Power Grids Portfolio Optimization

by

Jan Kevin Pluut

A thesis submitted in partial fulfillment for the
degree of Master of Science

in the

Chair of Entrepreneurial Risks
Department of Management, Technology and Economics

March 2017

Abstract

Chair of Entrepreneurial Risks
Department of Management, Technology and Economics

Master of Science

by Jan Kevin Pluut

A lot of utility firms may benefit from paying per use instead of paying for ownership for the nodes in the power grid. Interested investing parties, however, expose themselves by buying multiple nodes to additional systemic risk. The goal of this dissertation is to compute the systemic risk component that should be assigned to price to ensure fairness. Using complex network analysis and simulation methods, modes of failure will be emulated and resulting penalty costs will be computed. Also, different investment strategies will be benchmarked to each other to advise potential asset managers optimal investing portfolio (acquiring clusters, decentralized, and upstream, downstream). No evidence was found for superior heuristically investment strategies. The 1/N portfolio was benchmarked to be very close to Markowitz efficiency as long as frequently failing nodes were identifiable.

Contents

Abstract i

List of Figures iv

Acronyms vii

1 Introduction 1
- 1.1 Setting the scene . . . 1
- 1.2 Case transaction . . . 4
- 1.3 Goal of this dissertation . . . 5

2 Models 7
- 2.1 Technical sub-problem . . . 8
  - 2.1.1 The distribution of both $U_{i,t}$ and $Z_{i,t}$ . . . 8
  - 2.1.2 A primer on power network theory . . . 9
  - 2.1.3 Understanding the (technical) risk of power assets . . . 11
  - 2.1.4 Picking the right grid model . . . 12
  - 2.1.5 Branching process . . . 13
  - 2.1.6 The OPA model . . . 15
    - 2.1.6.1 A primer on SOC . . . 15
    - 2.1.6.2 The OPA model explained . . . 16
- 2.2 Portfolio sub-problem . . . 19
  - 2.2.1 Markowitz efficient frontier . . . 20
  - 2.2.2 Other risk measure . . . 21
  - 2.2.3 Moments . . . 21

3 Methods 23
- 3.1 Step 1: Grid models . . . 24
- 3.2 Step 2 & 3: Data generation . . . 25
- 3.3 Step 4 & 5: Revenue stream construction . . . 26
  - 3.3.1 Leasing price . . . 27
  - 3.3.2 Penalty fee . . . 27
  - 3.3.3 Statistical analysis . . . 28
- 3.4 Step 6 & 7: Portfolio optimization . . . 29
- 3.5 Step 8 & 9: Heuristic testing . . . 30

4 Results 33

4.1 Sub-question 1: Data generation 33
4.1.1 Real-world data analysis 33
4.1.2 Data comparison 38
4.2 Sub-question 2: Revenue stream construction 44
4.3 Sub-question 3: Portfolio optimization 48
4.4 Sub-question 4: Heuristic testing 50

5 Conclusion **54**
5.1 Sub-question 1: Data generation 54
5.2 Sub-question 2: Revenue stream construction 55
5.3 Sub-question 3: Portfolio optimization 56
5.4 Sub-question 4: Heuristic testing 57
5.5 Final judgement on overall methodology 58

6 Open questions **59**

A MATLAB Code of OPA Model **61**

B Grid models **67**

Bibliography **74**

List of Figures

1.1 A sale-leaseback can be separated in two different sub-transactions: 1) The Lessee sells their power asset(s) to the Lessor for their market value. 2) The Lessee pays for the use of the assets (lease) 3

1.2 In a fixed yearly lease model, the cash flows are constant and thus less risky resembling more the characteristic trait of debt. In contrast, the pay-as-you-model fits more a dividend paying share. The additional risk is thus comparable to the inherent risk of equity. 4

1.3 The case transaction involves only nodes of the grid (substation, load or power generator). Furthermore, we assume the residual economic value of the asset to be the price of the asset. For simplicity, we use a straight line depreciation to calculate this. 5

2.1 This particular problem is essentially a portfolio problem in which the inputs require some technical simulations. The reason for this is that day-to-day data of operational metrics in a power grid are not publicly available for privacy and safety reasons 7

2.2 The same IEEE 24 RTS system is displayed here above: the left picture is the typical power engineering format, while the right picture is generated using the included 'force' algorithm of MATLAB. From a mathematical point of view, they are the same though. Note that the redundancies (parallel lines) are not displayed in the right picture. 9

2.3 According to , the consensus is that a blackout is the product of a triggering event (or events) on a critical system state. Accordingly, the modes of failure are the transfer mechanism that facilitate the blackout, i.e. the blue lines. 14

2.4 A variant of the OPA model that takes the N-1 policy into account[5] . . 18

3.1 The general outline in methodology 24

3.2 Real-world data of the energy throughput of a 100-150 kV substation. The timescale is in hours and total time series amount to one year. There is clearly a seasonality element, which is of course a consequence of the higher energy consumption in both summer and/or winter. In fact, the seasonality applies not only to the first moment (mean), but also the second: there is clearly a higher variance visible whenever there is a higher average consumption. 26

3.3 Each node in the graph can be either exogenous or endogenous power inflow. Mathematically, this is the same for power outflow 31

4.1 The original data (blue line) is filtered using a symmetric 1001-term MA filter (red line) to deseasonalize the data (yellow line). 34

4.2 Autocorrelation plot of the whole data sample using hours as lag-scale. . . 34
4.3 A power spectral analysis of the sample data . The other two spikes at 3,5 hours and 2 days are speculated to result from the short (breaks every 3-4 hours from working) and long business (breaks in the week) hours cycle. 35
4.4 A histogram of the detrended series after centering. 35
4.5 The survival functions are both plotted on a log-linear scale. There is a huge asymmetry between the tails, which can be explained by (un)planned maintenance and/or power outages that skew the distribution. 36
4.6 Both tails on a log-log scale. 36
4.7 A QQ-plot of the emperical distributuion against the corresponding normal distribution. 37
4.8 Galtson-Watson parameter over time (partitioning the data over 25 bins). 39
4.9 The Mean Time Between Failure calculated again using 25 bins of partitioned data. 39
4.10 Compressed, detrended and centered real-world data 40
4.11 A plot of the daily relative standard deviations against time. One can see that in Regime 1, the standard deviations themselves deviate a lot. 40
4.12 Scaled down operational standard deviations of OPA model. From both grid models (system 24 and system 39) only the operational data of node 1 are plotted 41
4.13 Standard deviations of each node of system 39 averaged out over last 5000 days of data. We used the same filters as for the real-world data 41
4.14 Relative standard deviation of each node of system 24. Averaged out over the last 5000 days. 42
4.15 The skewness of each node of system 24. Averaged out over the last 5000 days. 43
4.16 A survival function plot on log-linear scale of node 1 of system 39 (randomly chosen). Both the left and right tails are plotted 43
4.17 Survival analysis of empirical data of only Regime 1. 44
4.18 Accumulated cash flow of the first year using the last 50000 days of simulation sample. Furthermore, we use 2.8 time standard deviations as hit rate for a penalty fee 45
4.19 Spatial autocorrelation (Moran's I) using amount of links as distance unit. The lines are a 95% confidence interval. The graph does not seem to show convergence over the time 45
4.20 Synchronization matrices of system 24. The data is sampled over only the whole 50000 days. One can spot that there is a group of lines that fail often. 46
4.21 The synchronization matrices of system 39 show even more frequently failing lines (again sampled over the whole 50000 days). 47
4.22 A histogram of the daily returns of a 1/N portfolio consisting of all the nodes of system 39 47
4.23 A histogram of the daily returns of a 1/N portfolio excluding all the frequently failing nodes of system 39. 48
4.24 Markowitz efficient frontier of all system 39 nodes excluding the frequently failing nodes 49
4.25 Markowitz efficient frontier of all system 24 nodes excluding the frequently failing node 49

4.26 Standard deviation of each node of system 24 benchmarked against Streamness and Katz Centralilty 51
4.27 Standard deviation of each node of system 39 benchmarked against Streamness and Katz Centralilty 51
4.28 A sample of 20 equally-distanced portfolios on the efficient frontier. Starting from number 1 (minimum variance) to number 20 (maximum return) each portfolio of system 24 is benchmarked against their Streamness and Katz Centralilty score. 52
4.29 A sample of 20 equally-distanced portfolios on the efficient frontier. Starting from number 1 (minimum variance) to number 20 (maximum return) each portfolio of system 39 is benchmarked against their Streamness and Katz Centralilty score. 53

Acronyms

AAM Alternative Asset Management

AUM Assets Under Management

CAGR Compound Annual Growth Rate

CAPEX CAPital EXpenditure

CAPM Capital Asset Pricing Model

DC Direct Current

ENS Energy Not Supplied

FACTS Flexible Alternating Current Transmission System

IID Independent and Identically Distributed

IRR Internal Rate of Return

KISS Keep It Simple and Stupid

MTBF Mean Time Between Failure

OPA Oak Ridge-PSERC-Alaska

OPF Optimal Power Flow

pdf probability distribution function

PWC PricewaterhouseCoopers

SOC Self-Organized Criticality

SPV Special Purpose Vehicle

VaR Value at Risk

Chapter 1

Introduction

1.1 Setting the scene

The power sector is in constant change: decentralized power markets, renewable power generation and the emergence of super grids are just some of the many recent developments that have had (or are having) an enormous impact on how the power value chain is delivering energy to society as a whole. Along with the technical developments mentioned above, the power industry is exploring new ways to conduct business. A good example is the decreasing cost of renewables (e.g., solar panels and wind turbines), which allows individual agents to sell their energy directly to the grid and/or to an end customer. This relatively new development in power generation, also known as distributed power generation, presents an opportunity for new types of marked-price agreements on the customer side that more efficiently allocate power demand to the power supply[18][6].

It is not only power agents actively looking for innovative ways to capture value from recent technological developments: the financial world as well has been watching closely how the power sector is changing. In particular, so-called Alternative Asset Management (AAM) funds have kept a close eye on the power industrys transformation. These funds are characterized by investing in assets that are typically very illiquid and difficult to value. Usually, these include real estate, infrastructure, and other objects of significant value like art collections. The main focus, therefore, is to invest in real-world assets as opposed to typical (virtual) financial assets. One of the main drivers behind this new trend is the recent financial crisis in 2009. According to PricewaterhouseCoopers

(PWC)s latest report[1] on AAM, rapid developments in the global economic environment have pushed asset management to the forefront of social and economic change. An important part of this change [is] the need for increased and sustainable long-term investment returns. Accordingly, they forecast this particular segment of the capital markets to grow in Assets Under Management (AUM) with 8.1% Compound Annual Growth Rate (CAGR) until 2020.

In particular, infrastructure has been a special subset of this asset class[13] and is typically characterized by a long investment horizon, low correlation with the market (beta), and low overall volatility[1]. Although the Internal Rate of Return (IRR) is generally lower than conventional equity (e.g., stocks), it offers professional investors a way to diversify their otherwise non-diversifiable portfolios market risk. Belonging to this subclass, power assets have become a very attractive investment opportunity for AAM funds. Since the liberalization of power generation markets, power generation has been mainly financed by private investors looking for nice returns. Explicit ownership of the revenue streams and associated risks of such an asset may be the reason why most private equity began there. The same cannot be said of substations or other mid-stream devices like Flexible Alternating Current Transmission System (FACTS) equipment. After all, how would investors measure the income generation of owning a substation (an interconnected power asset that mainly functions as an electrical gear by transforming voltage levels)? Before liberalization, this question would not have made sense as the whole grid would have been owned by one single entity. This was mainly due to regulation and natural monopoly dynamics in order to spread the extremely high fixed cost over as many end customers as possible. However, since the liberalization of the markets in many countries the rules of the game have changed to allow external players to invest in the grid. For example, by means of ancillary services, it is possible for a private fund to own an element or function across the whole grid. These services are, in essence, just a way for a grid operator to outsource a certain function of the grid like reactive power compensation[15]. These are just some of the many ways that an interested private equity fund can invest in the grid. Another, more direct, way is by acquiring a certain portion of the network from the grid operator by means of a so-called sale leaseback transaction 1.1.

[1]Source: https://www.pwc.com/jg/en/publications/alternative-asset-management-2020.pdf

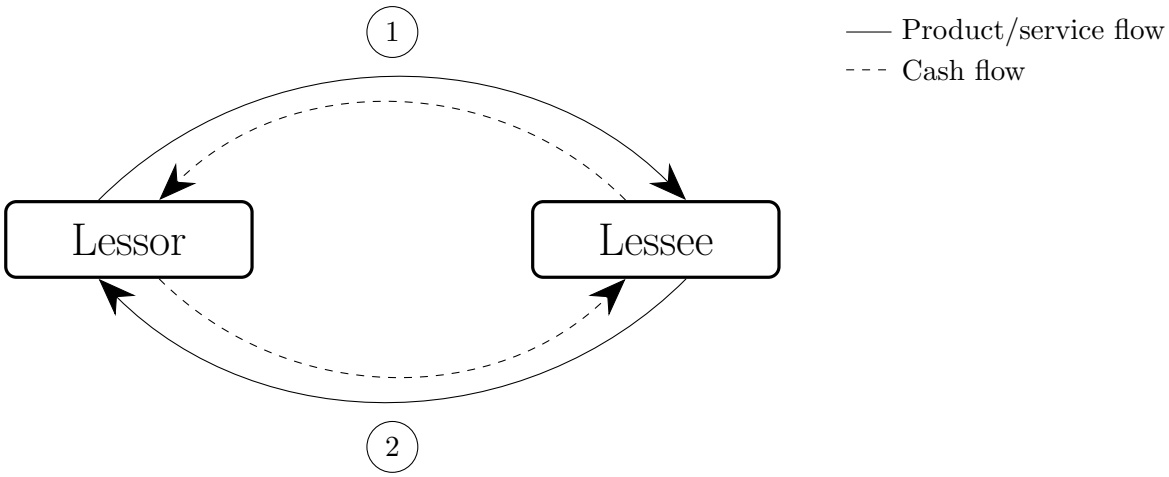

FIGURE 1.1: A sale-leaseback can be separated in two different sub-transactions: 1) The Lessee sells their power asset(s) to the Lessor for their market value. 2) The Lessee pays for the use of the assets (lease)

It should be noted that this type of financial transaction can be done in several ways. In particular, the payments of the second sub-transaction are something that can be tailored to the type of asset under consideration. One could, for example, set up a lease for a fixed number of years ($/year) as a type of bond Figure 1.2. This could be the way forward for assets where the value production is only indirectly measurable (like FACTS equipment[2]). Even better would be a pay-as-you go kind of model ($/MWh). Due to the inherent variability of energy throughput, the revenue stream resembles more a dividend-paying share (Figure 1.2). The latter would better fit an asset, like a substation, as the usage can be clearly measured. The modelling of the first kind of leaseback transactions should not be that difficult for the typical financial analyst. The second, though, requires already a significant degree of competence in electrical engineering to successfully model. However, if we ignore the interconnectedness of the asset, the asset valuation problem should also be doable from a technical point of view. It really gets interesting, however, when we extend the asset valuation problem to a portfolio problem, i.e., the lessor buys not just one, but multiple assets of the same grid. This adds a great layer of complexity to the original problem as the assets can now be expected to display some sort of interdependence . In fact, the interconnected nature of a grid guarantees some kind of non-linear interdependence as the probability

[2]Flexible AC Transmission System, better known as FACTS, is a power device that mainly functions as the regulator of the electrical power grid. In the most general case, it serves as a compensator for the reactive power generated by transmission of power.

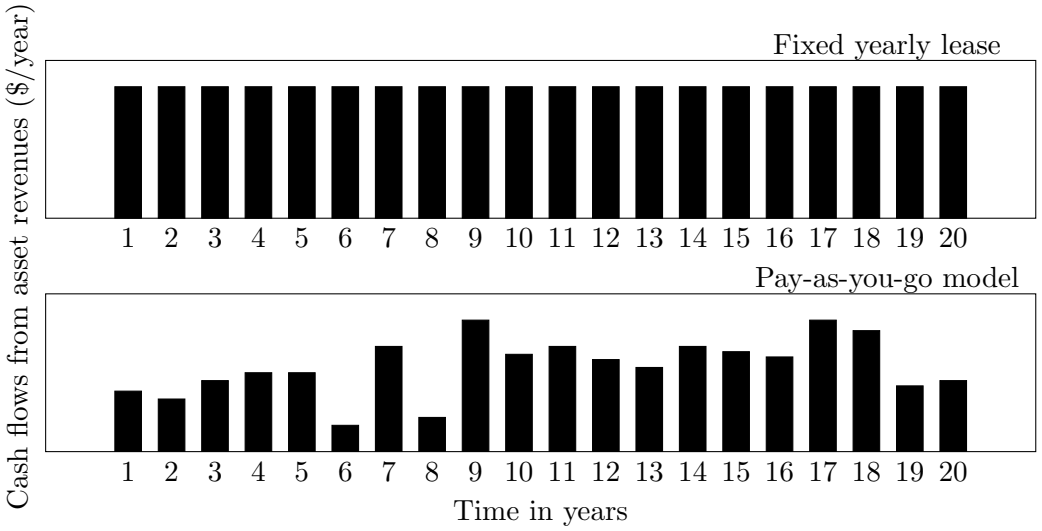

FIGURE 1.2: In a fixed yearly lease model, the cash flows are constant and thus less risky resembling more the characteristic trait of debt. In contrast, the pay-as-you-model fits more a dividend paying share. The additional risk is thus comparable to the inherent risk of equity.

distribution function (pdf) of blackout magnitude in several countries does seem to suggest [21][11][22][12].

1.2 Case transaction

For the sake of precision, we define here a case transaction that in our opinion offers the business and academic worlds the most value. A utility (the lessee), transfers their power asset (e.g., substation) off-balance sheet to a special purpose vehicle (the lessor). The lessor is obviously funded by an interested investor. Furthermore, we limit ourselves to the nodes (more on that in) in the grid: substations, loads and power generation assets. The transfer of ownership will be done in exchange for the residual economic value of the asset. Additionally, the lessee maintains the right to use the asset by paying a predetermined price to the lessor. This price is directly proportional to the amount of usage of the substation. It furthermore should include all the costs and risk premiums incurred by the lessor for owning the asset plus a healthy profit margin. A more graphical overview of this deal is shown in Figure 1.3.

Note that this type of financial transaction could also be done for new projects (so-called *greenfield* projects). In that case, the first sub-transaction would be omitted (Figure 1.3).

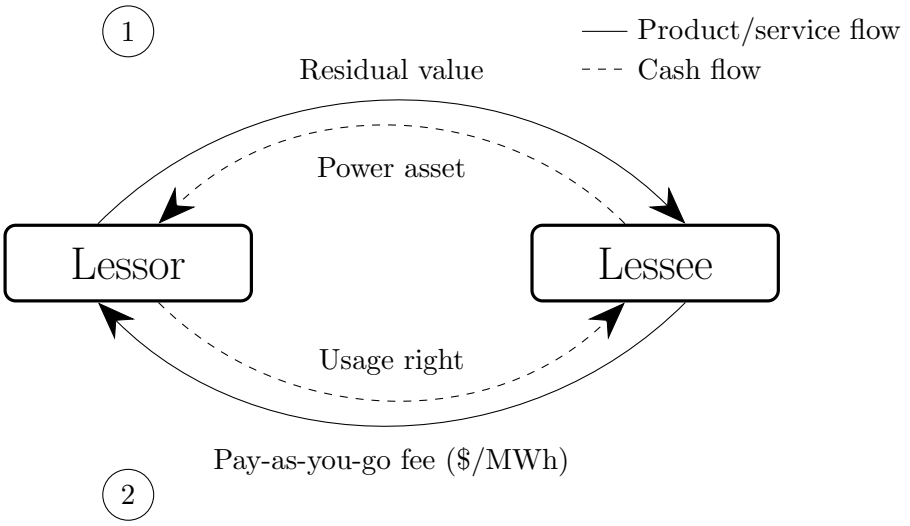

FIGURE 1.3: The case transaction involves only nodes of the grid (substation, load or power generator). Furthermore, we assume the residual economic value of the asset to be the price of the asset. For simplicity, we use a straight line depreciation to calculate this.

However, we will limit the scope of the thesis only to existing assets (so-called *brownfield* projects). This is done for a couple of reasons: first, greenfield deals are usually done one at a time, which limits the flexibility of tailoring a portfolio; second, from a modelling point of view, it is easier to model an existing grid than an evolving grid (in terms of additional nodes).

1.3 Goal of this dissertation

For theoretical reasons, we first define the main setting of the problem that is to be solved in this dissertation as follows: We imagine a hypothetical investor with budget B in a world where he can only invest in all N nodes of a given grid (and as well in some risk-free asset). Consequently, there exists N different lessors (Special Purpose Vehicles (SPVs)) and we assume that each lessor is infinitely divisible (i.e., it is possible to buy any specific amount of shares). Just as the Markowitz efficient portfolio theory prescribes a certain degree of inclusion or exclusion of particular assets by optimizing a portfolios variance (one of the many measures for risk) given an expected return, it might be of interest for the investor to understand the underlying dynamics of power grids to obtain a tailored portfolio. That is to say, how can the investor strategically tailor their portfolio given their risk appetite? Additionally, in this dissertation we will focus on the following key sub-questions:

1. Data generation: Are we able to simulate the statistics of the energy throughput of real-world data?

2. Revenue stream construction: If so, what are the statistics of the associated revenue streams?

3. Portfolio optimization: How can we select the most efficient portfolio based on the revenue streams of each asset?

4. Heuristic testing: Are there any rules-of-thumb strategies that consistently lead to better financial performance?

Chapter 2

Models

It already might have become obvious to the reader that this techno-economic problem can be divided into two separate sub-problems: a technical problem and an ordinary financial portfolio problem. In fact, these two separate sub-problems are weakly coupled, i.e., they can be modularized (Figure 2.1).

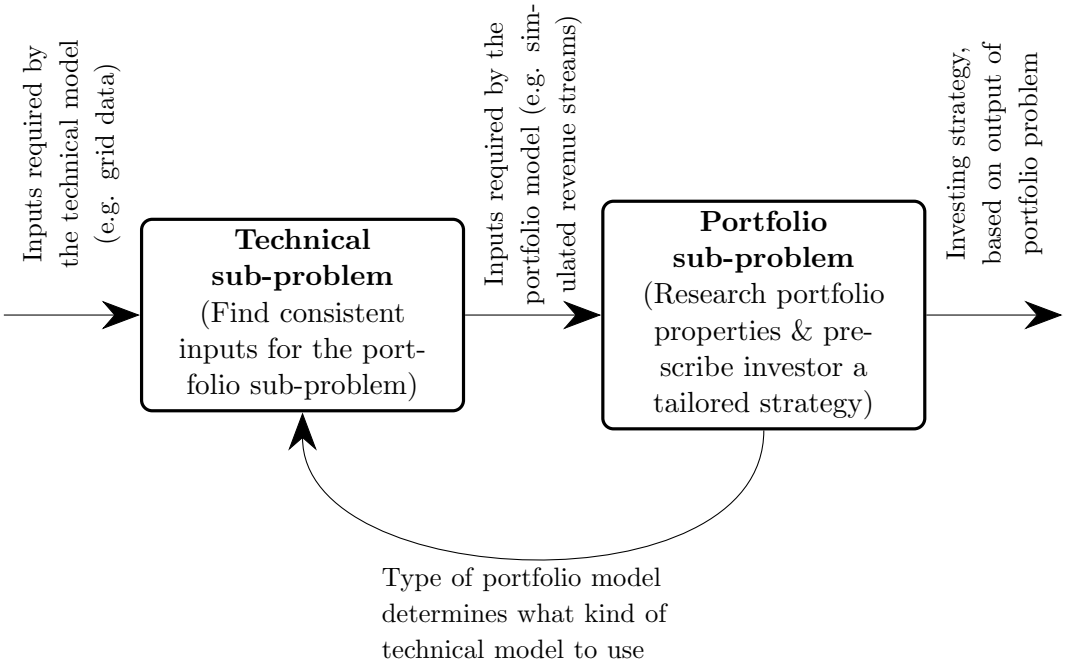

FIGURE 2.1: This particular problem is essentially a portfolio problem in which the inputs require some technical simulations. The reason for this is that day-to-day data of operational metrics in a power grid are not publicly available for privacy and safety reasons

It is worth taking note that there is some degree of coupling: the portfolio model determines what kind of technical model to use. If the goal is risk assessment based on small risks, then a grid model with some noise added will suffice. However, if the goal is to avoid large risks (e.g., blackouts), then a technical model should be used that captures the type of fast dynamics associated with blackouts and other catastrophic events.

2.1 Technical sub-problem

Assuming the case transaction, what is relevant for an investor from a portfolio point of view is the revenue stream of each asset (Equation 2.1:

$$R_{i,t} = U_{i,t} \cdot p_{i,t} - Z_{i,t} \tag{2.1}$$

where:

$R_{i,t}$ = Revenue of asset i on day t [\$]
$U_{i,t}$ = Usage of asset i on day t [MWh]
$P_{i,t}$ = Usage price of asset i on day t [\$/MWh]
$Z_{i,t}$ = Cost of power interruption for asset i on day t (binary) [\$/day]
i = amount of nodes ($\in N$)

Operational grid data like the energy throughput through a node ($U_(i,t)$) and outage of node ($Z_(i,t)$) are usually not disclosed by any grid operator for privacy and safety reasons, at least not in a time-scale that is relevant for longer-term investment problems. This is why we are obliged to simulate the data in order to analyze it.

It has to be noted that this is in stark contrast to what a typical investor would do. Usually, they rely on historical data for their investment decisions.

2.1.1 The distribution of both $U_{i,t}$ and $Z_{i,t}$

It is well known that power grids are prone to catastrophic events, i.e., blackouts. In fact, evidence for power law behavior in the pdf of blackout distribution has been found in many power grids through the world. Although the statistics given in the literature

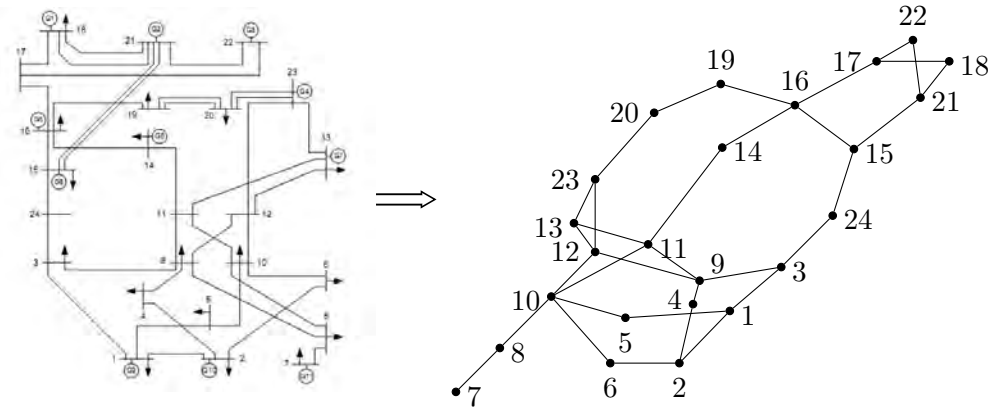

FIGURE 2.2: The same IEEE 24 RTS system is displayed here above: the left picture is the typical power engineering format, while the right picture is generated using the included 'force' algorithm of MATLAB. From a mathematical point of view, they are the same though. Note that the redundancies (parallel lines) are not displayed in the right picture.

usually refer to the energy shed (the fraction of energy that was not able to be delivered), it is not unreasonable to expect that both energy throughput ($U_(i,t)$) and node outage ($Z_(i,t)$) are similarly distributed or at least show some kind of interdependence between nodes.

2.1.2 A primer on power network theory

As with any network, it is conventional to model a power grid as a graph comprising nodes and links (Figure 2.2). In a very simplistic manner, the links represent transmission lines while the nodes usually represent a substation, generator, or load. The real grid, of course, is much more complex, but to understand the power flow dynamics this simplistic view is usually enough.

The power dynamics through a network are governed by Kirchhoff's circuit laws. As a consequence, the dynamics of power flow through the network can be modelled using the nodes real and reactive power balances (Equation 2.2):

$$\begin{aligned} P_k &= \sum_{j=1}^{N} |V_k||V_j| \left(G_{kj} cos(\Theta_k - \Theta_j) + B_{kj} sin(\Theta_k - \Theta_j)\right) \\ Q_k &= \sum^{N} |V_k||V_j| \left(G_{kj} sin(\Theta_k - \Theta_j) - B_{kj} cos(\Theta_k - \Theta_j)\right) \end{aligned} \tag{2.2}$$

where:

P_k = real power injection on node k

Q_k = reactive power injection on node k

V_k = magnitude of the voltage phasor of node k

G_{kj} = conductance of the line between node k and j

Θ_k = phase of the voltage phasor node k

B_{kj} = reactance of the line between node k and j

k, j = amount of nodes ($\in \mathbb{N}$)

In other words, power flow through the nodes is a function of the admittance of the lines and voltage phasor of the nodes, better known as the classic power flow problem.

To solve this non-linear problem is expensive computationally as it can only be solved using numerical methods like NewtonRaphson. However, by approximating the grid as a Direct Current (DC) system (by linearizing the voltage angles), the problem becomes linear and, as such, computationally efficient using SIMPLEX or other variants.

$$P_k = \sum_{j=1, j \neq k}^{N} \left(B_{kj} \left(\Theta_k - \Theta_j\right)\right) \tag{2.3}$$

Note that the reactive power is not given, as in the DC approximation for a well-engineered power grid: $P_k \ggg Q_k$.

It is usually the job of the grid operator to optimize the flows through the grid in a way that maximizes a given economic objective function. This has to respect not only the power flow dynamics, but also the conservation of energy and thermal line constraint equations. The general Optimal Power Flow (OPF) problem, as this mathematical problem is commonly called, consists of solving an optimization problem of the form equation 2.4:

$$\begin{aligned} &\min_x f(x) \\ &\text{subject to} \\ &\quad g(x) = 0 \\ &\quad h(x) = 0 \\ &\quad x_{\min} \leq x \leq x_{\max} \end{aligned} \tag{2.4}$$

For the DC approximation (DCOPF), this results in the following problem (Equation 2.5):

$$\min_{\Theta, P_g} \sum_{t=1}^{n_g} f_P^t(p_g^t)$$

subject to

$$\begin{aligned} & g_P(\Theta, P_g) = B_{\text{bus}}\Theta + P_{\text{bus,shift}} + P_d + G_{sh} - C_g P_g = 0 \\ & h_f(\Theta) = B_f\Theta + P_{f,\text{shift}} - F_{\max} \leq 0 \\ & h_t(\Theta) = -B_f\Theta - P_{f,\text{shift}} - F_{\max} \leq 0 \\ & \Theta_i^{ref} \leq \Theta_i \leq \Theta_i^{ref} \quad i \in x_{ref} \\ & p_g^{t,min} \leq p_g^t \leq p_g^{t,max} \quad i = 1 \cdots n_g \end{aligned} \tag{2.5}$$

where:

$f()$ = economic objective function
P_g = real power injection of generator g
$G_{k,j}$ = conductance of the line between node k and j
Θ = vector containing all phases of the nodes
$B_{k,j}$ = reactance of the line between node k and j
k, j = amount of nodes ($\in N$)

One may argue whether this kind of approximation is justifiable. The literature has shown that the error between an ACOPF solution and DCOPF is small enough to compensate for the enormous gain in calculation speed by solving a linear set of equations.

2.1.3 Understanding the (technical) risk of power assets

In general, the source of volatility of energy throughput of a node is both endogenous and exogenous. The former comes from the power dispatching activities that arise whenever supply and demand change dynamically. There are many factors that have an effect on the so called load profile. However, the biggest is when the grid operator sets an (economic) objective function to optimize (see OPF). Because of the complex network structure, it is then possible that a node that is far away will experience a

change in energy throughput. The exogenous factors include sudden line trips due to storms or other hazards. Planned and unplanned maintenance activities also could be considered exogenous factors. According to , these events (when we account for their monetary value of damage) are very much power law-distributed. Consequently, this means that their impact (although rare) can be much bigger than the volatility generated by normal operations. This statistical behavior can generally be produced through many mechanisms. However, the complex network structure of the system makes it very plausible that there is some kind interdependent relation that achieves some degree of criticality. That is to say, the probability of a node failing becomes higher when another node in the network already has failed. If this difference in probability is high enough, then there is a chance for total system collapse. This domino effect, often compared with avalanches, creates in a similar fashion interdependencies in the investment portfolio. However, the degree of independence may vary.

2.1.4 Picking the right grid model

The root of the previously identified interdependence is, of course, the severe complexity of a system like a power grid. To put this into perspective: the California system itself consisted of approximately 16,000 buses (and this was six years ago!). Add to this the stochastic demand of each customer and producer (renewable energy sources are known for their fluctuating energy supply) and we have a perfect example of a complex network system.

As always, when we deal with complexity, it is never the goal to make a model that replicates all the complexity. In fact, most of the time a Keep It Simple and Stupid (KISS) model is the only way feasible way to study these kind of systems: you only model the specific dynamics in the system that you are interested in and keep it as simple as possible. In this particular case, we are solely interested in the risk of the revenue streams as given by Equation 2.1. As we suspect some kind interdependencies, studying the variance in this time series is not enough. In other words, the model we use should be able to capture the non-linear dynamics of the grid that produces blackouts. This is because if the large events overshadow the small events (variance), then a portfolio optimization problem based solely on variance is incomplete and might paint a wrong picture. Given that our model of choice should be able to reproduce

the fast-dynamics associated with the large impact events mentioned above, we put our focus to models that reproduce outages. The literature has identified not one, but many modes of failure that are responsible for grid outages. Examples include overloaded lines, hidden failures, and frequency-induced trips. Please note that we are not referring here to the triggering events, see Figure 2.3. The difference is in the endogeneity: a triggering event is exogenous and a mode of failure is an endogenous mechanism that facilitates another event. This does not necessarily mean that each of these events will produce a blackout: the consensus is that a blackout is the combination of a stressed system state and a triggering event. Reliability engineers know this very well and, therefore, design enough redundancy to survive any triggering event that affects only one device, the so-called N-1 protection. From an economic and mathematical point of view, this is as far as we can go. For example, with a 16,000 bus grid the amount of redundancy grows from 16,000 to almost 256 million states that have to be individually protected in the of case N-2 protection[1] . This factorial increase is one of the fundamental reasons why even though protection technology is advancing, the rate of power grid blackouts is still not decreasing. In fact, some may argue that the probability of a large blackout is still increasing.

Nevertheless N-1 protection is still beneficial as a minimum engineering safety margin. It instantly rules out most the probable critical system states for a marginal cost in terms of redundancy. However, it neglects improbable system states, e.g., two different triggering events at the same time or one event that affects multiple devices. Given that most of the grid regulators have made N-1 protection compulsory, we can argue that most of the blackouts now are the result of a combination of a critical system state and an improbable (series of) triggering events: a perfect storm.

2.1.5 Branching process

A branching process is a generic stochastic model for processes that involve every form of propagation, e.g., offspring, cascades, and epidemics. The most simple and popular one is the GaltonWatson branching process, which is in the form (Equation 2.6):

[1]The amount of possible contingencies grow with approximately with N^k, where N is the amount of devices and k is the level of protection

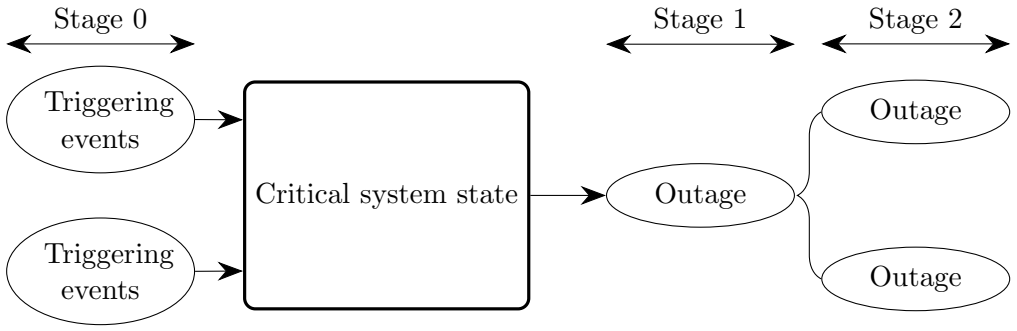

FIGURE 2.3: According to , the consensus is that a blackout is the product of a triggering event (or events) on a critical system state. Accordingly, the modes of failure are the transfer mechanism that facilitate the blackout, i.e. the blue lines.

$$X_{n+1} = \sum_{j=1}^{X_n} \xi_j^{(n)} \tag{2.6}$$

where:

$\left\{\xi_j^{(n)} : n, j \in \mathbb{N}\right\}$ = set of Independent and Identically Distributed (IID) natural number-valued random variables

If ξ_j is Poisson-distributed, then the extinction probability distribution function can be approximated with the following relation (Equation 2.7):

$$x_{n+1} = e^{\lambda(x_n - 1)} \tag{2.7}$$

This implies that when $\lambda > 1$, we can expect the process to explode (supercritical) in the long-term, although there still exists a possibility that the process will die out. In the subcritical case of $\lambda < 1$, the process is guaranteed to die out in the long-term. Perhaps the most interesting case is when λ approaches 1. Around this critical point, the model produces propagations that are power law-distributed with an exponent of 1.5.

In the case of power grid blackouts, evidence has been found for Poisson-distributed branching processes [9]. This implies that the state of the grid can be very roughly approximated by fitting their propagations with a GaltonWatson branching process. In fact, the resulting parametrization of λ can be used as a lagging system state indicator. It is expected that the more a system is stressed the more λ will approach 1.

2.1.6 The OPA model

Although there are plenty of models capable of replicating each of the failure modes, only a few can replicate the flow chart in Figure 2.3 over many system states.

One of the KISS models that has received particular appraisal is the so-called Oak Ridge-PSERC-Alaska (OPA) model. Created by [9]., the model has been well-received in the community due to its ability to replicate the long-term temporal statistics of real-world grids. The fundamental assumption of this model is that the power grid main blackout mechanism is governed by so-called Self-Organized Criticality (SOC).

2.1.6.1 A primer on SOC

The concept itself comes from the physics realm after it was proposed by Bak with the popular sand pile model, better known as the Abelian sand pile[23]. Since then, many variations on the original sand pile have been proposed and extensively analyzed, although the original is still the most popular one. To understand the Abelian sand model, we imagine a circular base with open borders onto which grains of sand are consecutively dropped[20]. After a certain period, a canonical shape will form: a sand pile. Of course, adding the grains will result, eventually, in avalanches of grains because the slope of the pile will be too steep for friction to hold the sand grains together. What is interesting from a physics point of view are the statistics of these avalanchesboth the temporal and spatial statistics of the avalanches reveal that the system grows from a non-critical to critical state. At the beginning, the size of the avalanches will be generally small. Then, after a certain period, bigger avalanches will occur in addition to the occasional smaller avalanches. This kind of phenomenon is usually an indicator that the system has reached a critical regime and has the remarkable property of producing scale-free statistics. In laymans terms, this simply means that in a critical regime, a big avalanche is just a small avalanche that did not stop. This property has the tendency to produce scale invariant probability distribution functions, i.e., demonstrates power law behavior. Aside from these long temporal correlations, there will also be a tendency in the system to produce long spatial correlations, i.e., there is a delicate balance between internal avalanches and avalanches that touch the boundaries.

It is important to understand that although SOC systems produce scale-free probability distributions, not every scale invariant pdf is the result from an SOC mechanism. The difficulty lies in validating such a model, as SOC dynamics can only be fully validated in an experimental setting. In cases where this is not possible, like in a power grid, an SOC mechanism can only be made plausible by comparing the statistics of an SOC system (e.g., the sand pile model) with the statistics of the respective system. In the case of self-organized criticality, the long-range temporal and spatial correlations are characteristics[10] that have to be compared with empirical data.

2.1.6.2 The OPA model explained

This very high-level model has been promising due to its ability to explain (or at least replicate) one of the many complex dynamics behind power outages. Surprisingly enough, the model has been made plausible to some extent by being able to replicate real-world blackout data on the California grid. To understand the general idea behind the OPA model, see Figure 2.4, it might be beneficial to keep the basic sand pile model in mind:

1. We take an existing grid, which is governed by Kirkhoff equations and, therefore, governed by the basic power flow equations.
2. Once we have the stationarity solutions, some noise in the power demand is added in every time-step to simulate different states of the system.
3. At each time-step, we multiply the overall power demand with $_g$, which is calibrated to simulate the annual power demand growth.
4. Given that power grids are engineered to tolerate a certain electrical capacity, usually linked to the thermal limits of the equipment, we expect the system to get ever-more stressed until some of the links are overloaded. Usually, this will happen due to a sudden (positive) shock of noise added in step 2. In the real world, this could be an exceptionally cold day, which naturally drains more energy.
5. Additionally, to the overloaded lines we add exogenous triggering events that cause a line to trip. In the real world, these sudden line trips resemble the improbable events that cause line trips, like falling trees, thunder, etc.

6. Once a line trips (either because of the endogenous or exogenous factors mentioned above), the power is redispatched using the standard optimal power flow algorithm given by Equations 2.5

7. If redispatch is not possible because of the system constraints, we shed load (i.e., generator power gets reduced and, consequently, end customers get cut off). Of course, this is only done as a last resort as the grid regulator applies penalties when this happens. Therefore, this is also implemented in the economic objective function.

8. Once redispatching is done, the algorithm looks if other lines are around their thermal limits. As the redispatch, in general, stresses the whole system, there is a possibility that other lines trip too. In this case, the power is redispatched again and again until, finally, a stable state is achieved. Taking the analogy of the sand pile model into consideration, it is immediately obvious that blackouts are, according this particular model, comparable to the same fast dynamics associated with the sand pile avalanches in the Abelian sand pile model.

9. To mimic the dissipating force of the sand pile model, an enforcement of the tripped lines is done so that in the next time-step, the lines are not only restored but also enforced.

It is arguable whether modelling the reinforcement as a simple capacity increase is justifiable in the following time-step for a couple reasons:

1. Grid planning, as grid capacity expansion process is commonly called, is usually done 510 years in advance.

2. Furthermore, adding capacity to a line usually implies building a parallel set of lines next to the current lines, which is, in most cases, extremely unlikely.

However, since the grid is progressively loaded, the expansion of the grid is also in constant growth. As this in itself takes place at a constant pace, it is not unreasonable to model the grid expansion as simple line and generator capacity growth.

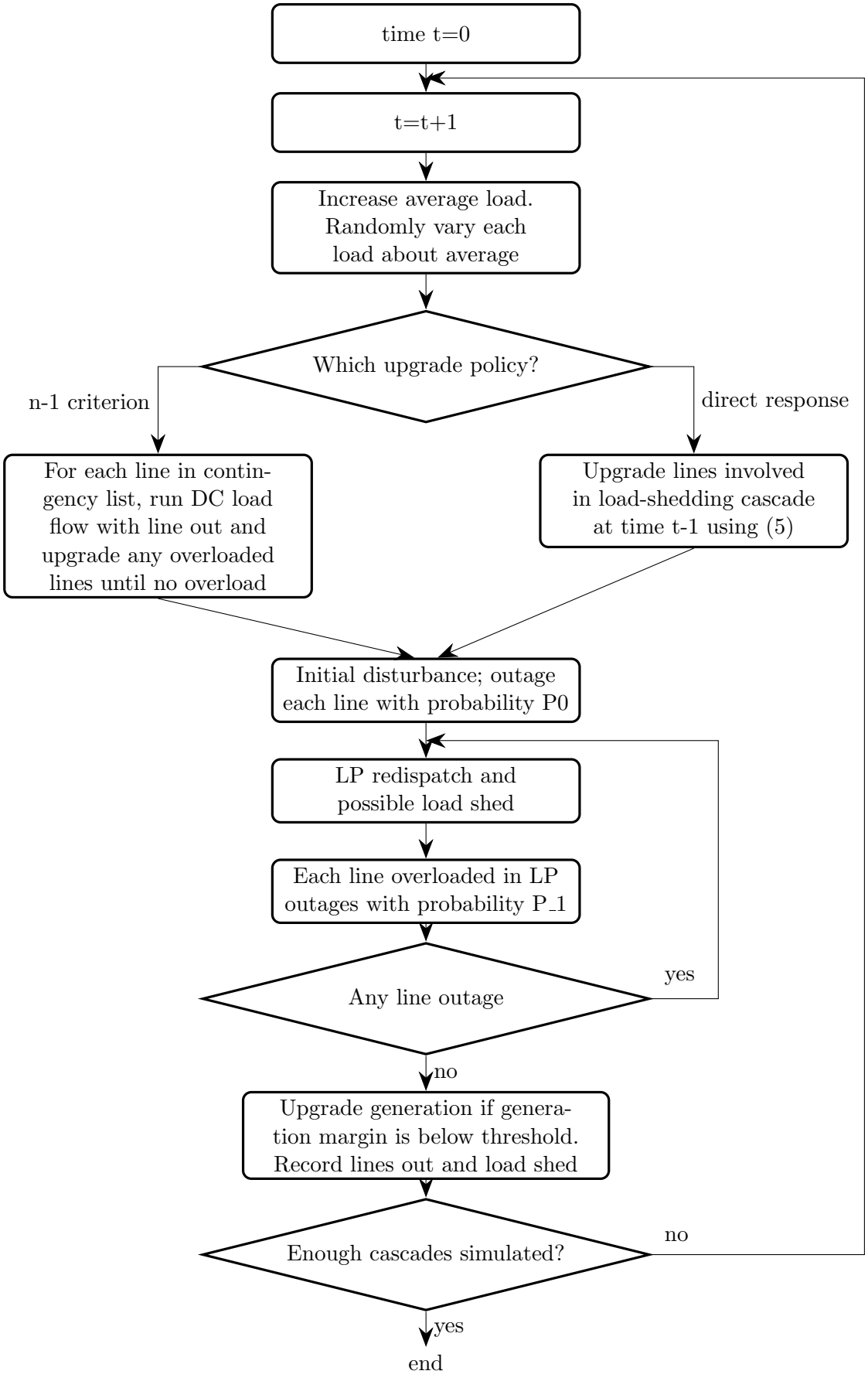

FIGURE 2.4: A variant of the OPA model that takes the N-1 policy into account[5]

2.2 Portfolio sub-problem

The grid data generated with the previous algorithms, of course, are operational only in nature. However, when we translate those values to money values, it is possible to construct the cash flows that are relevant to an investor. More on that, in Chapter 3 Once these are constructed, a rational investor is able to select or deselect which assets generate the most economic utility for them. This problem is, of course, known as an ordinary portfolio optimization problem.

Generally, such a problem is in the form of (equation 2.8):

$$\begin{aligned} &\min_x f(x) \\ &\text{subject to} \\ &\quad g(x) = 0 \\ &\quad h(x) = 0 \\ &\quad x_{\min} \leq x \leq x_{\max} \end{aligned} \tag{2.8}$$

where:

f = economic utility objective function, usually a risk measure
x = vector containing the weights of each respective asset
g = function that characterizes the expected return
h = other constraints that are relevant to the problem

The risk measure is usually a measure for the degree of dispersion of all (or some of) the possible asset states. In the most popular case, this is usually the centered second moment, better known as the variance. A variant of variance is the so-called semi-variance, which only takes dispersion below the center (i.e., the downside risk) into account. Other popular risk measures include: Value at Risk (VaR), expected shortfall, higher moments, or a combination of these. The type of risk measure to use depends generally on:

- The risk appetite of the investor: a risk-averse investor is more interested in avoiding potential losses and, therefore, would rather use the semi-variance than the variance as a risk measure of choice.

- More importantly, the mechanism behind the stochastic element of the asset. An asset that produces normal asset returns can be approximated easily by variance only, as higher-order effects are then guaranteed to be non-existent. However, if the asset has the tendency to produce fatter than normal tails, using only variance as a risk measure is typically not a good practice as the higher-order effects generally are not well-captured. More on that in paragraph 2.2.2.

Furthermore, if there is a risk-free asset in existence, the portfolio optimization problem can be fully tailored to the return requirements of the investor. Usually the desired return, the first moment of the portfolio pdf, is then set in g(x).

The other constraints, in the form of h(x), can include different limitations, such as transactions costs involved in switching between assets, as this seemingly small cost can have a big impact on the overall structure of the most efficient portfolio. This is especially true in dynamically reallocated portfolios in which the portfolio optimization problem is then reiterated over time. Lastly, there is an option to limit portfolios to contain only long (or short) positions by setting an additional constraint to the sign of the weight vector x.

2.2.1 Markowitz efficient frontier

In the original problem, Markowitz proposed portfolio optimization based only on variance as a risk measure[17]. If the optimization is then done for several return targets, then a locus on the return variance plot can be constructed that locates all the efficient portfolios, i.e., for a desired return, there is no other portfolio that offers a better return/risk ratio. This so-called efficient frontier virtually obliges every risk-averse rational investor to allocate their resources in a portfolio that is on the locus. Furthermore, if a risk-free asset is available, then the locus would reduce to a point: the tangent portfolio. In that case, the efficient frontier would simply be a line that captures all the possible combinations of the risk-free asset and the tangent portfolio.

The main assumption behind the math in the Markowitz efficient frontier is that asset returns are generally normally distributed and thus can be parametrized by the covariance matrix of the assets available to the investor. Unfortunately though, fitting historical data to this matrix ignores most of the higher-order dynamics that usually

are present in the regions with higher deviations, i.e. the tail. Although this approach simplifies the math a lot (and therefore is still one of the most widely used method to optimize portfolios), this is something that can only be done with prudence. By definition, large deviations have a bigger impact than small deviations and thus simply averaging them out together into one single metric that lies bigger on small deviations can have unexpected consequences.

2.2.2 Other risk measure

If the return data follows a non-Gaussian distribution, it is more accurate to use other risk measures that lay more emphasis on larger deviations. In particular, if these event are more impactful than the small events, this approach is usually more a prudent way to manage overall portfolio risks. Apart from other risk measures, like VaR, there is also the possibility to generalize the framework prescribed by Markowitz to higher order moments or cumulants. These measures take a bigger emphasis on the larger deviations and as such are a better way to optimize a portfolio when the assets are prone to fat tails.

2.2.3 Moments

In general, moments are implicitly defined by their moment generating function (equation 2.9).

$$\hat{P}(k) = \sum_{n=0}^{+\infty} \frac{(ik)^n}{n!} M_n \tag{2.9}$$

Where $\hat{P}$ Is the Fourier transform of the pdf of S.

$$\hat{P}(k) = \int_{-\infty}^{\infty} dS P(S) e^{ikS} \tag{2.10}$$

Equation 2.10 can be written explicitly as equation 2.11.

$$\mu_n = \int_{-\infty}^{\infty} (x - c)^n f(x) dx \tag{2.11}$$

That is to say: μ_n is the nth moment around c of the pdf of X. Whenever, we center the moments on the first moment, we call it the centered moment. It is immediately obvious that the first moment is the mean and the second centered moment is the variance. From there on we identify both the skewness and kurtosis as, respectively, the third and fourth centered moments.

Chapter 3

Methods

Looking back at our analysis of the problem in Chapter 2, we concluded that the main thesis problem is very much decomposable into both a power engineering problem and a financial portfolio problem. Therefore, although we will prescribe now an exact methodology, the reader should keep in mind that the models that will be combined here can be interexchanged with other models that produce similar outputs. The general recipe presented here is thus a guideline for further work.

The framework, visualized below in Figure 3.1, should serve as an outline for answering the sub-questions presented in Chapter 1 that ultimately will lead to answer the main question.

The choice is made to use the OPA algorithm as the operational data generator for its ability of being able to simulate the cascading dynamics of the grid and as well capture the longer term complex dynamics that (could) govern the grid. Although, there exist many other models that are able to generate the dataset that is required to construct consistent revenue streams, the OPA model has been chosen over the other algorithms because of its recent successful attempts to make the model more plausible by validating the output with real-world data.

In the following paragraphs the methodology of each step in the outline will be explained.

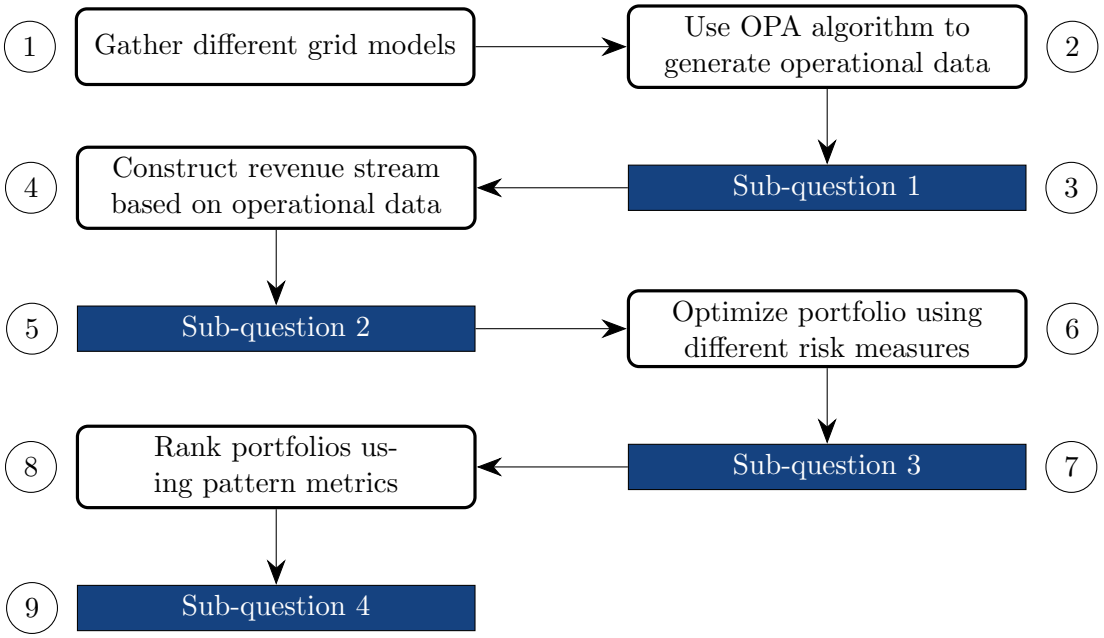

FIGURE 3.1: The general outline in methodology

3.1 Step 1: Grid models

The OPA algorithm is able to use any real-world grid model as input that is complete, i.e. the full topology and electrical properties are disclosed and thus the power flow problem can be solved. Additional to this, the transmission line limits should also be disclosed as these form the basis of the cascade generating mechanism of the model. Although there are many different power grid models publicized, the typical model does not include the transmission limits, mainly because of the antiquity of the models.

Fortunately, there are two relatively small grid models that fulfill the criteria above: the IEEE 24 RTS system and the New England 39 system. The data of these systems can be found in Appendix B.

Thanks to the MATPOWER project [25], the models are already pre-formatted and using the same software the power flow problem can easily be solved. Using the *dcopf* command both grids are solved numerically for the DC approximation. Once done, the inputs for the OPA algorithm are ready to be plugged in.

3.2 Step 2 & 3: Data generation

The OPA algorithm is coded using the recipe given by [19]. The script used to produce the results is given in Appendix **??**.

The choice is made to not model the generator margin to be constant (a small deviation from [19]), but instead to couple the generator capacity growth rate to the load growth rate.

Moreover, we choose the same parameters as used by [4], as those paramaters were validated and thus provide the most confidence in representing reality:

Parameter name	Parameter value
Simulation time	50000 days
Energy demand growth factor	1.00005 per day
Capacity growth factor	1.005 if failure
Probability of initial line failure	0.001
Probability of failure of overloaded line	0.15

Lastly, we remark that the OPA algorithm's original purpose is not to study the energy throughput through nodes. In fact, to our knowledge no public paper has used the load data between nodes in any of their research. The model's original goal was to explain the complex dynamics behind power grid outages. It achieves this by applying a Monte Carlo simulation over differnent states of the grid. It is therefore, in our opinion, not unreasonable to use the data for other purposes. However, we will first compare the statistics of the data to real world data, to add some confidence in this approach before analyzing the data.

The operational data generated by the OPA algorithm will thus be validated against real world data. The data is a courtesy of a consulting firm and as such is anonymized for privacy reasons. The load curves (which is the community term for the times-series of the energy-throughput) are a hourly sample of one full year and extracted from a 100-150 kV substation.

The moments of the real-world data, plotted in Figure 3.2, and the simulated data are then compared to each other to see if the there is a match in terms of statistical properties.

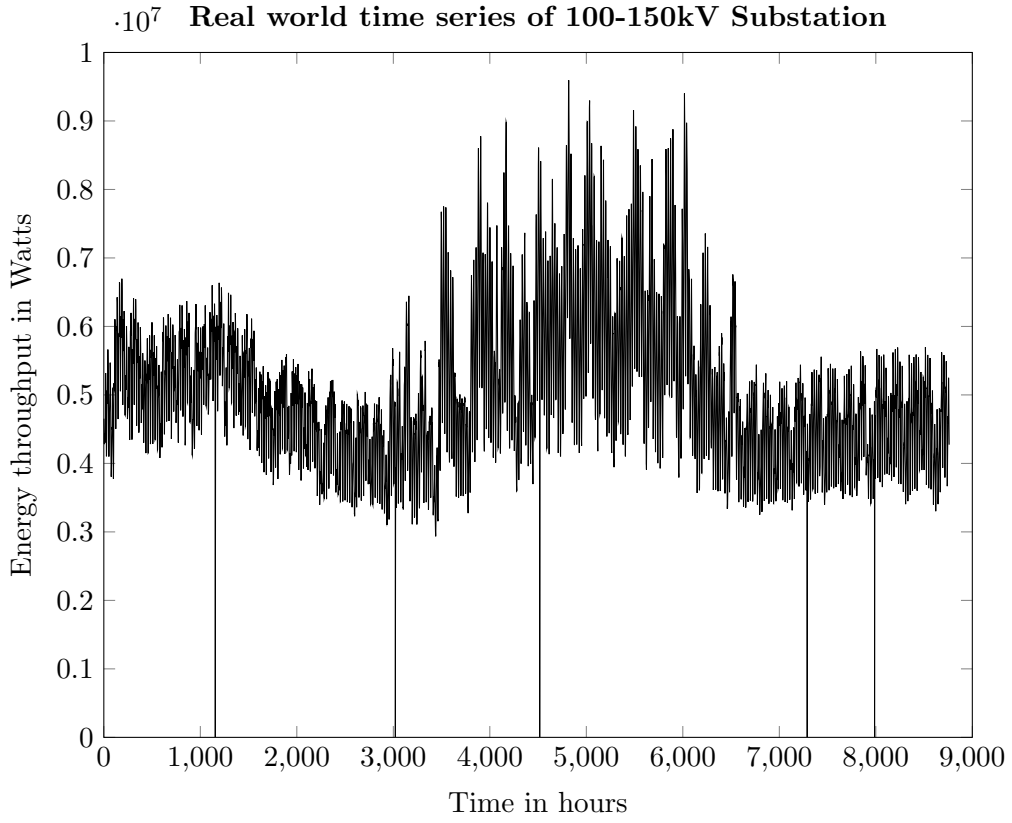

FIGURE 3.2: Real-world data of the energy throughput of a 100-150 kV substation. The timescale is in hours and total time series amount to one year. There is clearly a seasonality element, which is of course a consequence of the higher energy consumption in both summer and/or winter. In fact, the seasonality applies not only to the first moment (mean), but also the second: there is clearly a higher variance visible whenever there is a higher average consumption.

3.3 Step 4 & 5: Revenue stream construction

After validation of the OPA data, we are ready to convert the operational values into financial values.

Ultimately, what is important from an investor point of view is, not the operational data, but of course the revenue stream associated with owning the asset. We will used the simple model described in equation 2.1.

Once we have modelled the usage $U_{i,t}$ of all the assets in the grid, what rests us is thus to calculate the leasing price $p_{i,t}$ of each MWh transferred and as well to set a penalty fee $Z_{i,t}$ for each event of power outage.

3.3.1 Leasing price

In a real-world setting, the price that the lessor will ask from lessee is simply calculated by numerically finding the price that matches the expected IRR of all cash flows associated with owning the asset. This IRR is usually chosen in such a way that it is comparable to the return of assets with similar risk profiles.

It is also worth taking note that in a real world setting, there will be definitely some kind of bargaining going to take place that might skew the ultimate price.

For the sake of brevity, however, we will assume an identical linear model for each of the substations, i.e. we set a price per MWh that is valid for each of the substations. This price is assumed to be 2% of the energy price[1]. Taking the USA power markets into consideration, this figure is thus only 120 USD * 2% = 2.4 USD/MWh.[2]

3.3.2 Penalty fee

Whenever a grid operator is not able to transfer energy, the industry practice is to assign a penalty fee to the agent that is accountable for it. Like for example, agents that generate power are entitled to some kind of compensation fee as they suffer the most of transmission outages, since their high CAPital EXpenditure (CAPEX) structure requires uninterrupted operations to stay profitable. The magnitude of these so-called Energy Not Supplied (ENS) payments (in the unplanned case) amount to sometimes more than a million dollars per day, which is of course a serious tail risk for the grid operator.

However, it has to be said, that although we have defined a clear case transaction in the first chapter, the aspect of risk transfer was not mentioned. This is because a transaction like this will vary strongly on the Terms and Conditions from transaction to transaction. Therefore, for the sake of simplicity, we will assume that if failure occurs the accountable agent will pay fee to an exogenous entity, i.e. that there is no cash flow transfer between revenue streams of nodes.

[1] This price was calculated in a previous business case

[2] Source:https://www.ovoenergy.com/guides/energy-guides/average-electricity-prices-kwh.html

Lastly, we remark that the OPA algorithm does not simulate node outages, but only transmission line outages. To simulate this phenomenon, we will artificially 'manufacture' these node outages. We do this in the following way:

- We log the volatility of the energy throughput per node
- Whenever a node hits a certain (down-side) deviation, we count that as a 'failure' and assign a penalty fee to that event.

The threshold is set to 2.8 times the standard deviation of the node in question. This number was chosen so that most of the nodes are cash-flow positive (with a couple negative ones). This emulates a realistic business case, in which each node has big, but not full, probability of ending cash flow positive.

3.3.3 Statistical analysis

In the same way that the operational data has been analyzed using the methodology prescribed in the previous paragraph, the revenue streams will be analyzed to find if there are any other higher-order effects than just variance and the expected left-tail kurtosis generated by the penalty fees.

Additionally to that, the degree of interdependence between assets will be checked. Tools that will enable this analysis will include the correlation coefficient matrix, which measures the statistical linear interdependence between revenue streams. Although not sufficient to fully grasp the full structure of the data, the correlations coefficient matrix can rule out the hypothesis of fully independent assets. This will be simply visualized by the resulting efficient frontier that can reveal the non-diversifiable risk, i.e. system risk.

For spatial interdependence, we use Morans I using the link distance as distance measure (Equation 3.1) as proposed by [8].

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \hat{X})(X_j - \hat{X})}{\sum_i (X_i - \hat{X})^2} \tag{3.1}$$

where:

N = amount of nodes

w_{ij} = link distance between node i and j

X_i = binary variable that is 1 if node distance is i-j

$\hat{X}$ = mean distance

The statistical significance of Moran's I statistic can be approximated using a normal approximation under randomization [7] to give Equation 3.2.

$$
\begin{aligned}
V(I) &= \frac{ns_1 - s_2 s_3}{(n-1)(n-2)(n-3)\left(\sum_i \sum_j w_{ij}\right)^2} \\
s_1 &= (n^2 - 3n + 3)\left(0.5 \sum_i \sum_j (w_{ij} + w_{jj})^2\right) \\
&\quad - n\left(\sum_i (\sum_j w_{ij} + \sum_j w_{ij})^2\right) + 3\left(\sum_i \sum_j w_{ij}\right)^2 \\
s_2 &= \frac{n^{-1}\sum_i (y_i - \bar{y})^4}{(n^{-1}\sum_i (y_i - \bar{y})^2)^2} \\
s_3 &= 0.5 \sum_i \sum_j (w_i j + w_j i)^2 \\
&\quad - 2n\left(0.5 \sum_i \sum_j (w_{ij} + wji)^2\right) + 6\left(\sum_i \sum_j w_{ij}\right)^2
\end{aligned}
\tag{3.2}
$$

Furthermore, we check if there are clusters of lines that frequently fail together to check for interdependencies in clusters. We check this by constructing a so-called synchronization matrix, as proposed by [4], which tracks which lines fail together more than k times. This reveals potential clusters of lines (and their connecting nodes) that should be avoided including in the portfolio because of the high interdependent risk of failure.

3.4 Step 6 & 7: Portfolio optimization

Once the revenue stream of each particular node are constructed, we compute the cumulated cash flows over the whole investment horizon. We set this to be 10 years (3650 days).

We then check for nodes that fail frequently (and thus result in disproportionally high amount of penalty fees). From this point on, we make the distinction between two separate cases: a portfolio with these frequently failing nodes and one excluding these ones. This is done because the risk in owning an asset is severely asymmetric: there are only large negative deviations possible. This will guarantee them to be excluded of the efficient frontier, as their additional variance can only lower the expected return compared to their counterparts.

Using portfolio optimization techniques, we then compare different portfolios lying on the efficient frontier with the 1/N portfolio to see if there are financial gains feasible by performing such a portfolio optimization. This is done only for the case without the frequently failing nodes.

3.5 Step 8 & 9: Heuristic testing

Once we have found the efficient frontier, the question arises if there is a dominant (topological) pattern present in the locus of efficient portfolios.

Why do we ask this question if the methodology above prescribes the direct composition of the efficient portfolio? Well, the reason is that although the recipe given above should be consistent in the sense that if the operational data generated by the OPA algorithm is consistent with the spatio-statistics of real-world data, then the optimization problem should prescribe the optimal portfolio at all times. However, given that the OPA algorithm is high-level in nature, it is expected that the resulting moments of the revenue stream contain big uncertainties. In fact, the efficient frontier, in general, is so susceptible to uncertainties in inputs that some have even argued that the reason that the Capital Asset Pricing Model (CAPM) hypothesis has consistently been rejected, is because of mere uncertainties in inputs [24]. Especially, because small variations in input can change the portfolio structure dramatically. To combat this we check if there are any heuristics existing in choosing a better portfolio (note that we explicitly use better and not best).

Two potential heuristics that could be relevant to this kind of problems are:

- Mainly investing in nodes higher up in the power transfer chain. Because these usually tend to be less clustered and thus less susceptible to interdependent risk
- Mainly investing in nodes that are decentralized from each other. Again, because there the interdependent risk is suspected to be more diversified more than buying clustered nodes.

To test these two hypotheses, we introduce the following two metrics that should give a general idea on if there are any structural patterns in the efficient frontiers: streamness and modified Katz centrality score [14].

Streamness

We define Streamness using a self-created function 3.3 using a simplistic node mode, see Figure 3.3.

$$
\begin{gathered}
Streamness : P_{gen}, P_{up}, P_{down}, P_{load} \rightarrow [0,1] \\
\text{Only three cases possible} \\
\begin{aligned}
Case1 &: P_{gen} > 0 && \text{then } Streamness = P_{gen}/(P_{gen} + P_{up}) \\
Case2 &: P_{load} > 0 && \text{then } Streamness = P_{load}/(P_{load} + P_{down}) \\
Case3 &: P_{gen} = 0 \cap P_{load} = 0 && \text{then } Streamness = 0.5
\end{aligned}
\end{gathered}
\tag{3.3}
$$

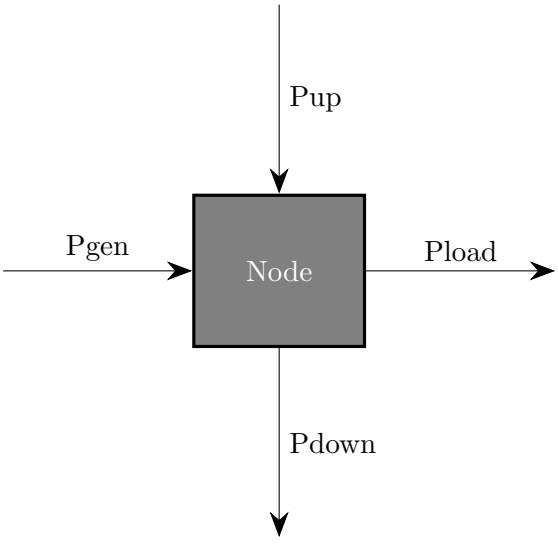

FIGURE 3.3: Each node in the graph can be either exogenous or endogenous power inflow. Mathematically, this is the same for power outflow

Once the streamness of each node is calculated, the overall portfolios degree of Streamness is computed by using a simple arithmetic weighted average (using the portfolios weight w_i calculated in the portfolio optimization problem) given by equation 3.4.

$$Streamness_{portfolio} = \sum_{i \in 1} w_i \cdot Streamness_i \tag{3.4}$$

Katz modified centrality

We define this modified version of Katz centrality measure in equation 3.5. Furthermore, we set the attenuation parameter α to 0.5.

$$C_{Katz,m}(i) = \sum_{k=1}^{\infty} \sum_{j=1}^{n} w_j \alpha^k \left(A^k\right)_{ji} \tag{3.5}$$

where:

w_j = weight of node j
α = attenuation parameters
A^k = adjecancy matrix for k -links between nodes
i, k = amount of nodes ($\in \mathbb{N}$)

The portfolio's clusterness is then simply the sum of the individual nodes clusterness scores. Although the score does not necessarily add up to 1, the score is nevertheless a good way to compare how centralized the portfolios are between each other.

Chapter 4

Results

4.1 Sub-question 1: Data generation

Before actually validating the operational OPA output with the real-world data, we first harmonize the latter so that we can compare 'apples with apples'. Then after this the moments of both data-sets are compared to each other.

The reason for this validation step is that the OPA output has only been validated for the probability distribution of blackouts measured in terms of energy shed. However, the operational data (energy-throughput of each node) has to our knowledge never been validated yet.

4.1.1 Real-world data analysis

To make a better comparison between the generated data and real world data, we de-seasonalize the data using an symmetric 1001-term moving average filter (Equation 4.1, Figure 4.1)

$$\mu_t = \frac{1}{1001} \sum_{i=-500}^{500} P_{t+i} \tag{4.1}$$

Two types of regimes are immediately obvious to the reader: a long period with relatively low volatility (Regime 2) and a shorter period with higher extremes and as well higher

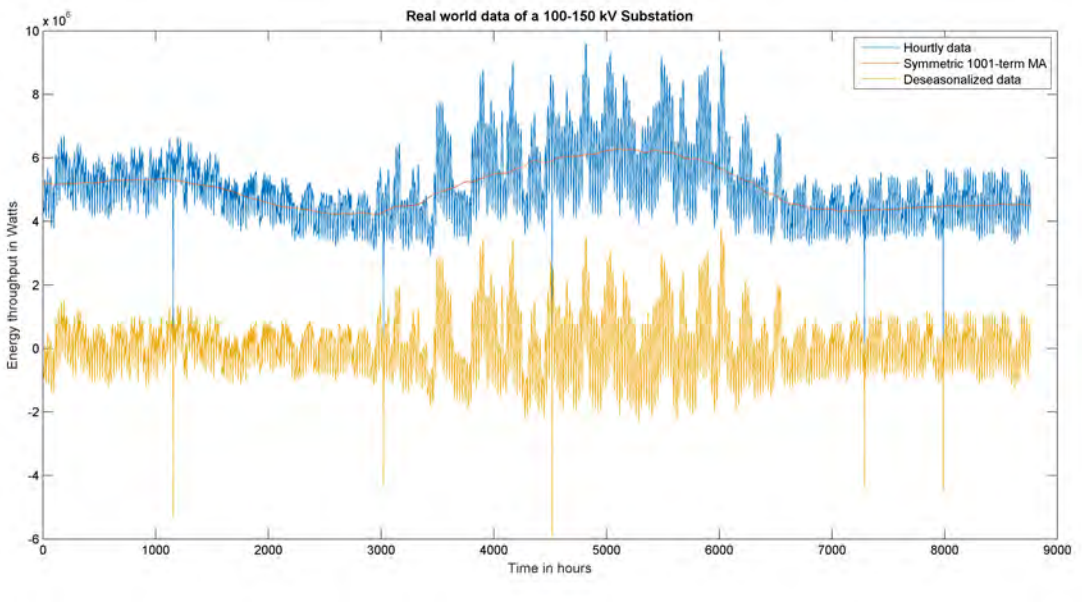

FIGURE 4.1: The original data (blue line) is filtered using a symmetric 1001-term MA filter (red line) to deseasonalize the data (yellow line).

volatility (Regime 1). Because the peak periods seem to have bigger variance as well, we check for temporal autocorrelations (Figure 4.2).

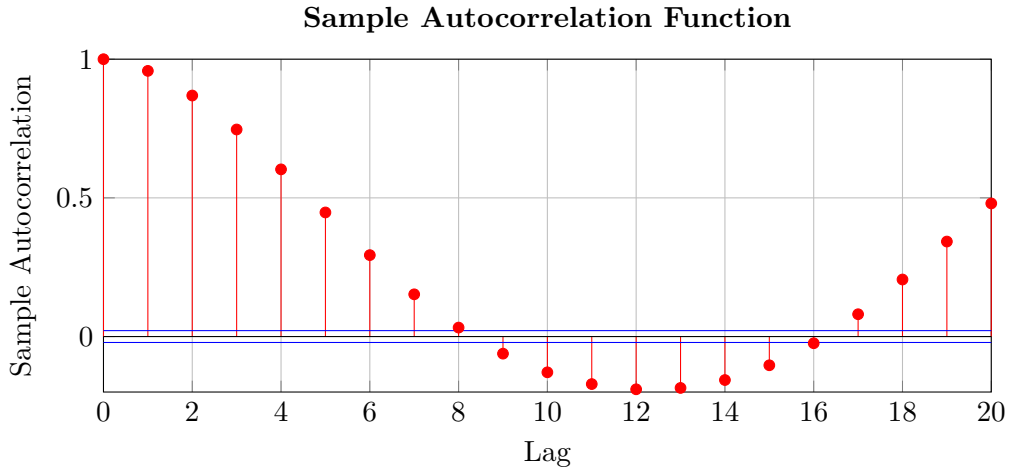

FIGURE 4.2: Autocorrelation plot of the whole data sample using hours as lag-scale.

The autocorrelation shows clearly a daily pattern, which is obviously a result from the intraday periodicity. This is however not a problem as the OPA model uses a daily time-scale. This means that the intraday effects in the real world data can be averaged out without any loss. To rule any other periodicity, we nevertheless make a power spectral analysis of the detrended data (Figure 4.3)

The power spectral analysis shows the obvious daily periodicity and rules out any other

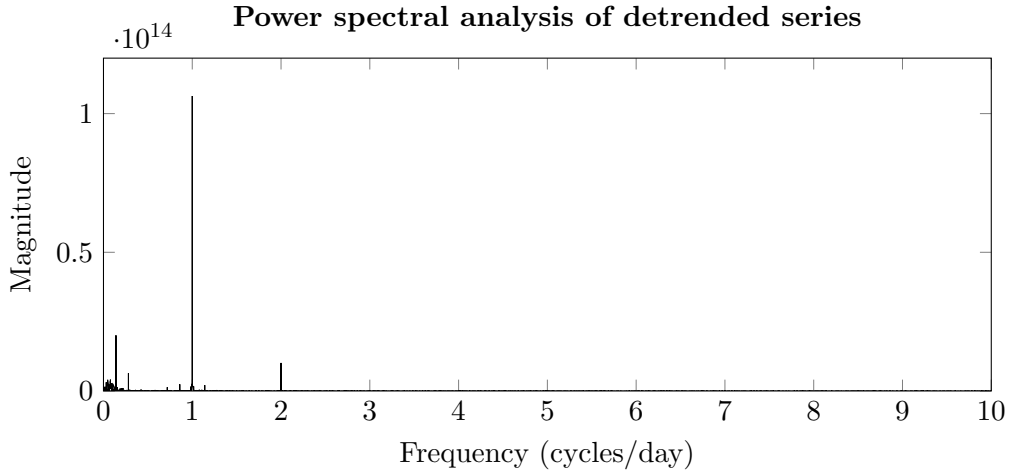

FIGURE 4.3: A power spectral analysis of the sample data . The other two spikes at 3,5 hours and 2 days are speculated to result from the short (breaks every 3-4 hours from working) and long business (breaks in the week) hours cycle.

(significant) periodicity. We then construct a histogram to visualize the probability distribution of energy-throughput after centering the resulting time series (Figure 4.4).

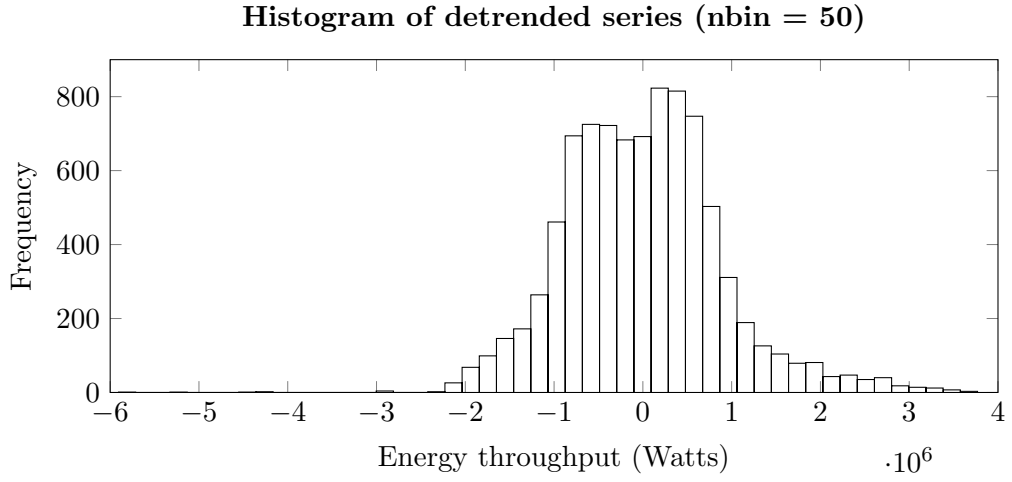

FIGURE 4.4: A histogram of the detrended series after centering.

Although a histogram is generally only a very rough estimation of the actual probability distribution function, it is however a helpful instrument to generate an easy visualization of the empirical frequency data. After observation, one can easily spot the fatter-than-normal left-tail and a huge asymmetry between tails. In particular, the left tail shows evidence of some higher-order effects in energy throughput. Moreover, we speculate the reason that the right tail suddenly stops at around 4 MW is that there are saturation effects build into the system that deaccelerate the consumption demand increase (e.g. higher energy prices)

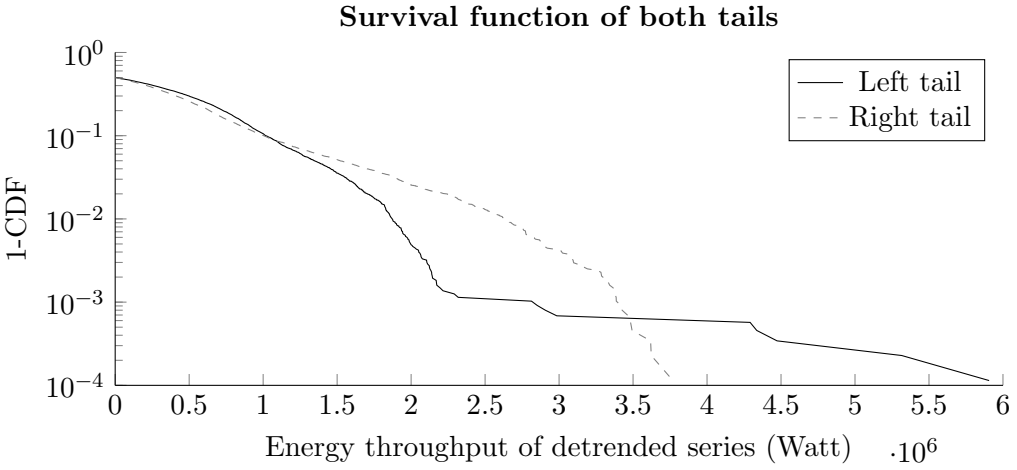

FIGURE 4.5: The survival functions are both plotted on a log-linear scale. There is a huge asymmetry between the tails, which can be explained by (un)planned maintenance and/or power outages that skew the distribution.

To support our assessment of the real-world data, we examine the tails of the empirical distribution function. We perform a survival analysis using the complement cumulative distribution function of both tails (Figure 4.5).

Figure 4.5 confirms our previous observation of asymmetry between tails. The right tail seems to be thinner than exponential and the left tail does on first sight seem to suggest some kind of exogenous factors that produce a discontinuity in the tail. In fact, one could argue that this discontinuity could also be the result of some power law generating mechanism. To check this, we plot the tails on a log-log scale (Figure 4.6).

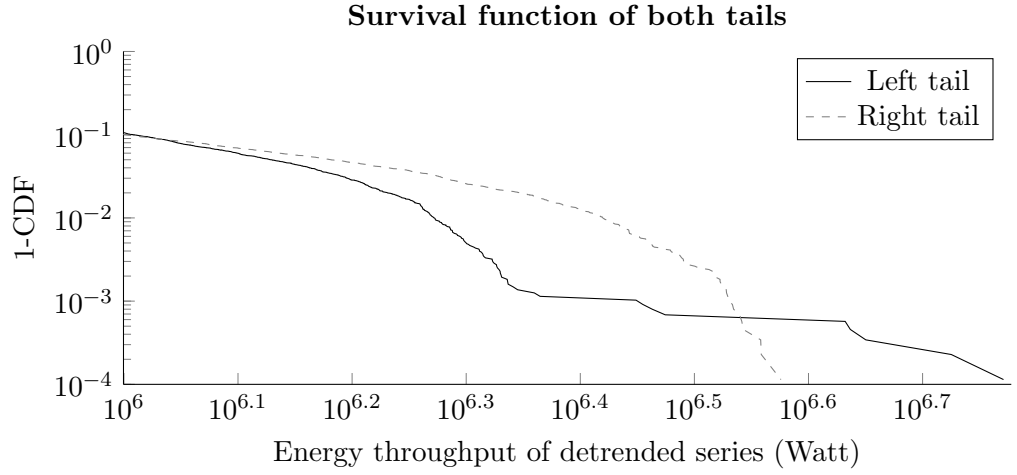

FIGURE 4.6: Both tails on a log-log scale.

The left tail does appear to be drawn from a power law distribution. The discontinuity in the plot is probably because there are multiple mechanisms generating this behavior:

the left-part of the blue line is probably because of the operational variance and the right part of the blue line (after the discontinuity) could be the result of outages and maintenance. It seems thus that there are exogenous factors (from a substation system point-of-view) that generate left-sided heavy tails.

To finalize our analysis, we construct a QQ plot of the sample energy-throughput (Figure 4.7).

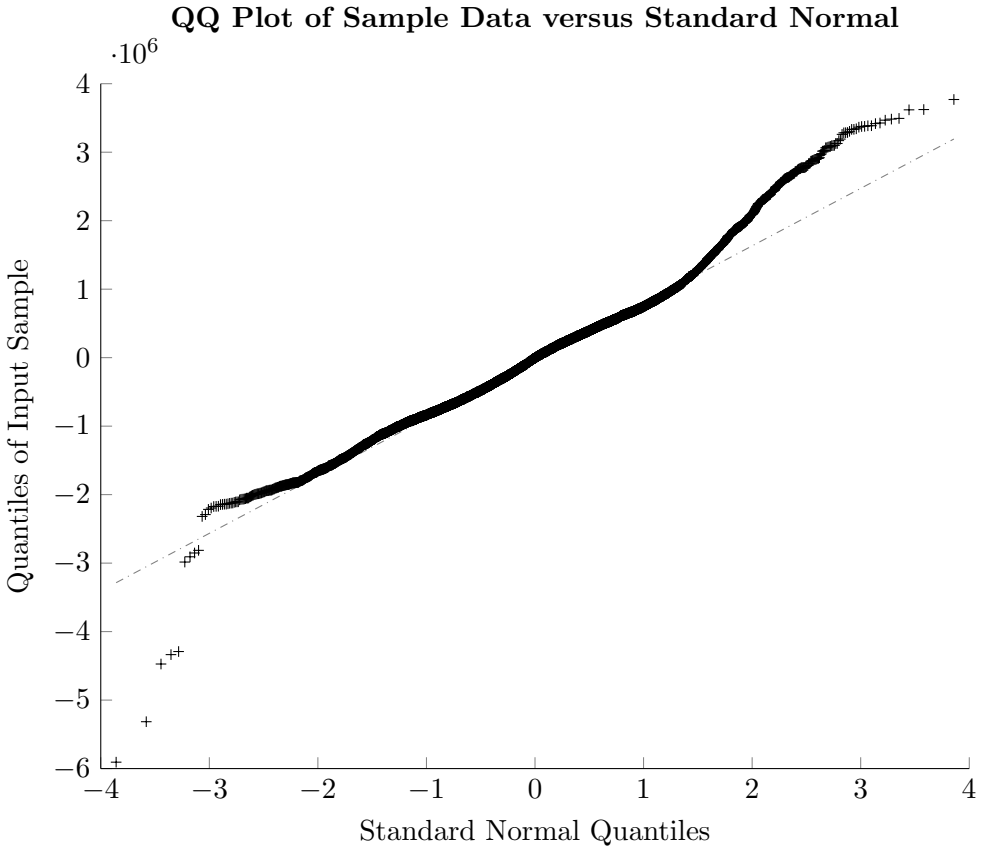

FIGURE 4.7: A QQ-plot of the emperical distributuion against the corresponding normal distribution.

The QQ plot does indeed show the same asymmetry between both tails. The left-tail is definitely heavier than the right tail Figure 4.7 supports our previous judgement on the asymmetry of the tails. The left tail seems absolutely not natural, in the sense that the discontinuity seems to be so diametrical that it suggest exogenous factors to be playing a role.

We tabulate the moments for both regimes, specified in Figure 4.1:

Scale-free moments	Regime 1	Regime 2
Relative Std. Deviation	1.6	0,23
Skewness	0.41	-0.37
Kurtosis	2.7	5.0

The moments of both regimes differ significantly:

- Regime 1 is more volatile, right skewed and less curved than a comparable normal distribution (Kurtosis¡3).
- Regime 2 is less volatile, left skewed and much more curved than a comparable normal distributio (Kurtosis¿3).

4.1.2 Data comparison

Now that the real-world data has been statistically analyzed, we compare both the operational data to the empirical data, so we can proceed with analysis.

Choosing the right time window

We first check the state of system over time. The simulation length is 50000 days (approximately 140 years), which is obviously a very long time horizon in a different time scale than the typical investment horizon. According to [**?**], the state of the system becomes more critical with the demand growth. Furthermore, a metric that parametrizes this criticality is derived from the Galtson-Watson process [3]. We compute λ over time to check if the state of system converges.

Looking at Figure 4.8, we observe that system 24 achieves convergence later than system 39, probably because of different initial states of criticality. According to [12], current grids are being operated around their operational limits, which instructs us to use only the last portion (last 5000 days) of the data. This ensure that we compare the simulated data that resembles the most the real-world data.

To do a last check of convergence, we calculate the Mean Time Between Failure (MTBF) statistic using again 50 partitions.

The convergence again suggest that using only the last portion of the data is the most suitable for comparison with the empirical data.

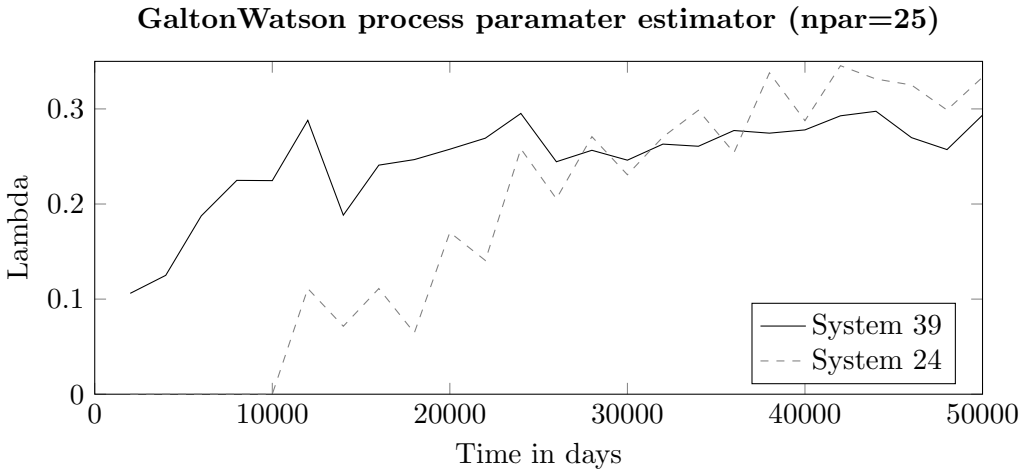

FIGURE 4.8: Galtson-Watson parameter over time (partitioning the data over 25 bins).

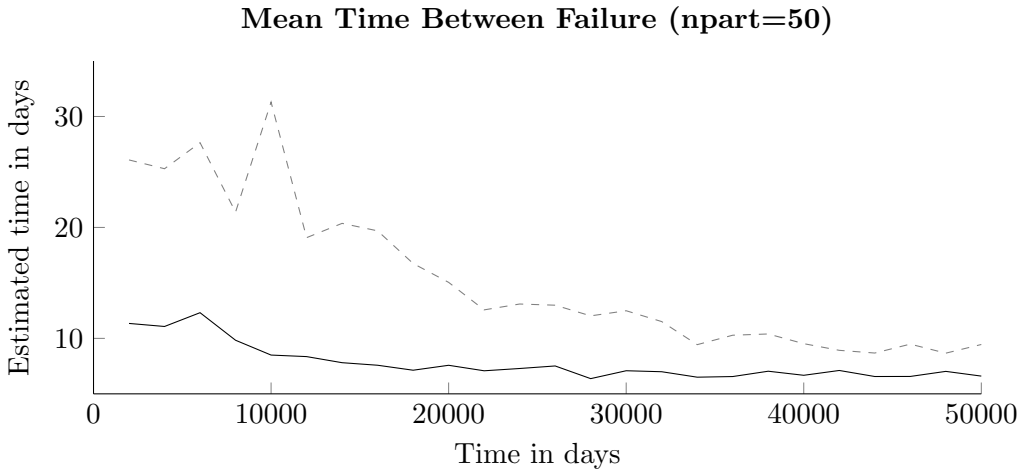

FIGURE 4.9: The Mean Time Between Failure calculated again using 25 bins of partitioned data.

Furthermore, we average out the intraday effects by compressing the hourly data into daily data (Figure 4.10).

We then compare the empirical moments to the moments from the simulated data to see if the data is usable.

Second moment

Using the following backwards looking moving variance filter, Equation 4.2, we plot the daily standard deviation to see how dispersed the deviations themselves are. (Figure 4.11.

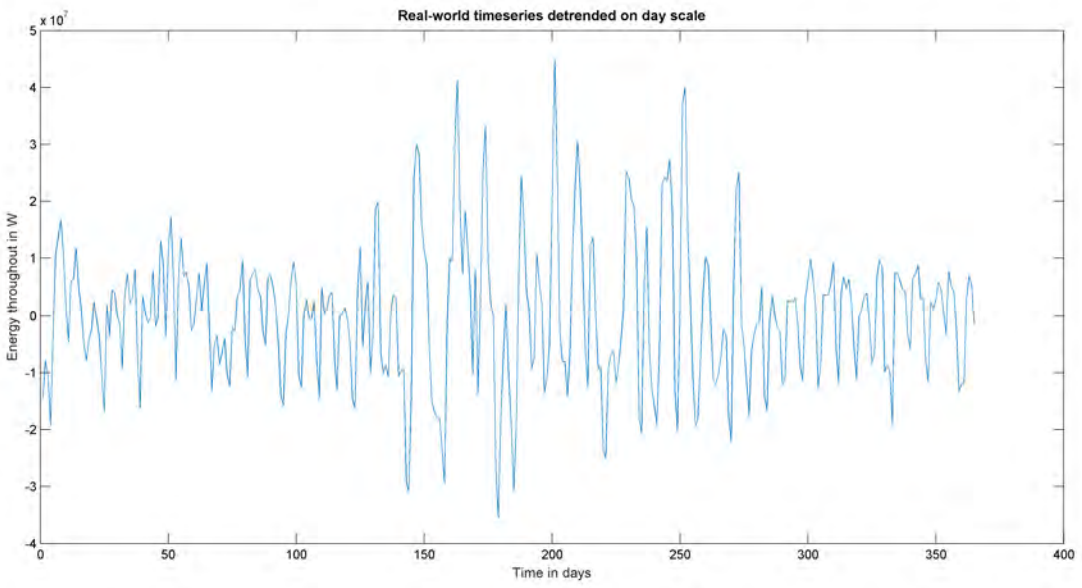

FIGURE 4.10: Compressed, detrended and centered real-world data

$$V = \frac{1}{50-1}\sum_{i=1}^{50}|A_i - \mu|^2 \tag{4.2}$$

where:

$$\mu = \frac{1}{50}\sum_{i=1}^{50} A_i \tag{4.3}$$

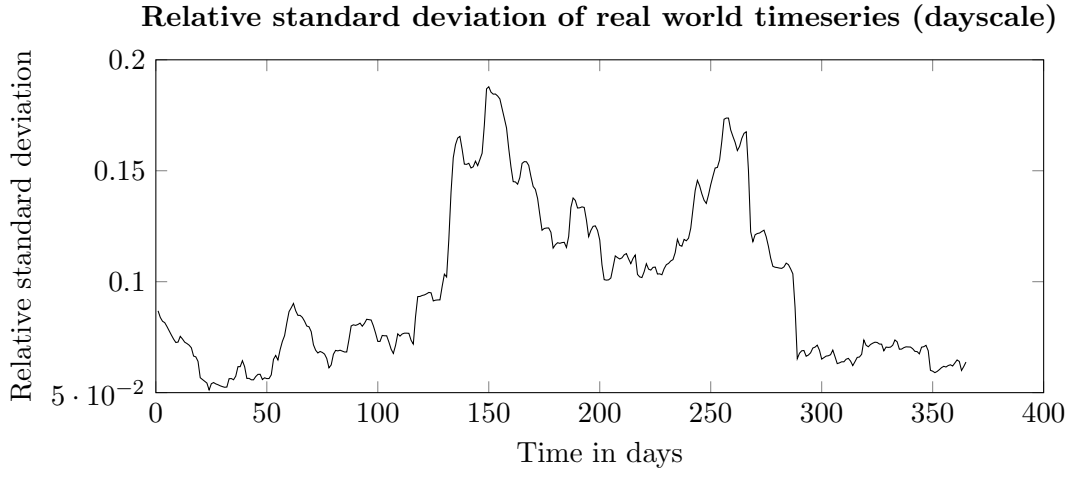

FIGURE 4.11: A plot of the daily relative standard deviations against time. One can see that in Regime 1, the standard deviations themselves deviate a lot.

We compare this with the operational data produced by the OPA algorithm, Figure 4.12.

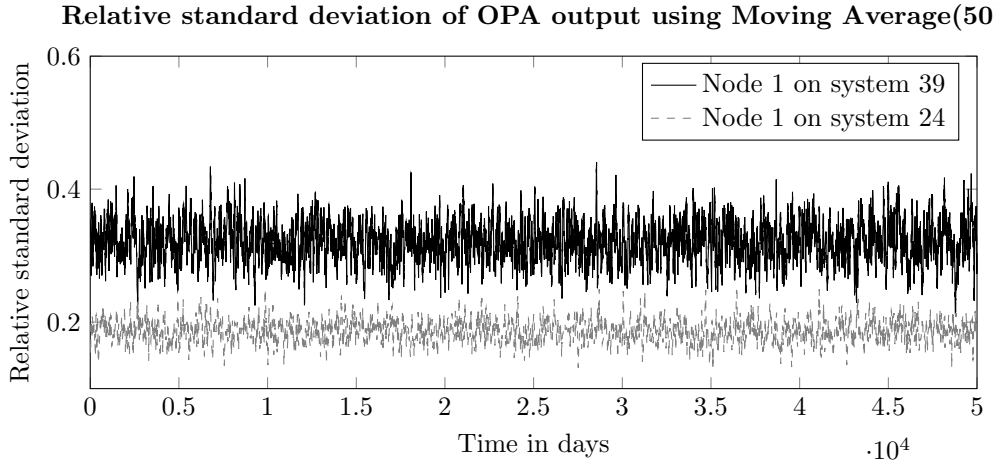

FIGURE 4.12: Scaled down operational standard deviations of OPA model. From both grid models (system 24 and system 39) only the operational data of node 1 are plotted

One can see that the average relative standard deviation of system 39 and system 24 are around, respectively, 0.33 and 0.19. These values are in the same order of magnitude of the values in regime 1 of the real-world data. However, the plot is of a random node (namely node 1). To see if there are differences between nodes, we plot the average relative standard deviations of each node over the last 5000 days (Figure 4.13).

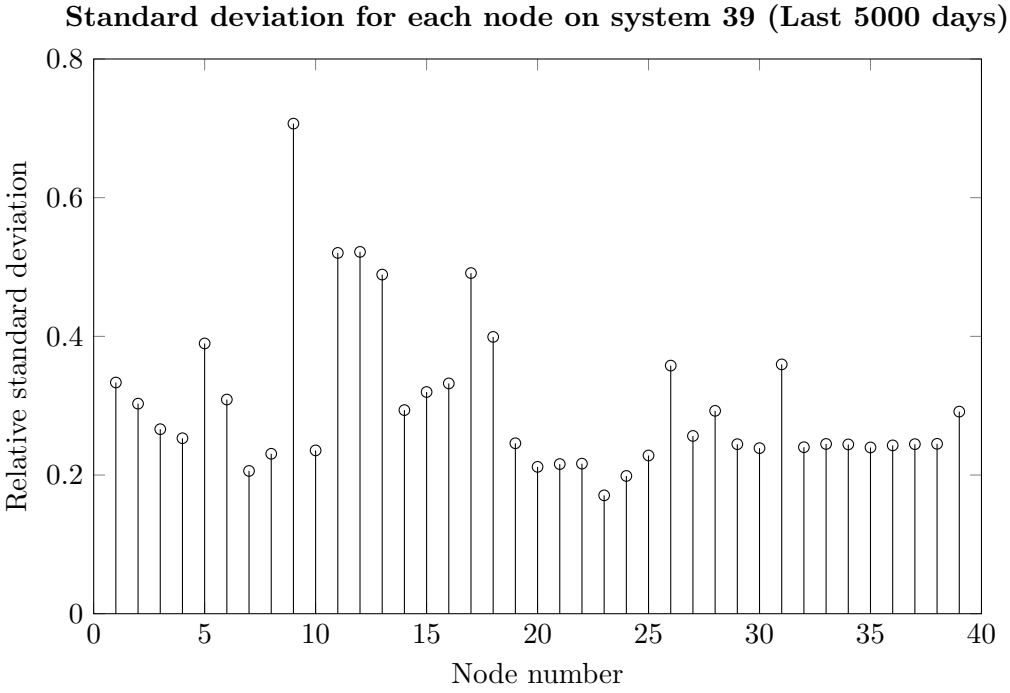

FIGURE 4.13: Standard deviations of each node of system 39 averaged out over last 5000 days of data. We used the same filters as for the real-world data

The relative standard deviation of each node seems to be around 0.28. This gives confidence in the model's output.

In fact, if we look to system 24, see Figure 4.14, our hypothesis that the operational data of OPA is realistic seems even more plausible:

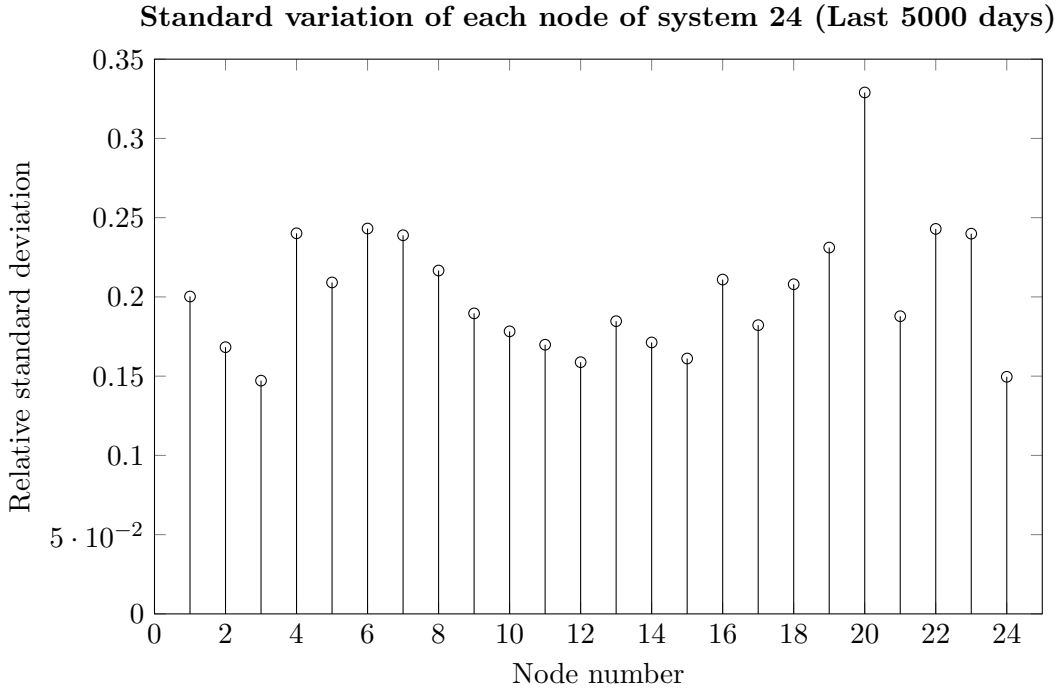

FIGURE 4.14: Relative standard deviation of each node of system 24. Averaged out over the last 5000 days.

With an average of around 0.18, the relative standard deviation seems very comparable to the deviations of the Regime 1 of the real world data.

Third moment

The skewness of the data is a metric for the asymmetry in the data series. We have observed that the sign of this moment seems to change from Regime 1 to Regime 2 in the empirical data.

We compute the skewness of each node of system 24 in Figure 4.15.

The average skewness seems to float around 0.19, which has not only the same sign as the empirical value of Regime 1, but is also in the same order of magnitude. This again seems to suggest that the OPA simulation output is statistically more similar to Regime 1.

Tail analysis

To finalize our comparison between the empirical data and simulation data, we make a plot of both tails of a random node of system 39 (Figure 4.16).

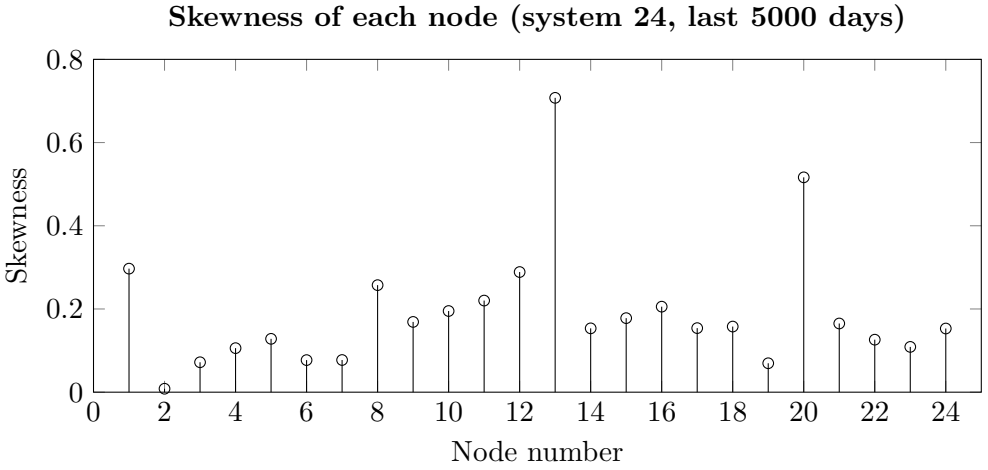

FIGURE 4.15: The skewness of each node of system 24. Averaged out over the last 5000 days.

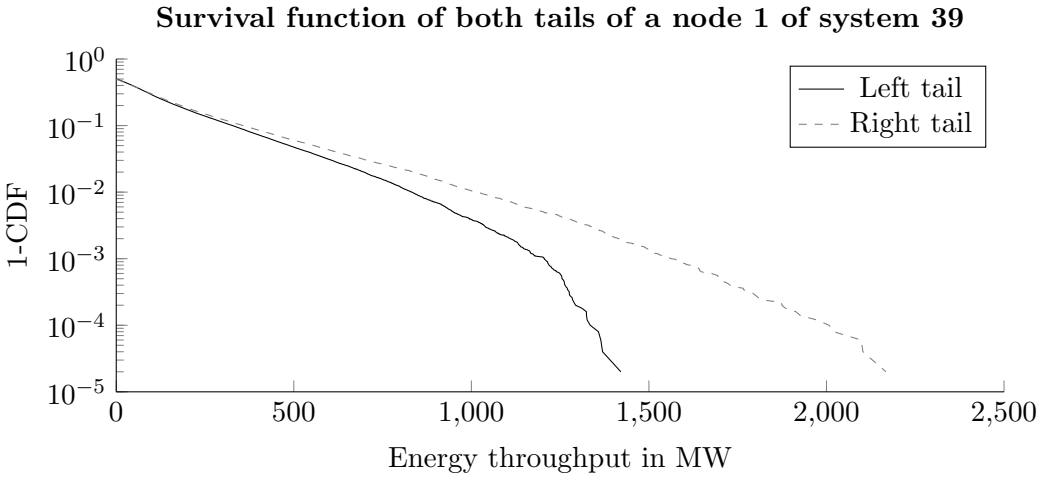

FIGURE 4.16: A survival function plot on log-linear scale of node 1 of system 39 (randomly chosen). Both the left and right tails are plotted

The positive skewness can be clearly identified. However, this the opposite of what is seen in the tail plot of the empirical data. The reason for this is that the survival plot of the empirical data, was not made by make a distinction between Regime 1 and Regime 2. As the positive skewness of Regime 1 seems to suggest, we could have expected a similar plot if we only had taken Regime 1 into consideration. We check this by plotting the survival curves of only Regime 1 (Figure 4.17).

Now the tails seem more comparable:

- The right tail of the simulated data seems to be drawn from an exponential distribution, which is definitely fatter-than-normal and similar to Regime 1.

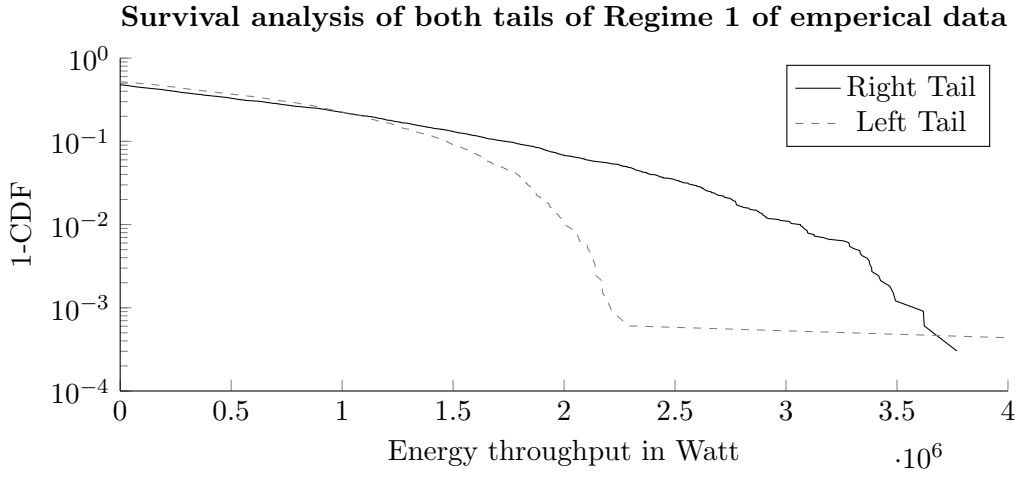

FIGURE 4.17: Survival analysis of empirical data of only Regime 1.

- The left tail still contains the discontinuity, which the operational data misses. The OPA algorithm does not produces node outages, which should be the reason that the discontinuity is not present in the simulated data.

4.2 Sub-question 2: Revenue stream construction

Using equation 2.1 and the parameters given in Chapter 3, we construct for each node the revenue stream.

The stacked bar chart in Figure 4.18 are the accumulated cash flows of owning a node. Note that they exclude the initial investment. Also note that in the first year, only a handful of node do indeed experience outages.

To check for interdependencies between assets, we start first by looking at spatial relationships (Figure 4.19).

We observe a statistically significant spatial autocorrelation at distance 1. After that, we dont see any evidence for spatial autocorrelation.

In fact, if we look over time the autocorrelation function doesnt seem to converge. We do see a continuous progression that is cyclical in nature. That element seems to be consistent over each distance level, which suggest that there could be some statistically significant correlations, which are not captured by the normal approximation of equation.

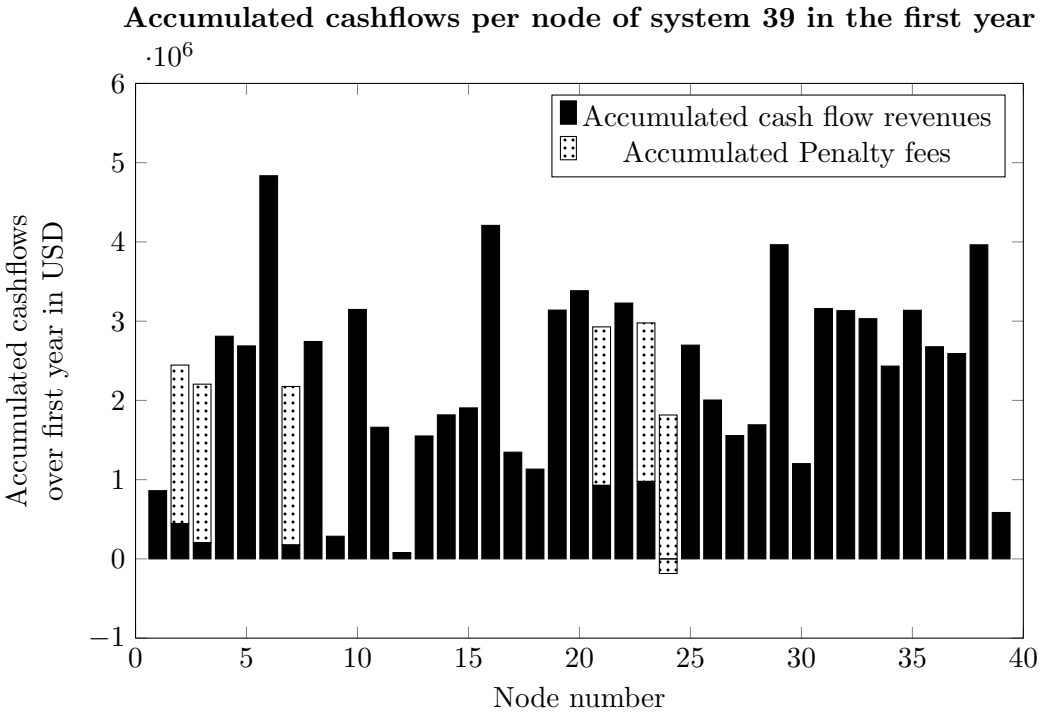

FIGURE 4.18: Accumulated cash flow of the first year using the last 50000 days of simulation sample. Furthermore, we use 2.8 time standard deviations as hit rate for a penalty fee

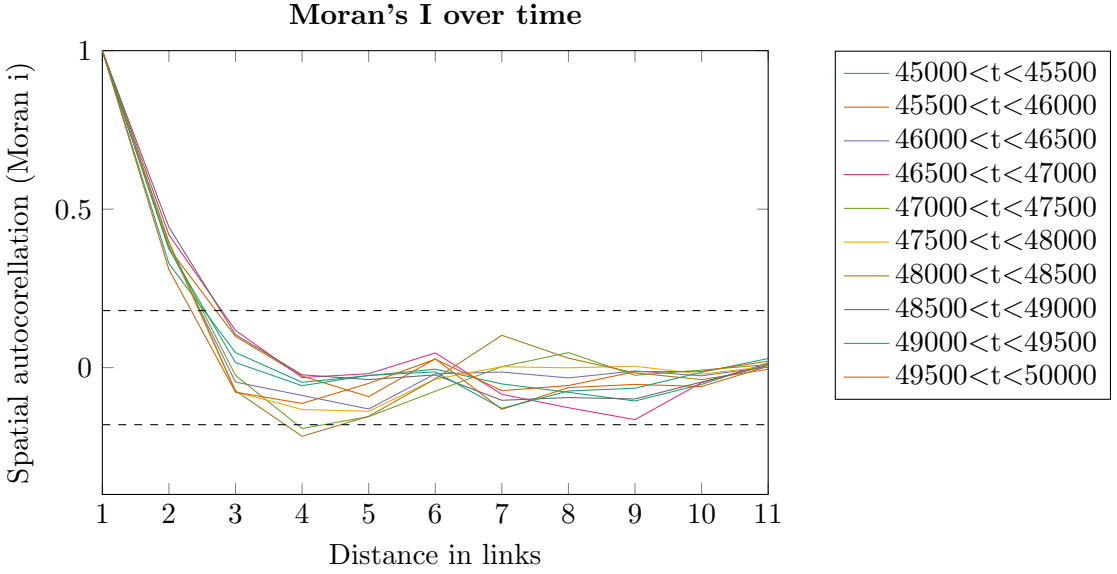

FIGURE 4.19: Spatial autocorrelation (Moran's I) using amount of links as distance unit. The lines are a 95% confidence interval. The graph does not seem to show convergence over the time

To check further for interdependence, we go back to the operational data and check for synchronization in line outages. Using a so-called synchronization matrix, we plot which lines tend to fail frequently together. By setting a certain threshold, we check if there indeed clusters of lines that tend fail frequently. If that is the case, then there is a strong case for not rejecting the hypothesis of interdependence.

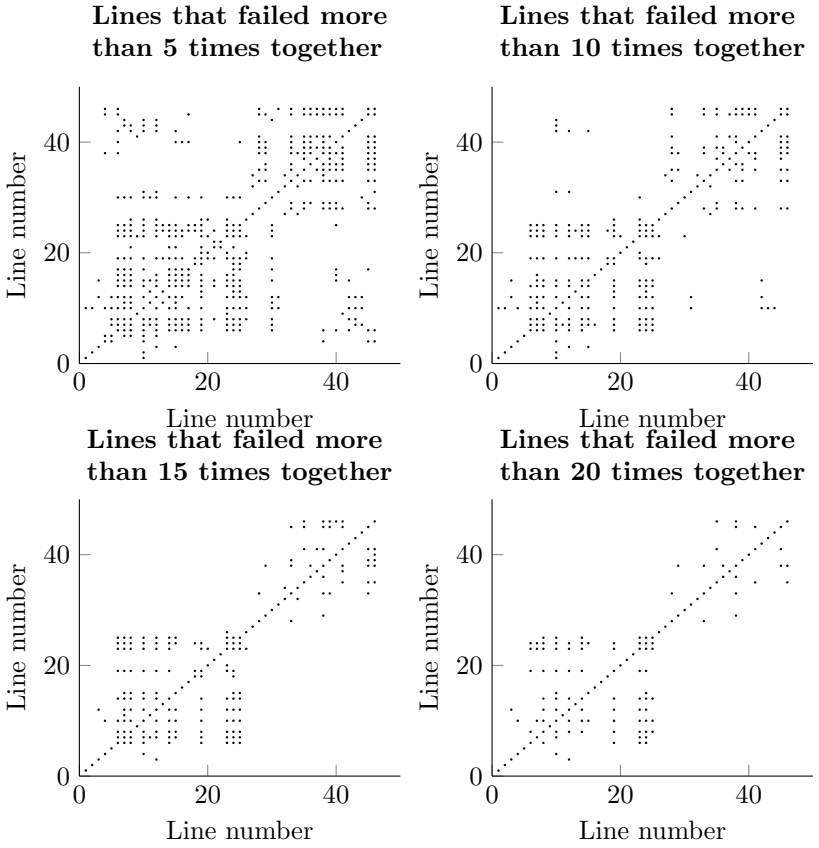

FIGURE 4.20: Synchronization matrices of system 24. The data is sampled over only the whole 50000 days. One can spot that there is a group of lines that fail often.

After inspection, the case of interdependency has been made stronger: often frequently failing group of lines indicate that the failure modes have influence over a group of nodes.

When we construct the histogram of a portfolio returns, see Figure 4.22 consisting of equally weighted assets (also known as the 1/N portfolio), we immediately see a negative skewness and a long left tail. This is of course the product of the inherent asymmetry of penalty fees: there are no performance bonuses when delivering more energy than possible.

From the histogram it is clear that the portfolio with the often-frequent nodes contains a serious tail risk for the investor. The long left-tail could be a reason to not use variance

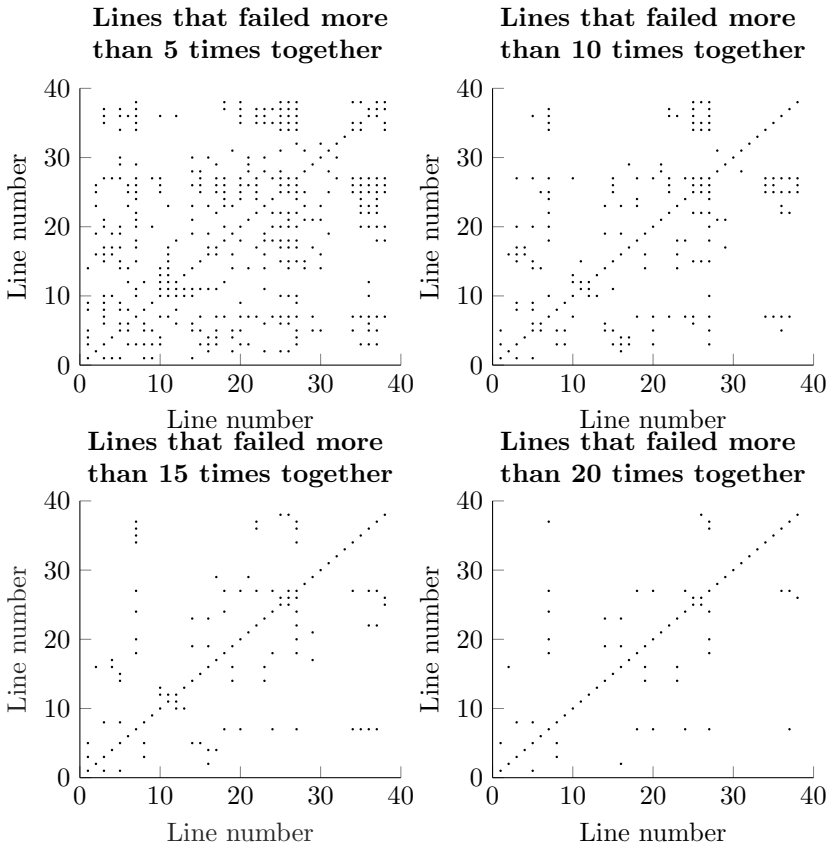

FIGURE 4.21: The synchronization matrices of system 39 show even more frequently failing lines (again sampled over the whole 50000 days).

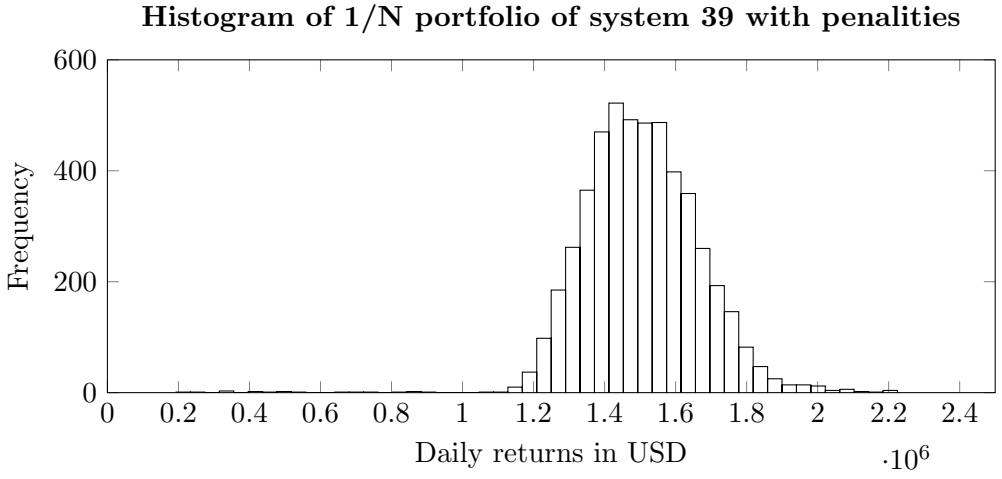

FIGURE 4.22: A histogram of the daily returns of a 1/N portfolio consisting of all the nodes of system 39

as risk measure; a higher moment risk measure could be more beneficial.

4.3 Sub-question 3: Portfolio optimization

From the previous paragraph, it is clear that the best portfolio optimization strategy is obviously omitting the frequently failing nodes: their long tail risk is only downwards directed and thus will never be part of any efficient portfolio. In fact, just detecting these group of frequently failing nodes and omitting them from the portfolio can positively skew the return distribution, as can be seen from Figure 4.23.

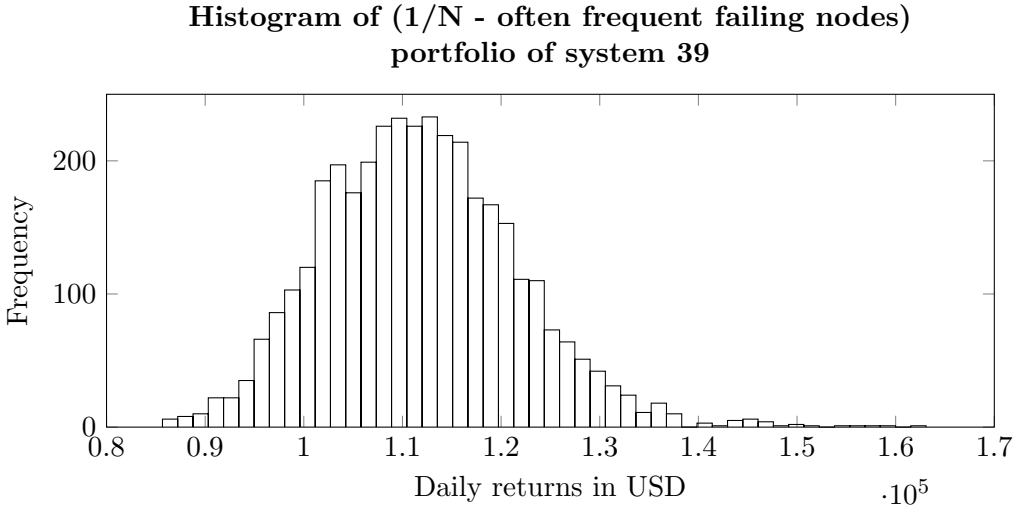

FIGURE 4.23: A histogram of the daily returns of a 1/N portfolio excluding all the frequently failing nodes of system 39.

Omitting the frequently failing nodes is thus very rewarding both in terms of returns (higher mean) as risk (only upwards risk). However, we can go one step further and optimize the restant of the portfolio using Markowitz efficient portfolio theory, see Figure 4.24 and Figure 4.25. By explicitly omitting the frequently failing nodes, we are basically only diversifying the operational risk, which is subject to the noise that the power demand change generates.

Interestingly enough, the 1/N portfolio is in both cases very close to the frontier. Furthermore, we remark that the risk-free asset that is usually present in these kind of portfolio optimization analyses is omitted on purpose: we assume that the investor has other assets in his portfolio that might change the overall frontier. As such, constructing a tangent portfolio in this particular case doesnt make sense.

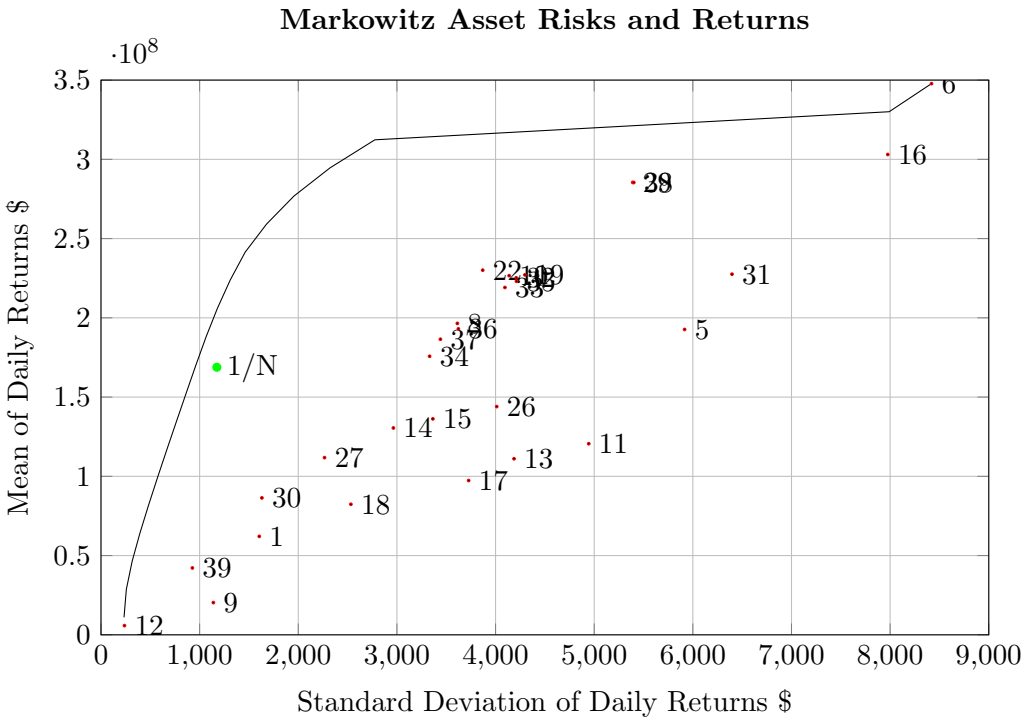

FIGURE 4.24: Markowitz efficient frontier of all system 39 nodes excluding the frequently failing nodes

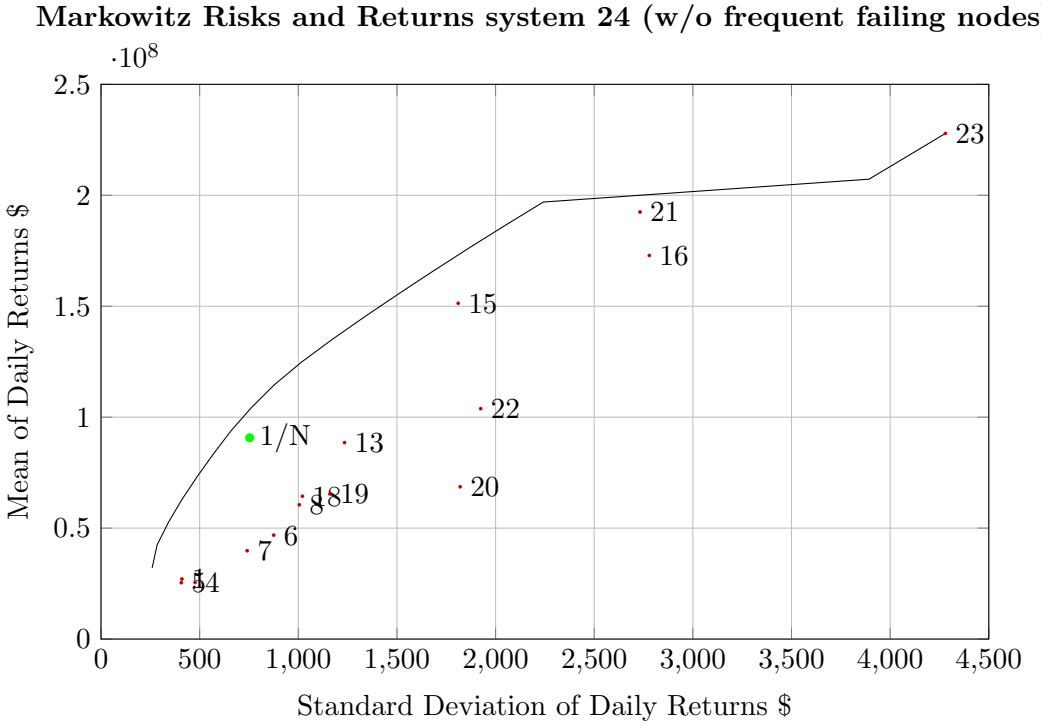

FIGURE 4.25: Markowitz efficient frontier of all system 24 nodes excluding the frequently failing node

4.4 Sub-question 4: Heuristic testing

With the understanding that the previously constructed frontier is very sensitive to change in errors in inputs, we will research if there are certain rules-of-thumb that advise the investor on if there are certain dominant investing strategies based on the topology of the grid. The main patterns that we will look for are :

Streamness: a measure on how to upstream (or downstream) a node is located, see equation 3.3.

Katz Centrality: a measure on how clustered the portfolio is. A high number means that the portfolio is relatively more clustered than a portfolio with a small number, see equation 3.5.

Pattern testing on solely variance

First we benchmark each node in terms of their standard deviation for both Katz Centrality and streamness. Note that the interdependence of each node is here neglected; we look for patterns only in the diagonal of the covariance matrix.

Although not instantly visible in Figure 4.26, there is a statistical significant (linear) correlation between Katz centrality and standard deviation of 0.62. Streamness though doesnt seem to have a statistical significant impact on the standard deviation.

We perform the same benchmarking for system 39, see Figure 4.27 to see if that correlation only holds for that particular system or is more general in nature.

In the case of system 39, we did not find any statistical significant (linear) correlation between any of the two predictors.

Pattern testing on the efficient frontier

Lastly, we test if there is any pattern in the efficient frontier; we benchmark a sample of equally spaced portfolios on the efficient frontier using the same metrics, see Figure 4.28. Note that we use the efficient frontier produced in the previous sub-question, i.e. they exclude the frequently failing nodes.

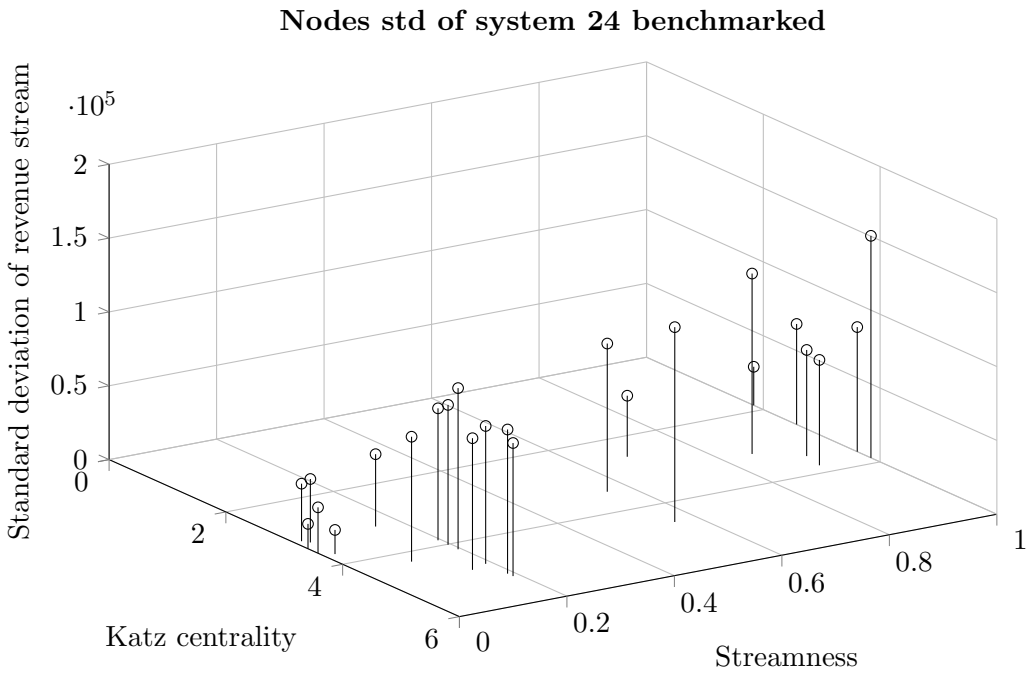

FIGURE 4.26: Standard deviation of each node of system 24 benchmarked against Streamness and Katz Centralilty

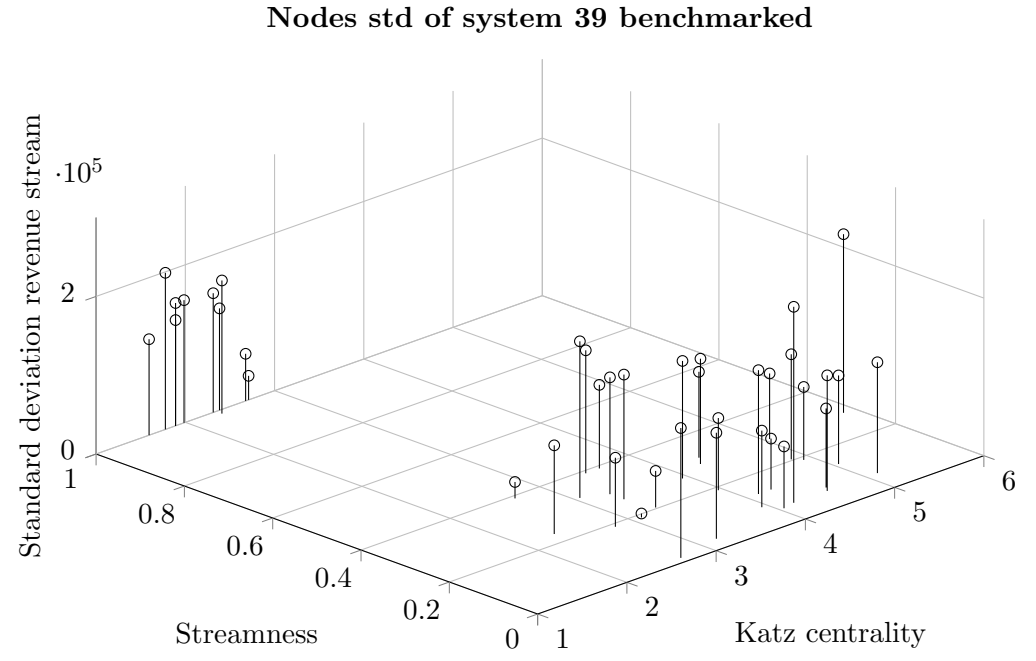

FIGURE 4.27: Standard deviation of each node of system 39 benchmarked against Streamness and Katz Centralilty

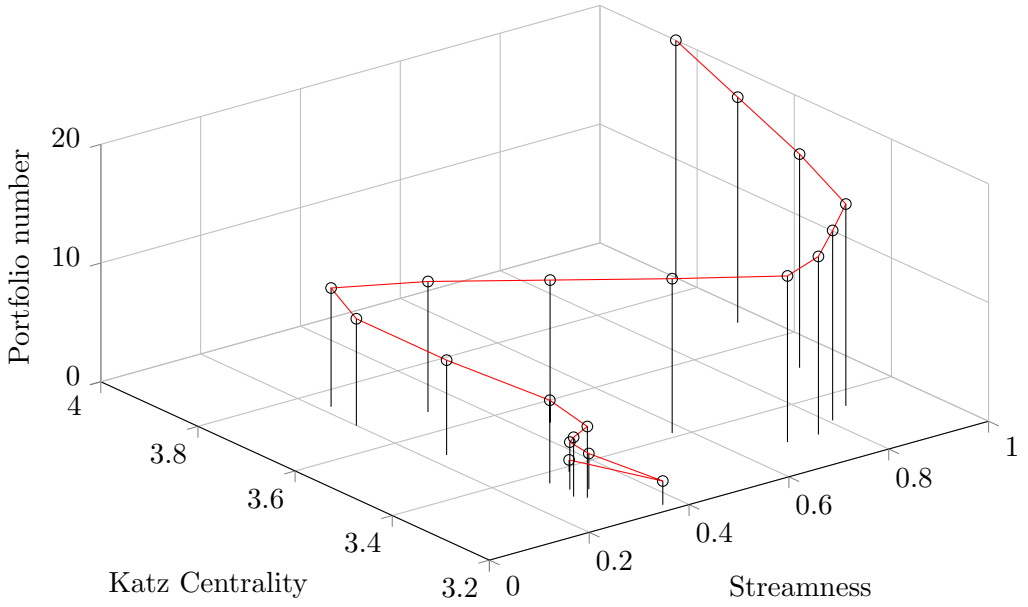

FIGURE 4.28: A sample of 20 equally-distanced portfolios on the efficient frontier. Starting from number 1 (minimum variance) to number 20 (maximum return) each portfolio of system 24 is benchmarked against their Streamness and Katz Centralilty score.

We find no statistically linear correlation for neither one of the predictors (or their second-order combinations). However, a funny S-shape appears, that we think might be the result of some quadratic relationship.

Again we find no statistical significant (linear) correlations in any of the predictors while using system 39, see Figure 4.29. However, the similar locus shape appears in this plot . The sign seems to be inversed, though, which indicates that this pattern might be pure coincidence.

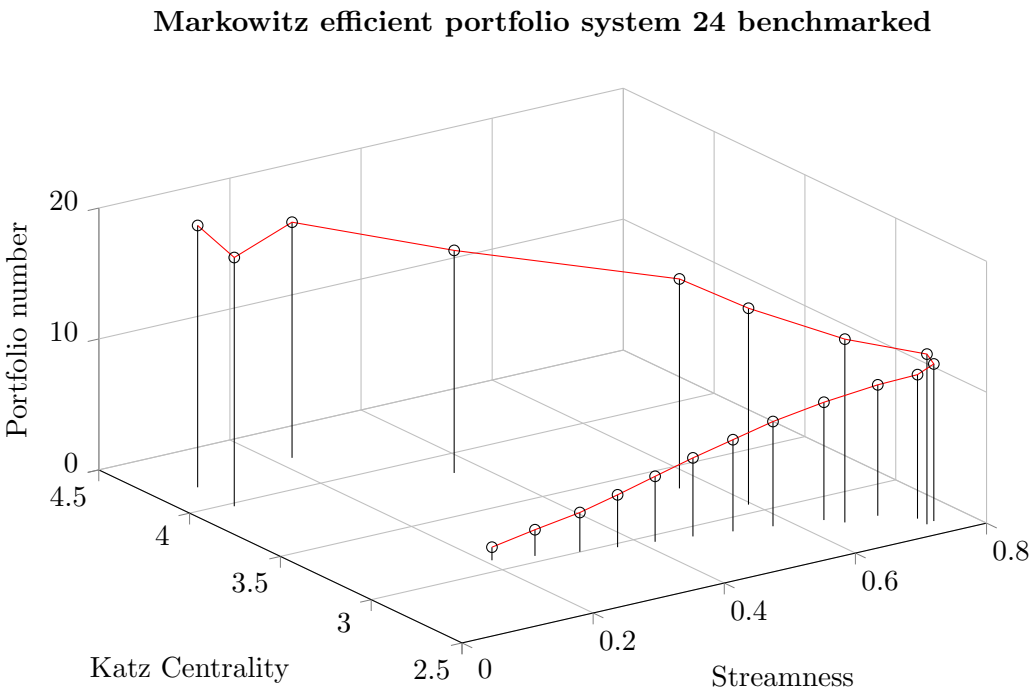

FIGURE 4.29: A sample of 20 equally-distanced portfolios on the efficient frontier. Starting from number 1 (minimum variance) to number 20 (maximum return) each portfolio of system 39 is benchmarked against their Streamness and Katz Centralilty score.

Chapter 5

Conclusion

Before finalizing this dissertation, an attempt will be made to use the previous collected results to answer the questions raised in the first chapter. After that, a final judgement will be made on the overall methodologys results.

5.1 Sub-question 1: Data generation

The data generated by the OPA model seems to suggest that there is a critical regime change over time: The Galton-Watson fit parameter increases monotonically and converges at a much higher value than at the beginning of simulation. This indicates that the systems state grows, in terms of criticality, to a certain steady-state independently from the grid model inputs. This is, of course, a consequence of the OPA models fundamental assumption that power grids, in the long term, are governed by some kind of self-organized criticality. With the knowledge that most of the grids now-a-days are running at their maximum engineering capacity, we chose therefore to sample only the last 10 years of the model simulation. This sample was the basis of comparison between the empirical and model data.

The results of the analysis of the empirical data revealed that there is enough evidence to conclude that the energy-throughput is definitely not normally distributed. Interesting enough, the empirical data is measured to be consisting of several periodic elements: both yearly and daily oscillation have identified. In respect to yearly seasonality, we

have identified two clear regimes in which all the first four moments differ by a significant difference. Leaving the present seasonality aside, we conclude that the tails are asymmetric (although the sign does seem to change per regime). Furthermore, the tails differ in their kurtosis, which could be an indicator that there are exogenous factors that take place (such as maintenance dynamic pricing) that stabilizes the device.

Comparing the operational data to the empirical data, we conclude that only Regime 1 seems to be consistent with the temporal statistics of the operational data generated by the OPA algorithm. This could be because the OPA algorithm is designed to simulate an over-stressed system, which is more in line with the higher mean and variance measured in Regime 1. In particular, most of the nodes of system 24 seem to be statistically similar to the empirical temporal statistics.

Only the left tail seems to differ substantially from the operational data. This was expected as that side of the pdf stands for device shutdown (either maintenance or outage) and the OPA model is not designed to simulate these events. This means that to successfully use the model data as input for the revenue stream construction, this element needed to be back-engineered.

All in all, we can conclude that the operational data generated by the OPA model doesnt differ too much, in terms of standardized moments, to the empirical data. This gave good faith in using the operational data, as long as the node outages (as opposed to line outages) were carefully back-engineered. Especially, because these outage events were subject to penalty fees, this last step was crucial.

5.2 Sub-question 2: Revenue stream construction

Given that the operational data is statistically similar to empirical data, we proceeded with the revenue stream construction. We generalized this stream to be in the form of Equation. The difficulty was the following two points:

1. Estimating the prices and penalty fees. These figures are very much tied to the negotiated Terms Conditions and are thus expected to vary significantly across deals.
2. The OPA model does not generate node outages (only line outages). This means that to assign penalty fees can only be done if outages are back-engineered.

For simplicity, the first point was solved by assuming a certain price (and penalty fee) for energy-transferred that is valid for all the nodes. The second-point was resolved by back-engineering node outages. The mechanism that generated was simply a hit-a-threshold function that generates an outage (and thus also a penalty) whenever the load profile hits a certain (downwards) threshold. The deviation that is associated with the threshold was chosen to make the accumulated cash flows consistent with a real business case, i.e. most nodes will end up cash flow positive, while a few might end up in the negative.

Furthermore, the cash-flows (excluding the penalties) were checked for interdependence. In particular, spatial auto-correlation was computed to see if there are dependencies between neighboring links. A plot of Morans I, using link distance as unit, revealed statistical significant autocorrelations at distance 1. After that, there was no evidence of further spatial autocorrelation. Furthermore, the plot was done for different (time step) partitions of the data sample, to check for convergence. Although, there was no evidence for convergence, there was definitely some element of continuity. This could be weak evidence of oscillations in the criticality state of the system.

To further support the hypothesis of interdependence, synchronization matrices of both grid models were constructed. The plots, which reveal which lines have tendency to fail together, revealed that there are definitely cluster of lines failing frequently. This is in line with the notion of interdependence, as a line outage will for sure affect the connected nodes volatility.

Lastly, we constructed a histogram of the 1/N portfolio of all the revenue streams associated with grid model 39. The plot shows the large (left) long tail risk that is generated by the accumulation of penalty fees. This is consistent with reality, as this kind of risk are asymmetrical in nature (there are no positive penalty fees).

5.3 Sub-question 3: Portfolio optimization

Noting that there a handful of frequently failing nodes that are the root of the negative long tail risk, our portfolio optimization started omitting these nodes. Their (negatively) asymmetrical risk is not compensated with any additional return and thus will never be in an efficient portfolio.

Excluding these nodes already made the histogram of returns skew to the right, which was of course already a great improvement. We checked if the resulting portfolio still had room for improvement in terms of efficiency. For this their Markowitz efficient frontier was constructed to see how far their distance to it was. In both cases, the 1/N portfolio (excluding frequently failing nodes) was, surprisingly enough, very close to the frontier. This suggest that the covariance matrix is quite symmetric, in the sense that diversification over all nodes almost eliminates only idiosyncratic risk. In fact, one could argue that the 1/N strategy is the most robust one (with negligible cost in efficiency), given that other strategy are susceptible to uncertainties.

5.4 Sub-question 4: Heuristic testing

Although previously concluded that the 1/N strategy (by excluding the frequently failing nodes) is the most robust (and thus preferable) one, we nevertheless checked if there are other dominant strategies present. In particular, we checked for the following two rule-of-thumb strategies:

1. Buy only upstream (or downstream) nodes 2. Buy a decentralized (or centralized) portfolio

Moreover, for additional granularity, we even checked for second-order strategies (combinations of both strategies).

Using the streamness and Katz (modified) centrality metrics, first all the nodes were benchmarked according their relative standard deviations. A statistical linear correlation between standard deviation and Katz centrality measure was found in system 24. However, a double check with system 39 (where no correlation was found) suggest that this dependency was probably idiosyncratic in nature. Furthermore, in both cases a weak (useless) negative correlation between both metrics was found, i.e. upstream nodes tend to be more decentralized and vice-versa.

After this, the same analysis was done for the efficient frontier portfolios. Taking an equally-distanced sample of the efficient frontier, produced in the previous section, each of these portfolios was benchmarked in line with the metrics. Both loci seemed to have a similar S shape, but not no other mutually consistent correlations. This suggest that,

even have checked for second-order strategy combinations, there is no evidence for a dominant heuristically investment strategy.

5.5 Final judgement on overall methodology

If a potential investor would use the presented methodology, then we have found no evidence of contradiction between using generated data as opposed to emperical data. This is assuming that an investor would have access to the relevant grid data, which because of security reasons, might be a gross assumption.

In the case of not having access to this data, then we can rely on the main conclusion of this thesis: if the frequently failing nodes could be detected (e.g. by looking at their history), then a 1/N portfolio would almost be as efficient as Markowitz recipe would prescribe. This strategy has the huge benefit, however, of being robust against measuring errors and is thus the preferred strategy.

Chapter 6

Open questions

The methodology proposed is a combination of multiple techniques existing in different sub-fields. Therefore, it is not expected that this thesis has produced any novel insight, in the sense that it has pushed the academic frontier in any way. However, if there would be someone who would be interested in this cross-field area, there are some items that are still left for additional research.

First of all, the effect of different pricing for each node has not been studied. That is to say, by setting a fixed price for each asset, we have effectively randomized the revenue streams. In a realistic scenario, this would of course not be the case, as bigger nodes would have a higher price to account for their higher CAPEX. This could have the changed the efficient frontier in total different way.

Secondly, the OPA model is not designed to simulate node outages. That is why it was needed to back-engineer this events into the revenue streams. However, from a technical point-of-view, there is no reason to not modify the OPA model to include these modes of failure in the simulation itself. Doing so, would create a much richer data sample that is perhaps reveals other dynamics that are not directly obvious to us. However, this has been to our knowledge never been done before.

Lastly, to arrive to our main conclusion, we did have to assume that frequently failing nodes are detectable. In the case that this would not be possible, or perhaps because there is a spectrum in the frequency of failure, then other risk measures than variance would be more applicable for risk management. Value at Risk or Expected Shortfall would in that case be a better risk measure.

Or perhaps using higher-cumulant portfolio theory (as they are more sensitive to the long tail risk) would be a more applicable way to capture the risk[16]? The downside of this approach, however, is that the math complicates a lot whenever the assets are correlated, i.e. dependent on each other. In that case, the Markowitz approach of doing a Gaussian fit through use of the covariance matrix is not possible anymore. To avoid this problem, Malevergne and Sornette proposed a framework where the Gaussian fit is extended to a modified Weibull distributions, which is able to capture the higher order moments, in particular when fat tails are present. In the interdependent case, however, the formulas for the multivariate fit, becomes extremely complex. Therefore they serve more as qualitatively tool than for quantitative weight prescription. However, by approximating the dependency as independent assets, it should be possible to strongly simplify the math and construct other frontiers similar to Markowitz efficient frontier[2]. This approach could reveal a better strategy than the proposed 1/N strategy if the frequently failing nodes are indeed undetectable.

Appendix A

MATLAB Code of OPA Model

```
%% OPA Model
% By Jan Kevin Pluut 15.11.16

clear all
clc
tic
% Define MATPOWER column names
define_constants

% Load en solve base case using MATPOWER
case_original=ext2int(case24_ieee_rts);
results=rundcpf(case_original);

% Set model parameters
dg=1.000049;          % Demand growth mean

cg=1.005;            % Capacity growth average

h0=0.001;            % Probability of initial line failure
h1=0.15;             % Probability of failure of overloaded line

kt=50000;               % Amount of days (simulation length)
ptol=0.1;            % Load shedding tolerance


% Add extra capicity to overloaded lines (optional)
M1=abs(results.branch(:, PF))./results.branch(:, RATE_A);
o=find(M1>1);
results.branch(o,RATE_A)=results.branch(o,RATE_A)+20;

% Get model inputs
rate=results.branch(:,RATE_A);                  % Rate of line
```

```
33
34 m=size(results.branch,1);                          % Amount of branches
35 n=size(results.bus,1);                             % Amount of nodes
36
37 typeg=results.bus(:,2)==2;                         % Get nodes types (1=Load, 2=
       Generator, 3=Reference/slack
38 typel=results.bus(:,2)==1;
39 types=results.bus(:,2)==3;
40
41 ig=find(typeg);                                    % Get generator indices
42 il=find(typel);                                    % Get load indices
43 is=find(types);
44
45 M=zeros(m,kt);
46 M(:,1)=M1;                                         % Get basecase fractional
       overloads
47
48 F=zeros(m,kt);
49 F(:,1)=results.branch(:, PF);                      % Get basecase branch flows
50
51 P=zeros(n,kt);
52 P(:,1)=-results.bus(:,PD);                         % Get basecase load
53
54 for i=1:n                                          % Get basecase generators
55     for j=1:size(results.gen,1)
56         if results.gen(j,1)==i
57             P(i,1)=P(i,1)+results.gen(j,PG);
58         end
59     end
60 end
61
62
63
64
65 % Calculate OPF matrices
66 fat=size(ig,1);                                    % Size of additional vatiables
67
68 obj=ones(1,n+fat)*100;                             % Make objective function
69 obj(ig)=0;
70 obj(n+1:end)=1;
71 obj(is)=0;
72
73 diagl=diag(typel);                                 % Make (3-4) inequality
       constraint matrix
74 matl=[diagl zeros(n,fat)];
75 matl=matl(any(matl,1),:);
76
```

```
77  diagg=diag(typeg);                                  % Make (4-8) inequality
        constraint matrix
78  matg=diagg(any(diagg,1),:);
79
80  gmax=zeros(n,kt);                                   % Get generation bus maximum
        outputs
81  for i=1:n
82      for j=1:size(results.gen,1)
83          if results.gen(j,1)==i
84              gmax(i,1)=gmax(i,1)+results.gen(j,9);
85          end
86      end
87  end
88
89
90
91  PTm=makePTDF(case_original);                        % Get PTDF matrix
92  aPTm=[PTm -PTm(:,ig)];                              % Make (1-2) inequality matrix
93
94  % Merge inequality matrices to one A matrix -> A*x<=b
95  A1=-aPTm;
96  A2=aPTm;
97  A3=-matl;
98  A4=matl;
99  A5=[-matg -matg(:,ig)];
100 A6=[matg zeros(fat)];
101 A7=[-matg zeros(fat)];
102 A8=[matg -matg(:,ig)];
103 A=[A1;A2;A3;A4;A5;A6;A7;A8];
104
105
106 % Make equality matrices Aeq*x=beq
107 Aeq=ones(1,n+fat);
108 Aeq(n+1:end)=-1;
109 beq=0;
110
111 % Set output vatiables
112 exit=ones(kt,1);                                    % Redispatch log
113 BO=zeros(kt,1);                                     % Blackout log
114 OL=zeros(kt,1);                                     % Amount of overloaded lines
115 prop=zeros(kt,m);
116
117 shed=zeros(length(il),kt);                          % Amount of load shed
118
119 % Begin OPA siumulation
120
121 ioverload=[];
122 outaged=zeros(m,kt);
```

```
123 % Slow dynamics (days)
124 for k=2:kt
125     s=0;
126     % Gemerate model random variable
127                                             % Load growth
128     o1rand=rand(m,1);                                                   % Auxillary
         outage randon vector #1
129     newcase=case_original;                                              % Load
        basecase/repair broken links
130
131     if isempty(ioverload)~=0 || sum(outaged(:,k-1))~=0
132         ioutage=find(outaged(:,k-1));                                           %
        Increase rate of overloaded/outaged lines
133         rate([ioutage;ioverload])=cg*rate([ioutage;ioverload]);
134
135     end
136
137     % Increase model variables (slow dynamics)
138     P(:,k)=unifrnd(0.6,1.4,n,1).*P(:,1).*dg^(k-1);    % Increase of node load
139     F(:,k)=PTm*P(:,k);                              % Increase branch load flow
140     M(:,k)=abs(F(:,k))./rate;                       % Calculate fractional overload
        of lines
141
142     gmax(:,k)=gmax(:,1)*dg.^(k-1);       % Increase generator capacity
143
144     % Look if overload and/or initial outage
145     ioverload=find(abs(F(:,k))./rate>0.99);
146     ioutage=find(o1rand<h0);
147     prop(k,1)=sum(o1rand<h0);
148
149     % Set fast dynamics variables
150     f=F(:,k);
151     p=P(:,k);
152
153     % Start fast dynamics (cascades)
154
155       %  Calculate power redispatch if either: (1)One or more lines are
        overloaded
156       %                                        (2)One or more lines are
157       %                                        outaged (initial outage)
158       %                                        (3)Generator flows are
159       %                                        bigger than maximum
160
161     rate1=rate;
162
163     while isempty(ioverload)==0 || isempty(ioutage)==0
164
165         % Calculate updated inequality matrix A -> A*x<=b
```

```
166
167             newcase.branch(ioutage,5)=newcase.branch(ioutage,5)/1000;
168             rate1(ioutage)=rate1(ioutage)/1000;
169             outaged(ioutage,k)=1;
170
171             nPTm=makePTDF(newcase);                    % Get PTDF matrix
172             aPTm=[nPTm -nPTm(:,ig)];                          % Make (1-2)
    inequality matrix
173
174             A1=-aPTm;
175             A2=aPTm;
176             A3=-matl;
177             A4=matl;
178             A5=[-matg -matg(:,ig)];
179             A6=[matg zeros(fat)];
180             A7=[-matg zeros(fat)];
181             A8=[matg -matg(:,ig)];
182             A=[A1;A2;A3;A4;A5;A6;A7;A8];
183
184         % Calculate updated inequality vector b -> A*x<=b
185             b=[rate1+F(:,k);rate1-F(:,k);zeros(length(il),1);-1*P(il,k);zeros(fat
    ,1);gmax(ig,k)-P(ig,k);P(ig,k); zeros(fat,1)];
186
187         % Calculate redispatch with LP
188             [lp,lpval,exit(k),lpopt]=linprog(obj,A,b,Aeq,beq);
189
190         % Transform incremental node flows backs to normal variables
191             p=lp(1:n)+P(:,k);
192             p(ig)=p(ig)-lp(n+1:end);
193         % Calcualte new branch flows
194             f=nPTm*p;
195             f=f.*(outaged(:,k)-1)*-1;
196         % Start reiteration if line was overloaded
197                 ioverload= find(abs(f)./rate1>0.99);  % Find overloaded lines
198                 o2rand=rand(m,1);                     % Auxillary outage randon
    vector #2
199                 ioutage=find((o2rand<h1).*(abs(f)./rate1>0.99));
200                 s=s+1;
201                 prop(k,s)=sum((o2rand<h1).*(abs(f)./rate1>0.99));
202                 outaged(ioutage,k)=1;
203                 % Check if there are cascades
204                 if isempty(ioutage)==1
205                     shed(:,k)=p(il)-P(il,k);
206
207
208                     BO(k)=1;                          % Set output variables
209                     OL(k)=sum(outaged(:,k));
210                     break
```

```
211                 end
212
213
214         end
215
216
217 end
218
219 toc
```

Appendix B

Grid models

```
%% IEEE 24 RTS
%% bus data
%   bus_i   type    Pd  Qd  Gs  Bs  area    Vm  Va  baseKV  zone    Vmax    Vmin
mpc.bus = [
    1   2   108 22  0   0   1   1   0   138 1   1.05    0.95;
    2   2   97  20  0   0   1   1   0   138 1   1.05    0.95;
    3   1   180 37  0   0   1   1   0   138 1   1.05    0.95;
    4   1   74  15  0   0   1   1   0   138 1   1.05    0.95;
    5   1   71  14  0   0   1   1   0   138 1   1.05    0.95;
    6   1   136 28  0   -100    2   1   0   138 1   1.05    0.95;
    7   2   125 25  0   0   2   1   0   138 1   1.05    0.95;
    8   1   171 35  0   0   2   1   0   138 1   1.05    0.95;
    9   1   175 36  0   0   1   1   0   138 1   1.05    0.95;
    10  1   195 40  0   0   2   1   0   138 1   1.05    0.95;
    11  1   0   0   0   0   3   1   0   230 1   1.05    0.95;
    12  1   0   0   0   0   3   1   0   230 1   1.05    0.95;
    13  3   265 54  0   0   3   1   0   230 1   1.05    0.95;
    14  2   194 39  0   0   3   1   0   230 1   1.05    0.95;
    15  2   317 64  0   0   4   1   0   230 1   1.05    0.95;
    16  2   100 20  0   0   4   1   0   230 1   1.05    0.95;
    17  1   0   0   0   0   4   1   0   230 1   1.05    0.95;
    18  2   333 68  0   0   4   1   0   230 1   1.05    0.95;
    19  1   181 37  0   0   3   1   0   230 1   1.05    0.95;
    20  1   128 26  0   0   3   1   0   230 1   1.05    0.95;
    21  2   0   0   0   0   4   1   0   230 1   1.05    0.95;
    22  2   0   0   0   0   4   1   0   230 1   1.05    0.95;
    23  2   0   0   0   0   3   1   0   230 1   1.05    0.95;
    24  1   0   0   0   0   4   1   0   230 1   1.05    0.95;
];

%% generator data
```

```
32 %   bus Pg  Qg  Qmax    Qmin    Vg  mBase   status  Pmax    Pmin    Pc1 Pc2
       Qc1min  Qc1max  Qc2min  Qc2max  ramp_agc    ramp_10 ramp_30 ramp_q  apf %
       Unit Code
33 mpc.gen = [
34     1   10  0   10  0   1.035   100 1   20  16  0   0   0   0   0   0   0   0   0
          0   0;  %   U20
35     1   10  0   10  0   1.035   100 1   20  16  0   0   0   0   0   0   0   0   0
          0   0;  %   U20
36     1   76  0   30  -25 1.035   100 1   76  15.2    0   0   0   0   0   0   0   0
          0   0   0;  %   U76
37     1   76  0   30  -25 1.035   100 1   76  15.2    0   0   0   0   0   0   0   0
          0   0   0;  %   U76
38     2   10  0   10  0   1.035   100 1   20  16  0   0   0   0   0   0   0   0   0
          0   0;  %   U20
39     2   10  0   10  0   1.035   100 1   20  16  0   0   0   0   0   0   0   0   0
          0   0;  %   U20
40     2   76  0   30  -25 1.035   100 1   76  15.2    0   0   0   0   0   0   0   0
          0   0   0;  %   U76
41     2   76  0   30  -25 1.035   100 1   76  15.2    0   0   0   0   0   0   0   0
          0   0   0;  %   U76
42     7   80  0   60  0   1.025   100 1   100 25  0   0   0   0   0   0   0   0   0
          0   0;  %   U100
43     7   80  0   60  0   1.025   100 1   100 25  0   0   0   0   0   0   0   0   0
          0   0;  %   U100
44     7   80  0   60  0   1.025   100 1   100 25  0   0   0   0   0   0   0   0   0
          0   0;  %   U100
45     13  95.1    0   80  0   1.02    100 1   197 69  0   0   0   0   0   0   0   0
          0   0   0;  %   U197
46     13  95.1    0   80  0   1.02    100 1   197 69  0   0   0   0   0   0   0   0
          0   0   0;  %   U197
47     13  95.1    0   80  0   1.02    100 1   197 69  0   0   0   0   0   0   0   0
          0   0   0;  %   U197
48     14  0   35.3    200 -50 0.98    100 1   0   0   0   0   0   0   0   0   0   0
          0   0   0;  %   SynCond
49     15  12  0   6   0   1.014   100 1   12  2.4 0   0   0   0   0   0   0   0   0
          0   0;  %   U12
50     15  12  0   6   0   1.014   100 1   12  2.4 0   0   0   0   0   0   0   0   0
          0   0;  %   U12
51     15  12  0   6   0   1.014   100 1   12  2.4 0   0   0   0   0   0   0   0   0
          0   0;  %   U12
52     15  12  0   6   0   1.014   100 1   12  2.4 0   0   0   0   0   0   0   0   0
          0   0;  %   U12
53     15  12  0   6   0   1.014   100 1   12  2.4 0   0   0   0   0   0   0   0   0
          0   0;  %   U12
54     15  155 0   80  -50 1.014   100 1   155 54.3    0   0   0   0   0   0   0   0
          0   0   0;  %   U155
55     16  155 0   80  -50 1.017   100 1   155 54.3    0   0   0   0   0   0   0   0
          0   0   0;  %   U155
```

```
56      18  400 0   200 -50 1.05    100 1   400 100 0   0   0   0   0   0   0   0   0
           0   0;  %   U400
57      21  400 0   200 -50 1.05    100 1   400 100 0   0   0   0   0   0   0   0   0
           0   0;  %   U400
58      22  50  0   16  -10 1.05    100 1   50  10  0   0   0   0   0   0   0   0   0
           0   0;  %   U50
59      22  50  0   16  -10 1.05    100 1   50  10  0   0   0   0   0   0   0   0   0
           0   0;  %   U50
60      22  50  0   16  -10 1.05    100 1   50  10  0   0   0   0   0   0   0   0   0
           0   0;  %   U50
61      22  50  0   16  -10 1.05    100 1   50  10  0   0   0   0   0   0   0   0   0
           0   0;  %   U50
62      22  50  0   16  -10 1.05    100 1   50  10  0   0   0   0   0   0   0   0   0
           0   0;  %   U50
63      22  50  0   16  -10 1.05    100 1   50  10  0   0   0   0   0   0   0   0   0
           0   0;  %   U50
64      23  155 0   80  -50 1.05    100 1   155 54.3    0   0   0   0   0   0   0   0
           0   0   0;  %   U155
65      23  155 0   80  -50 1.05    100 1   155 54.3    0   0   0   0   0   0   0   0
           0   0   0;  %   U155
66      23  350 0   150 -25 1.05    100 1   350 140 0   0   0   0   0   0   0   0   0
           0   0;  %   U350
67  ];
68
69  %% branch data
70  %   fbus    tbus    r   x   b   rateA   rateB   rateC   ratio   angle   status
        angmin  angmax
71  mpc.branch = [
72      1   2   0.0026  0.0139  0.4611  175 250 200 0   0   1   -360    360;
73      1   3   0.0546  0.2112  0.0572  175 208 220 0   0   1   -360    360;
74      1   5   0.0218  0.0845  0.0229  175 208 220 0   0   1   -360    360;
75      2   4   0.0328  0.1267  0.0343  175 208 220 0   0   1   -360    360;
76      2   6   0.0497  0.192   0.052   175 208 220 0   0   1   -360    360;
77      3   9   0.0308  0.119   0.0322  175 208 220 0   0   1   -360    360;
78      3   24  0.0023  0.0839  0   400 510 600 1.03    0   1   -360    360;
79      4   9   0.0268  0.1037  0.0281  175 208 220 0   0   1   -360    360;
80      5   10  0.0228  0.0883  0.0239  175 208 220 0   0   1   -360    360;
81      6   10  0.0139  0.0605  2.459   175 193 200 0   0   1   -360    360;
82      7   8   0.0159  0.0614  0.0166  175 208 220 0   0   1   -360    360;
83      8   9   0.0427  0.1651  0.0447  175 208 220 0   0   1   -360    360;
84      8   10  0.0427  0.1651  0.0447  175 208 220 0   0   1   -360    360;
85      9   11  0.0023  0.0839  0   400 510 600 1.03    0   1   -360    360;
86      9   12  0.0023  0.0839  0   400 510 600 1.03    0   1   -360    360;
87      10  11  0.0023  0.0839  0   400 510 600 1.02    0   1   -360    360;
88      10  12  0.0023  0.0839  0   400 510 600 1.02    0   1   -360    360;
89      11  13  0.0061  0.0476  0.0999  500 600 625 0   0   1   -360    360;
90      11  14  0.0054  0.0418  0.0879  500 625 625 0   0   1   -360    360;
91      12  13  0.0061  0.0476  0.0999  500 625 625 0   0   1   -360    360;
```

```
92      12   23   0.0124   0.0966   0.203    500 625 625 0   0   1   -360    360;
93      13   23   0.0111   0.0865   0.1818   500 625 625 0   0   1   -360    360;
94      14   16   0.005    0.0389   0.0818   500 625 625 0   0   1   -360    360;
95      15   16   0.0022   0.0173   0.0364   500 600 625 0   0   1   -360    360;
96      15   21   0.0063   0.049    0.103    500 600 625 0   0   1   -360    360;
97      15   21   0.0063   0.049    0.103    500 600 625 0   0   1   -360    360;
98      15   24   0.0067   0.0519   0.1091   500 600 625 0   0   1   -360    360;
99      16   17   0.0033   0.0259   0.0545   500 600 625 0   0   1   -360    360;
100     16   19   0.003    0.0231   0.0485   500 600 625 0   0   1   -360    360;
101     17   18   0.0018   0.0144   0.0303   500 600 625 0   0   1   -360    360;
102     17   22   0.0135   0.1053   0.2212   500 600 625 0   0   1   -360    360;
103     18   21   0.0033   0.0259   0.0545   500 600 625 0   0   1   -360    360;
104     18   21   0.0033   0.0259   0.0545   500 600 625 0   0   1   -360    360;
105     19   20   0.0051   0.0396   0.0833   500 600 625 0   0   1   -360    360;
106     19   20   0.0051   0.0396   0.0833   500 600 625 0   0   1   -360    360;
107     20   23   0.0028   0.0216   0.0455   500 600 625 0   0   1   -360    360;
108     20   23   0.0028   0.0216   0.0455   500 600 625 0   0   1   -360    360;
109     21   22   0.0087   0.0678   0.1424   500 600 625 0   0   1   -360    360;
110 ];
111
112 %% New England 39 System
113 %% bus data
114 %   bus_i   type    Pd   Qd   Gs   Bs   area    Vm   Va   baseKV   zone    Vmax    Vmin
115 mpc.bus = [
116     1   1   97.6    44.2    0   0   2   1.0393836   -13.536602  345 1   1.06
        0.94;
117     2   1   0   0   0   0   2   1.0484941   -9.7852666  345 1   1.06    0.94;
118     3   1   322 2.4 0   0   2   1.0307077   -12.276384  345 1   1.06    0.94;
119     4   1   500 184 0   0   1   1.00446 -12.626734  345 1   1.06    0.94;
120     5   1   0   0   0   0   1   1.0060063   -11.192339  345 1   1.06    0.94;
121     6   1   0   0   0   0   1   1.0082256   -10.40833   345 1   1.06    0.94;
122     7   1   233.8   84  0   0   1   0.99839728  -12.755626  345 1   1.06    0.94;
123     8   1   522 176.6   0   0   1   0.99787232  -13.335844  345 1   1.06    0.94;
124     9   1   6.5 -66.6   0   0   1   1.038332    -14.178442  345 1   1.06    0.94;
125     10  1   0   0   0   0   1   1.0178431   -8.170875   345 1   1.06    0.94;
126     11  1   0   0   0   0   1   1.0133858   -8.9369663  345 1   1.06    0.94;
127     12  1   8.53    88  0   0   1   1.000815    -8.9988236  345 1   1.06    0.94;
128     13  1   0   0   0   0   1   1.014923    -8.9299272  345 1   1.06    0.94;
129     14  1   0   0   0   0   1   1.012319    -10.715295  345 1   1.06    0.94;
130     15  1   320 153 0   0   3   1.0161854   -11.345399  345 1   1.06    0.94;
131     16  1   329 32.3    0   0   3   1.0325203   -10.033348  345 1   1.06    0.94;
132     17  1   0   0   0   0   2   1.0342365   -11.116436  345 1   1.06    0.94;
133     18  1   158 30  0   0   2   1.0315726   -11.986168  345 1   1.06    0.94;
134     19  1   0   0   0   0   3   1.0501068   -5.4100729  345 1   1.06    0.94;
135     20  1   680 103 0   0   3   0.99101054  -6.8211783  345 1   1.06    0.94;
136     21  1   274 115 0   0   3   1.0323192   -7.6287461  345 1   1.06    0.94;
137     22  1   0   0   0   0   3   1.0501427   -3.1831199  345 1   1.06    0.94;
```

```
138     23  1   247.5   84.6    0   0   3   1.0451451   -3.3812763  345 1   1.06
        0.94;
139     24  1   308.6   -92.2   0   0   3   1.038001    -9.9137585  345 1   1.06
        0.94;
140     25  1   224 47.2    0   0   2   1.0576827   -8.3692354  345 1   1.06    0.94;
141     26  1   139 17  0   0   2   1.0525613   -9.4387696  345 1   1.06    0.94;
142     27  1   281 75.5    0   0   2   1.0383449   -11.362152  345 1   1.06    0.94;
143     28  1   206 27.6    0   0   3   1.0503737   -5.9283592  345 1   1.06    0.94;
144     29  1   283.5   26.9    0   0   3   1.0501149   -3.1698741  345 1   1.06
        0.94;
145     30  2   0   0   0   0   2   1.0499  -7.3704746  345 1   1.06    0.94;
146     31  3   9.2 4.6 0   0   1   0.982   0   345 1   1.06    0.94;
147     32  2   0   0   0   0   1   0.9841  -0.1884374  345 1   1.06    0.94;
148     33  2   0   0   0   0   3   0.9972  -0.19317445 345 1   1.06    0.94;
149     34  2   0   0   0   0   3   1.0123  -1.631119   345 1   1.06    0.94;
150     35  2   0   0   0   0   3   1.0494  1.7765069   345 1   1.06    0.94;
151     36  2   0   0   0   0   3   1.0636  4.4684374   345 1   1.06    0.94;
152     37  2   0   0   0   0   2   1.0275  -1.5828988  345 1   1.06    0.94;
153     38  2   0   0   0   0   3   1.0265  3.8928177   345 1   1.06    0.94;
154     39  2   1104    250 0   0   1   1.03    -14.535256  345 1   1.06    0.94;
155 ];
156
157 %% generator data
158 %   bus Pg  Qg  Qmax    Qmin    Vg  mBase   status  Pmax    Pmin    Pc1 Pc2
        Qc1min  Qc1max  Qc2min  Qc2max  ramp_agc    ramp_10 ramp_30 ramp_q  apf
159 mpc.gen = [
160     30  250 161.762 400 140 1.0499  100 1   1040    0   0   0   0   0   0   0   0
          0   0   0   0;
161     31  677.871 221.574 300 -100    0.982   100 1   646 0   0   0   0   0   0   0
          0   0   0   0   0;
162     32  650 206.965 300 150 0.9841  100 1   725 0   0   0   0   0   0   0   0   0
          0   0   0;
163     33  632 108.293 250 0   0.9972  100 1   652 0   0   0   0   0   0   0   0   0
          0   0   0;
164     34  508 166.688 167 0   1.0123  100 1   508 0   0   0   0   0   0   0   0   0
          0   0   0;
165     35  650 210.661 300 -100    1.0494  100 1   687 0   0   0   0   0   0   0   0   0
          0   0   0   0;
166     36  560 100.165 240 0   1.0636  100 1   580 0   0   0   0   0   0   0   0   0
          0   0   0;
167     37  540 -1.36945    250 0   1.0275  100 1   564 0   0   0   0   0   0   0   0   0
          0   0   0   0;
168     38  830 21.7327 300 -150    1.0265  100 1   865 0   0   0   0   0   0   0   0   0
          0   0   0   0;
169     39  1000    78.4674 300 -100    1.03    100 1   1100    0   0   0   0   0   0   0
          0   0   0   0   0   0;
170 ];
171
```

```
%% branch data
%   fbus    tbus    r   x   b   rateA   rateB   rateC   ratio   angle   status
    angmin  angmax
mpc.branch = [
    1   2   0.0035  0.0411  0.6987  600 600 600 0   0   1   -360    360;
    1   39  0.001   0.025   0.75    1000    1000    1000    0   0   1   -360
    360;
    2   3   0.0013  0.0151  0.2572  500 500 500 0   0   1   -360    360;
    2   25  0.007   0.0086  0.146   500 500 500 0   0   1   -360    360;
    2   30  0   0.0181  0   900 900 2500    1.025   0   1   -360    360;
    3   4   0.0013  0.0213  0.2214  500 500 500 0   0   1   -360    360;
    3   18  0.0011  0.0133  0.2138  500 500 500 0   0   1   -360    360;
    4   5   0.0008  0.0128  0.1342  600 600 600 0   0   1   -360    360;
    4   14  0.0008  0.0129  0.1382  500 500 500 0   0   1   -360    360;
    5   6   0.0002  0.0026  0.0434  1200    1200    1200    0   0   1   -360
    360;
    5   8   0.0008  0.0112  0.1476  900 900 900 0   0   1   -360    360;
    6   7   0.0006  0.0092  0.113   900 900 900 0   0   1   -360    360;
    6   11  0.0007  0.0082  0.1389  480 480 480 0   0   1   -360    360;
    6   31  0   0.025   0   1800    1800    1800    1.07    0   1   -360    360;
    7   8   0.0004  0.0046  0.078   900 900 900 0   0   1   -360    360;
    8   9   0.0023  0.0363  0.3804  900 900 900 0   0   1   -360    360;
    9   39  0.001   0.025   1.2 900 900 900 0   0   1   -360    360;
    10  11  0.0004  0.0043  0.0729  600 600 600 0   0   1   -360    360;
    10  13  0.0004  0.0043  0.0729  600 600 600 0   0   1   -360    360;
    10  32  0   0.02    0   900 900 2500    1.07    0   1   -360    360;
    12  11  0.0016  0.0435  0   500 500 500 1.006   0   1   -360    360;
    12  13  0.0016  0.0435  0   500 500 500 1.006   0   1   -360    360;
    13  14  0.0009  0.0101  0.1723  600 600 600 0   0   1   -360    360;
    14  15  0.0018  0.0217  0.366   600 600 600 0   0   1   -360    360;
    15  16  0.0009  0.0094  0.171   600 600 600 0   0   1   -360    360;
    16  17  0.0007  0.0089  0.1342  600 600 600 0   0   1   -360    360;
    16  19  0.0016  0.0195  0.304   600 600 2500    0   0   1   -360    360;
    16  21  0.0008  0.0135  0.2548  600 600 600 0   0   1   -360    360;
    16  24  0.0003  0.0059  0.068   600 600 600 0   0   1   -360    360;
    17  18  0.0007  0.0082  0.1319  600 600 600 0   0   1   -360    360;
    17  27  0.0013  0.0173  0.3216  600 600 600 0   0   1   -360    360;
    19  20  0.0007  0.0138  0   900 900 2500    1.06    0   1   -360    360;
    19  33  0.0007  0.0142  0   900 900 2500    1.07    0   1   -360    360;
    20  34  0.0009  0.018   0   900 900 2500    1.009   0   1   -360    360;
    21  22  0.0008  0.014   0.2565  900 900 900 0   0   1   -360    360;
    22  23  0.0006  0.0096  0.1846  600 600 600 0   0   1   -360    360;
    22  35  0   0.0143  0   900 900 2500    1.025   0   1   -360    360;
    23  24  0.0022  0.035   0.361   600 600 600 0   0   1   -360    360;
    23  36  0.0005  0.0272  0   900 900 2500    1   0   1   -360    360;
    25  26  0.0032  0.0323  0.531   600 600 600 0   0   1   -360    360;
    25  37  0.0006  0.0232  0   900 900 2500    1.025   0   1   -360    360;
    26  27  0.0014  0.0147  0.2396  600 600 600 0   0   1   -360    360;
```

```
217     26  28  0.0043  0.0474  0.7802  600 600 600 0   0   1   -360    360;
218     26  29  0.0057  0.0625  1.029   600 600 600 0   0   1   -360    360;
219     28  29  0.0014  0.0151  0.249   600 600 600 0   0   1   -360    360;
220     29  38  0.0008  0.0156  0   1200    1200    2500    1.025   0   1   -360
        360;
221 ];
```

Bibliography

[1] The risk profile of infrastructure investments: Challenging conventional wisdom. *CFA Digest*, 43(1):117–119, 2013.

[2] J.V. Andersen and D. Sornette. Have your cake and eat it, too: Increasing returns while lowering large risks! *The Journal of Risk Finance*, 2(3):70–82, 2001.

[3] B.A. Carreras, D.E. Newman, and I. Dobson. Determining the vulnerabilities of the power transmission system. In *45th Hawaii International Conference on System Sciences*, 2012.

[4] Benjamin A. Carreras, David E. Newman, Ian Dobson, and Naga S. Degala. Validating opa with wecc data. In *46th Hawaii International Conference on System Sciences*, 2013.

[5] Hui Ren ; Ian Dobson ; Benjamin A. Carreras. Long-term effect of the n-1 criterion on cascading line outages in an evolving power transmission grid. In *IEEE Transactions on Power Systems*, volume 23, 2008.

[6] Hung-Po Chao and Stephen Peck. A market mechanism for electric power transmission. *Journal of Regulatory Economics*, 10(1):25–59, 1996.

[7] A.D. Cliff and J.K. Ord. *Spatial processes : models & applications / A.D.* 1981.

[8] Li D, Jiang Y, Kang R, and Havlin S. Spatial correlation analysis of cascading failures: congestions and blackouts. *Sci Rep.*, 4:5381–5381, 2014.

[9] I. Dobson, K.R. Wierzbicki, B.A. Carreras, V.E. Lynch, and D.E. Newman. An estimator of propagation of cascading failure. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, 2006.

[10] Vladyslav A. Golyk. Self-organized criticality.

[11] Bei Gou, Hui Zheng, Weibiao Wu, and Xingbin Yu. Probability distribution of power system blackouts. In *Power Engineering Society General Meeting, 2007. IEEE*, pages 1–8, 2007.

[12] Paul Hines, Jay Apt, and Sarosh Talukdar. Trends in the history of large blackouts in the united states. In *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, 2008.

[13] Georg Inderst. Infrastructure as an asset class. *EIB Papers*, 15(1):70–105.

[14] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

[15] Deepak Paramashivan Kaundinya, P. Balachandra, and N.H. Ravindranath. Grid-connected versus stand-alone energy systems for decentralized powera review of literature. *Renewable and Sustainable Energy Reviews*, 13(8):2041–2050, 2009.

[16] Yannick Malevergne and Didier Sornette. Higher-moment portfolio theory. *The Journal of Portfolio Management*, 31(4):49–55, 2005.

[17] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[18] Amulya K.N. Reddy. Barriers to improvements in energy efficiency. *Energy Policy*, 19(10):953 – 961, 1991.

[19] Hui Ren, Qunjie Chen, Fei Wang, David Watts, and Chenjun Sun. Critical transitions analysis in self-organized power systems by opa. In *2016 IEEE International Conference on Power System Technology (POWERCON)*, 2007.

[20] Sornette. Sandpile model. *Encyclopedia of Nonlinear Science*, pages 821–823, 2005.

[21] Xiaofeng Weng, Yiguang Hong, Ancheng Xue, and Shengwei Mei. Failure analysis on china power grid based on power law. *Journal of Control Theory and Applications*, 4(3):235–238, 2006.

[22] Rafał Weron and Ingve Simonsen. *Blackouts, risk, and fat-tailed distributions*, pages 215–219. Springer Tokyo, Tokyo, 2006.

[23] Per Bak; Chao Tang; Kurt Wiesenfeld. Self-organized criticality: an explanation of 1/ noise. *Physical Review Letters*, 59(4):381384, 1987.

[24] Xiaohui Ni; Yannick Malevergne; Didier Sornette; Peter Woehrmann. Robust reverse engineering of cross-sectional returns and improved portfolio allocation performance using the capm. *The Journal of Portfolio Management*, 37(4):76–85, 2011.

[25] Ray Daniel Zimmerman, Carlos Edmundo Murillo-Sanchez, and Robert John Thomas. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. In *IEEE Transactions on Power Systems*, volume 26, 2011.