middel van stochastische opwekkers. De resultaten voor het geval van windturbines tonen dat de droop control strategie het best geschikt is voor spanningscontrole bij dergelijke eenheden. In het algemeen leidt de integratie van stochastische opwekking tot een toename van de variabiliteit in de vermogenstromen in het systeem. De voorgestelde analyse maakt kwantificatie van deze variabiliteit mogelijk en vormt de basis voor een systeemontwerp gebaseerd op een inschatting van het risico op overschrijding van de veiligheidsgrenzen. De resultaten tonen aan dat de integratie van stochastische opwekking tot op zekere hoogte leiden tot een afname van de systeemverliezen. Echter, vanaf een bepaald penetratieniveau doen omgekeerde vermogenstromen de verliezen weer toenemen.

De onzekerheidsanalysetechnieken die in deze thesis worden voorgesteld zijn noodzakelijk voor de analyse van de bedrijfsvoering van systemen met een hoog penetratieniveau van stochastische energiebronnen. Deze aanpak vormt de basis voor het herdefiniëren van fundamentele systeemstudies met het oog op de analyse van toekomstige energievoorzieningsystemen. Uitbreiding van opwekking, operationele planning en stabiliteitsstudies moeten geherdefinieerd worden in het kader van onzekerheidsanalyse. Het werk wat hier wordt voorgesteld legt de funderingen voor deze volgende stappen.

List of Publications

Journal Papers

1. G. PAPAETHYMIU, P. H. SCHAVEMAKER, L. VAN DER SLUIS, W. L. KLING, D. KUROWICKA, AND R. M. COOKE, *Integration of stochastic generation in power systems*, Electrical Power and Energy Systems, 28 (2006), pp. 655–667.
2. M. REZA, G. PAPAETHYMIU, AND P. H. SCHAVEMAKER, *An example of utilizing stochastic approach for investigating network constraints (congestions) on horizontally operated power systems with distributed generation*, Indonesian Electrical Engineering Journal (Jurnal Teknik Elektro), 6(2) (2006), pp. 79–86.
3. G. PAPAETHYMIU AND B. KLÖCKL, *MCMC for wind power simulation*, accepted for publication in IEEE Transactions on Energy Conversion.
4. G. PAPAETHYMIU, A. TSANAKAS, P. H. SCHAVEMAKER, AND L. VAN DER SLUIS, *Stochastic bounds for power system uncertainty analysis part i: Theory*, submitted to IEEE Transactions on Power Systems.
5. G. PAPAETHYMIU, A. TSANAKAS, P. H. SCHAVEMAKER, AND L. VAN DER SLUIS, *Stochastic bounds for power system uncertainty analysis part ii: Application on wind power*, submitted to IEEE Transactions on Power Systems.
6. B. KLÖCKL AND G. PAPAETHYMIU, *Local marginal based upsampling in MCMC models of stochastic generation*, submitted to Solar Energy.

Conference Papers

1. G. PAPAETHYMIU, J. VERBOOMEN, AND L. VAN DER SLUIS, *Estimation of power system variability due to wind power*, in PowerTech 2007 Conference, Lausanne, Switzerland, July 1-5 2007.
2. P. PINSON, G. PAPAETHYMIU, B. KLÖCKL, AND H. A. NIELSEN, *Generation of statistical scenarios of short-term wind power production*, in PowerTech 2007 Conference, Lausanne, Switzerland, July 1-5 2007.
3. G. PAPAETHYMIU, J. VERBOOMEN, P. H. SCHAVEMAKER, AND L. VAN DER SLUIS, *Impact of stochastic generation in power systems contingency analysis*, in

- 9th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Stockholm, Sweden, June 11-15 2006.
4. G. PAPAETHYMIU AND L. VAN DER SLUIS, *Modeling of stochastic generation in power systems: Principles and fallacies*, in 3rd IEEE Young Researchers Symposium in Electrical Power Engineering, Gent, Belgium, April 27-28 2006.
 5. M. REZA, G. PAPAETHYMIU, AND W. L. KLING, *Investigating transient stability impacts of a 'vertical-to-horizontal' transformation of power systems*, in 3rd IEEE Young Researchers Symposium in Electrical Power Engineering, Gent, Belgium, April 27-28 2006.
 6. G. PAPAETHYMIU, *Power Systems Uncertainty Analysis*, in Proceedings of the Doctoral Workshop on Future Energy Systems, ETH Zurich, March 2 2006.
 7. G. PAPAETHYMIU, A. TSANAKAS, D. KUROWICKA, P. H. SCHAVEMAKER, AND L. VAN DER SLUIS, *Probabilistic power flow methodology for the modeling of horizontally-operated power systems*, in International Conference on Future Power Systems (FPS2005), November 16-18 2005.
 8. B. KLÖCKL AND G. PAPAETHYMIU, *An effort to overcome the chronological modeling methods for energy storage devices*, in International Conference on Future Power Systems (FPS2005), November 16-18 2005.
 9. G. PAPAETHYMIU, P. H. SCHAVEMAKER, L. VAN DER SLUIS, W. L. KLING, D. KUROWICKA, AND R. M. COOKE, *Integration of stochastic generation in power systems*, in 15th Power Systems Computation Conference PSCC2005, Liege, Belgium, Invited Paper, August 22-26 2005.
 10. G. PAPAETHYMIU, A. TSANAKAS, M. REZA, P. H. SCHAVEMAKER, AND L. VAN DER SLUIS, *Stochastic modelling and analysis of horizontally-operated power systems with a high wind energy penetration*, in 2005 St. Petersburg PowerTech Conference, St. Petersburg, Russia, June 27-30 2005.
 11. G. PAPAETHYMIU, J. MORREN, P. H. SCHAVEMAKER, W. L. KLING, AND L. VAN DER SLUIS, *The role of power electronic converters as intelligent interfaces in distributed power systems: Stochastic steady state analysis*, in CIGRE 2005 Symposium "Power Systems with Dispersed Generation", Athens, Greece, April 17-20 2005.
 12. M. REZA, G. PAPAETHYMIU, P. H. SCHAVEMAKER, W. L. KLING, AND L. VAN DER SLUIS, *Stochastic transient stability analysis of power systems with distributed energy systems*, in CIGRE 2005 Symposium "Power Systems with Dispersed Generation", Athens, Greece, April 17-20 2005.
 13. G. PAPAETHYMIU, A. TSANAKAS, M. REZA, P. H. SCHAVEMAKER, AND L. VAN DER SLUIS, *Reliability assessment of HV/MV transformer-links for distributed power systems planning*, in 3rd IEE International Conference on Reliability of Transmission and Distribution Networks (RTDN 2005), London, UK, February 15-17 2005.

14. M. REZA, G. PAPAETHYMIU, P. H. SCHAVEMAKER, W. L. KLING, AND L. VAN DER SLUIS, *Investigating network constraints (congestions) of horizontally-operated power systems with stochastic distributed generation*, in 3rd IEE International Conference on Reliability of Transmission and Distribution Networks (RTDN 2005), London, UK, February 15-17 2005.
15. G. PAPAETHYMIU, A. TSANAKAS, P. H. SCHAVEMAKER, AND L. VAN DER SLUIS, *Design of 'distributed' energy systems based on probabilistic analysis*, in 8th International Conference on Probability Methods Applied to Power Systems (PMAPS), Iowa State University, Ames, Iowa, USA, September 12-16 2004.
16. G. PAPAETHYMIU, A. TSANAKAS, P. H. SCHAVEMAKER, AND L. VAN DER SLUIS, *Design of wind energy distributed power systems: Investigation of stochastic bounds using monte carlo simulation*, in 4th IASTED International Conference on Power and Energy Systems, Rhodes, Greece, June 28-30 2004.
17. G. PAPAETHYMIU, P. H. SCHAVEMAKER, AND L. VAN DER SLUIS, *Survey on the modelling of the uncertainty in 'distributed' power systems*, in IEEE Young Researchers Symposium, Delft, the Netherlands, March 18-19 2004.
18. M. REZA, G. PAPAETHYMIU, P. H. SCHAVEMAKER, AND L. VAN DER SLUIS, *Potential for transmission lines losses reduction in electrical power system operation with distributed generation*, in ISSM 2004, Delft, the Netherlands, October 9-10 2003.
19. A. D. TSANAKAS, G. I. PAPAETHYMIU, AND D. P. AGORIS, *Pollution flashover fault analysis and forecasting using neural networks*, in 39th CIGRE Session, Group 15: Materials for Electrotechnology, Preferential Subject 1, Paris, France, August 25-30 2002.
20. G. I. PAPAETHYMIU, A. D. TSANAKAS, D. P. AGORIS, AND M. KOVAIOS, *Influence of climatic factors on the fault occurrence due to insulators' flashover at networks of 15kv on islands*, in Session of the Greek National Committee of CIGRE, Athens, Greece, December 9-10 1999.
21. A. D. TSANAKAS, G. I. PAPAETHYMIU, AND D. P. AGORIS, *An approach to the effect of rainfall on the pollution performance of insulators in island networks*, in 11th International Symposium on High Voltage Engineering (ISH), London, UK, August 23-27 1999.

Acknowledgements

This thesis contains the results of my research performed at the Electrical Power Systems (EPS) Group of the Faculty of Electrical Engineering, Mathematics and Computer Science of the Delft University of Technology, the Netherlands. I would like to express my gratitude to all those who helped in direct or indirect ways to make this thesis possible.

First of all, I would like to thank my promotor Prof. ir. L. van der Sluis (Lou) for providing me with the opportunity to work in the Electrical Power Systems Group of Delft University of Technology. Many thanks go to my daily supervisor dr. ir. P.H. Schavemaker (Pieter) for his guidance and support. I would also like to thank dr. A. Tsanakas for his help and stimulating discussions from the first steps of my research, as well as dr. D. Kurowicka for being always there willing to help me with all my questions and problems.

The research has been performed within the framework of the research program 'Intelligent Power Systems' of the IOP-EMVT Program (*Innovatiegerichte Onderzoekprogramma's - ElektroMagnetische VermogensTechniek*). I would like to thank Prof. ir. M. Antal, the chairman of the IOP-EMVT program committee, ir. G. W. Boltje, the program coordinator, and the leaders and supervisors of the program Prof. W. L. Kling, Prof. L. van der Sluis, dr. ir. P. H. Schavemaker and dr. ir. J. M. A. Myrzik. My thanks are also to the other Ph.D. students of the project, Muhamad Reza, Johan Morren, Frans Provoost, Anton Ischenko, Andrej Jokic, Roald de Graaff, Jody Verboomen, Cai Rong and Sjef Cobben, for their collaboration and the nice time we spent together during these years.

Many thanks go to my officemate Muhamad Reza, for the friendship and the stimulating discussions around technical and non-technical subjects, my colleague Jody Verboomen for the nice discussions and for being the 'help-desk' in numerous technical problems I encountered, as well as the rest of the EPS-mates Bob Paap, Marjan Popov, Madeleine Gibescu, Ezra van Lanen, Arjan van Voorden, Johan Vijftigschild, Bart Ummels, Ralph Hendriks, Han Sloomweg, Tirza Drizi, Boukje Ypma, Barbara Slagter, Deborah Dongor, with whom I experienced a highly stimulating and pleasant working atmosphere during these years.

I wish to thank also the colleagues from the Vision of Future Energy Networks (VoFEN) project from the High Voltage and Power Systems Laboratories in ETH Zürich, especially Bernd Klöckl, for the fruitful collaboration.

I greatly acknowledge the financial support from the Greek State Scholarships Foundation (IKY).

I would like to specially express my gratitude to the person that initially recommended

me to come to TU Delft, Prof. D.P. Agoris. His advice, as all his advices in my life, was more than correct. His memory will always be with me.

Concerning the non-academic side of my life here, I have to thank a great number of people for their friendship and for the time we spent together. To avoid confusion and misunderstandings, they will remain nameless, but of course they know who they are.

I wish to thank my sister Vasiliki and my brother Spiros for their support during these years. Last but not least, I wish to thank my parents Ioannis and Eirini, for what can only be referred to as everything.

Georgios Papaefthymiou

Delft, May 2007

Biography

George Papaefthymiou was born on September 1, 1975, in Mesolongi, Greece. He obtained his diploma from the Department of Electrical Engineering & Computer Science of the University of Patras, Greece in 1999. After finishing his military service in 2001, he participated in the national examinations of the Greek State Scholarship's Foundation and received the scholarship on the field of "Renewable Energy Sources Technology". In 2001 and for 1 year he worked as GSM Cell Planning Engineer in Telestet Telecommunication Company, Athens, Greece. On December 2002 he joined the Electrical Power Systems research group at the faculty of Electrical Engineering, Mathematics & Computer Science of the Delft University of Technology, as a Ph.D. researcher within the framework of the research program 'Intelligent Power Systems'.