



Norwegian University of
Science and Technology

Simplification algorithms for large-scale power system transmission grids

Tomas Kruijer

Innovative Sustainable Energy Engineering

Submission date: September 2015

Supervisor: Olav B Fosso, ELKRAFT

Co-supervisor: Christian Skar, EPT
Francesco Dalla Longa, ECN

Norwegian University of Science and Technology
Department of Electric Power Engineering

Master thesis report

Simplification algorithms for large-scale power system transmission grids

Conducted with the purpose to obtain the degree of:

MSc Innovative Sustainable Energy Engineering

Under supervision of:

Prof. Dr. Olav Barte Fosso

Department of Electric Power Engineering, Norwegian University of Science and Technology

Dipl.-Ing. André Ortner

Energy Economics Group, Technical University of Vienna

Dipl.-Ing. Dr. Gustav Resch

Energy Economics Group, Technical University of Vienna

Submitted to the Department of Electric Power Engineering

by

Tomas Kruijer

Student dual MSc Innovative Sustainable Energy Engineering

Abstract

The simulation of large-scale power system models including transmission grid representations is limited by available computation power and time. Therefore, the reduction of transmission grid models is of paramount importance. This report proposes and tests a new method of reducing power grids. This method allows to reduce an existing transmission grid model to any desired size, however, at the cost of accepting increasing levels of inaccuracy.

An algorithm is developed that reforms the total number of nodes in the full grid representation into a smaller number of clusters, which are then connected by equivalent power lines. The algorithm is designed in such way that those power lines remain in the transmission grid that are likely to form a bottleneck, thus restraining power flows.

The accuracy of this method is measured by comparing the power flows in the reduced power grid to the flows from the original grid. The power flows of the full and reduced model are calculated by applying linear approaches based on PTDF matrices. PTDF matrices are commonly used in transmission grid analysis, linking node injections to power flows. The power flows are calculated based on a pre-defined set of injections, which represent cases of realistic power plant dispatches. The PTDF matrix for the reduced matrix is derived with the method proposed by Shi et al [1]. The reduced matrix is operating point dependent, based on a set of reference injections.

The results show that every country can be reduced up to 37.5% of its original size, when maintaining an allowable error of 20% of the available transfer capacity of a power line. Most countries can even be reduced further before they exceed the set accuracy benchmark.

In addition to this, the report researches whether guidelines can be identified to which extend power grids can be reduced within preset limits of accuracy. Power grids in countries have different properties like topologies and grid characteristics, possibly leading to differing error behavior. The results show that no clear relation can be identified between the properties of a country's power grid and its error behavior, but is dependent on case specific

situations in which node injections play a particularly decisive role. As the node capacity and node generation are random though, so are the occurring errors.

Finally, the relation between accuracy and gained computation time in optimization simulations is identified. The relation between the number of variables in an optimization model and required computation time is exponential. This relation shows that without exceeding the preset boundaries of accuracy significant gains in computation time can be acquired. A grid reduced to 22.5% of its original sizes does not exceed the error limit in power lines, while the computation time reduces by a factor of 261.

Table of Contents

Abstract	1
1 Introduction	7
2 Methodology	10
2.1 Description of the cluster process.....	10
2.2 Calculation of PTDF matrices	13
2.3 PTDF calculation of the reduced grid	15
2.4 Power flow error calculation.....	17
2.5 Calculating computation time.....	20
3 Application of the simplification algorithm to the ENTSO-E transmission grid model	22
3.1 Grouping countries into zones	22
3.2 Modeling the power grid	23
3.3 Introducing transformers in the model	24
3.4 Calculation of node injections.....	26
4 Results	32
4.1 Validation cluster algorithm.....	32
4.2 Reduction process of the case study	36
4.3 Error analysis.....	38
4.4 Time vs. Accuracy balance	44
5 Critical reflection.....	46
6 Conclusion	47
7 Literature overview	49

List of figures

- Figure 1. Schematic overview of the cluster process.....11
- Figure 2. Process overview of the clustering process.12
- Figure 3. Multiple lines in the full grid representation are combined to a singular line18
- Figure 4. Map of zone division over countries.23
- Figure 5. Schematic overview of the process reforming a node with several voltages connect to it ...25
- Figure 6. Example of the assignment process of plants to nodes for Macedonia.....26
- Figure 7. Diagram exemplifying node PV generation.28
- Figure 8. Example (Austria) of the variability of the production of hydro run-of-river in a country29
- Figure 9. The 12-node test grid for which the clustering algorithm is tested.32
- Figure 10. Reduction process of a 12 node example grid.35
- Figure 11. The full grid at 100%, before any reduction has occurred.....36
- Figure 12. The grid reduced to 50% of its original size.37
- Figure 13. The grid reduced to 0%, thus becoming a one cluster per zone situation.37
- Figure 14. The error diagram per zone, for reduction from 100% to 0%.....39
- Figure 15. The error diagram per zone, for reduction from 25% to 0%.39
- Figure 16. Two opposing cross boarder lines are clustered to one line.40
- Figure 17. Allowable reduction per zone.41
- Figure 18. Reactances of power lines of the UK ordered according to their values.....42
- Figure 19. The cutoff reactance vs. the allowable reduction in each zone.....42
- Figure 20. The relation between reduction and the average error of the line flows44

Nomenclature (According to order of appearance)

Symbol	Units	Description
ξ	%	Granularity of the grid, as a percentage of the full grid size
$n_{nodes,k}$		Total number of nodes in zone k
$n_{clusters,k}$		Number of clusters in the reduced grid in zone k
c_n	Ohm	Condition value of line n
X_n	Ohm	Reactance of line n
$\Delta P_{inj,i,j}$	MW	Difference in injection between two connected nodes i and j
TC_n	MW	Thermal capacity of line n
γ	Ohm	Clustering threshold
N		Number of nodes in the full grid excluding slack node
N_R		Number of clusters in the reduced grid excluding slack node
L		Number of lines in the full grid
C		Incidence matrix describing connection between nodes and lines
B_{branch}		Susceptance matrix describing connection between reactance of lines and nodes
B_{bus}		Susceptance matrix describing connection between reactance of lines and nodes
ϑ	°	Vector containing the voltage angles in the set of nodes
P_{inj}	MW	Vector containing the injections in the set of nodes
P_{flow}	MW	Vector containing the flow in the set of lines
Φ		Power transfer distribution factor matrix
Π_{bus}		Matrix describing relation between nodes and clusters
$(P_{inj})_R$	MW	Vector containing the injections in the set of clusters
L_R		Number of equivalent lines in the reduced grid
ψ		Matrix describing relation between cluster injections and the inter-cluster flows
Π_{flow}		Matrix describing relation between the full grid lines and reduced grid lines
$\mathbf{P}_{flow}^{inter-zonal}$	MW	Vector containing the inter-cluster power flow
Φ_R		Reduced power transfer distribution factor matrix
$E_{abs,i}$	%	Absolute error of a reduced power line i
$E_{rel,i}$	%	Relative error of a reduced power line i
$ATC_{R,i}$	MW	Available transfer capacity of a reduced power line i

$\bar{E}_{rel,i}$	%	Average relative error of a reduced power line i over all time steps
$number_{lines R,k}$		Number of inter-zonal lines in the reduced grid for zone k
CT	sec	Required computation time for a model of a certain size
n_{var}		Number of variables (clusters) in a reduced grid
\bar{E}	%	Average error over all inter-zonal power lines and time steps
$E_{i,i,inter-zonal}$	%	Error in inter-zonal power line i
$n_{i,i,inter-zonal}$		Number of inter-zonal power lines in the reduced transmission grid
X_i	Ohm	Reactance of a power line i
x_p	Ohm/m	Reactance per meter for a power line with voltage level p
l_i	km	Length of power line i
RL_k		Number of removed lines in zone k at the point of allowable reduction

1 Introduction

Electricity networks will be challenged more and more in the future. The growth of renewable energy capacity will cause increasing volatility in electricity production. This volatility will partly have to be absorbed by increased flexibility in the grid. On top of this, energy markets will become further internationally integrated, leading to more cross-country power trading. The capacity of power lines is restricted by their respective thermal limits though. To obtain insight in what investment- or policy decisions are required in order to keep the power grid reliable and stable under those changing circumstances, power grid simulations can be a useful tool. However, model simulations of large power grid systems often exceed available computation power or desired computation time. The reduction of computation requirements of grid models is therefore of paramount importance in power grid modeling. Various methods have been developed for several dedicated modeling purposes, which can be divided in static and dynamic methods. Static reduction has been described as ‘the reduced model represents a snapshot of the system [...] appropriate for power flow calculations, for operational and planning analysis.’ [1] The dynamic reduction method ‘is used for the analysis of dynamic effects, such as ‘(a) large scale power system off-line transient stability analysis, (b) large scale power system off-line dynamic stability analysis, (c) large scale power system on-line security assessment.’ [2]

As this paper is focusing on power flow calculations, exclusively static reduction methods will be applied. Typically, the nodes (or busses) of a grid system are simplified to the level of countries or zones. In the last decade, using net transfer capacities (NTC’s) has been a widely applied method of restricting the available trade capacity between countries. Although it is computationally efficient, using NTC’s has several flaws. Amongst others, the internal constraints within a country are just taken into account indirectly via incorporating a worst case scenario in the calculation of the NTC value. Thus, the power grid will not be used most efficiently at any time [3]. A power system can be reduced to any desired level of detail though, ranging from full-scale models describing all power lines and nodes to the one-node-per-country scale. But for any level of detail, reduction goes at the cost of accuracy. There are very few studies on the validity and the benefits of using reduction techniques in large-scale models though. This report will contribute to solve the lack of insight in this field, by answering to the following three research questions:

1. Power grid models can be clustered (reduced) to a certain degree, ranging from full-scale, unreduced models to models with just few nodes. What is the relation between granularity and error size?
2. Topologies of grid models vary in several characteristics. Can we reduce the maximum errors of reduction algorithms when we make use of certain characteristics of the particular power grid?
3. The reduction of power grid models will lead to improved computation times. What is the relation between error size and required calculation time?

For the calculation of the power flows of full-sized and reduced grids several methods are available. Among the static reduction methods belong the WARD method developed by Ward [4] and the REI method by Dimo [5], both improve calculation time by reducing the size of the nodal admittance matrix. Alternatively, market based reduction techniques based on the local marginal prices have been proposed by e.g. Singh et al [6]. In this report however, the power flows are estimated using power transfer distribution factors (PTDF) of the aggregated equivalent network. Various scholars have used and tested PTDF's (e.g. Sheng [7] and Liu [8]). Both the power flows for the full and reduced grids are calculated using a PTDF matrix. A PTDF matrix describes the relation between power injections in system nodes and the line flows between them. PTDF's offer several options to represent the grid; the transportation, AC power flow and DC power flow formulation. The AC power flow formulation is the most accurate, while the transportation formulation requires the fewest computation time. However, it is widely recognized that the DC power flow formulation offers both significantly reduced computation time while still delivering relatively accurate results under normal operating conditions [4]. Therefore, the model used in this report will apply a DC power flow approach. In the DC power flow formulation the relation between injections and power flows becomes linear.

A number of methods have been proposed for reduced PTDF calculations of large power systems. One has been brought forward by Shi et al. [1], from which the results showed that '(a) under the base case, the equivalent-system power flows exactly match those calculated using the full-network-model (b) as the operating conditions change, errors in line flows are minimized (c) the method is more computationally efficient than other bus aggregation methods proposed before' [1]. This method has only been applied on a small, six node

model though. In this report the method will be applied on a large scale grid model of the European transmission network. Additionally, the nodes in [1] are assigned manually to clusters, whereas in this report the nodes are conditionally clustered according to an algorithm. Only lines that are least likely to form bottlenecks in the grid are removed from the model, as they do not constrain power flows.

The resulting power flows of the reduced models are compared to the results of the unreduced grid, from which the error is being calculated for each inter-cluster flow. We calculate errors based on a sample of 960 characteristic hours, in which the injections in the nodes vary over time while the topology remains fixed.

A deeper analysis will be conducted on the error behavior of each country to answer the second research question. As every country has its own typical characteristics, it is likely that different error patterns can be identified. Countries differ in topologies of the grid, voltage levels, and interconnectivity. Linking the allowable reduction of a country to those characteristics will indicate what characteristics are important to consider when performing grid reductions.

Based on the simplified (or reduced) grid model we determine the achievable computation time reduction for any particular model size through applying an optimum power flow analysis for both, the full and the simplified grid model. From this, the relation between accuracy and computation time can be derived, which will contribute to answering the third research question. This relation will serve researchers to find the proper balance between those two factors in their future modeling.

The research questions are tested on a model of the ENTSO-E power grid, including Turkey and the North-African countries.

2 Methodology

In this section a brief overview of the applied methodology is given. We start with a description of the cluster algorithm in section 2.1, i.e. according to what criteria nodes are aggregated. All power flows of the original system and reduced systems are calculated via the use of PTDF matrices. Section 2.2 will therefore describe how the PTDF matrix for the full grid representation is derived. Then, section 2.3 explains how the PTDF matrix of the reduced grid is calculated, based on the theory of [1]. With the power flows for the original and for the reduced grid, the error caused by the reduction process is calculated. The method for this is described in section 2.4. Finally, section 2.5 explains how the relation between required computation time and accuracy is established.

2.1 Description of the cluster process

The clustering process reduces the power grid by aggregating two nodes into one cluster and removing the existing power line between them in the process (see figure 1). The clustering process is performed separately for each zone. A zone in this report is defined as a country or a designated set of countries. In section 3.1 all zones for this case study are defined. Inter-zonal lines cannot be removed in the clustering process; only lines with both nodes located within the respective zone are eligible for removing.

The extent to how far zones are reduced is dependent on the particular desired zone grid size, which can be defined by the targeted granularity denoted by ξ . The granularity is a percentage ranging from 0% to 100%, indicating the targeted model size. A ξ of 100% represents the full grid in a zone without any reduction, whereas a percentage of 0% represents a one-cluster-per-zone situation by definition. From the targeted granularity ξ and the number of nodes $n_{nodes,k}$ in zone k , the targeted number of clusters $n_{clusters,k}$ can be obtained with equation (1).

$$\begin{cases} n_{clusters,k} = \xi \cdot n_{nodes,k} \\ \text{if } \xi \cdot n_{nodes,k} < 1, n_{clusters,k} = 1 \end{cases} \quad (1)$$

As an example, figure 1 illustrates an imaginative grid clustered to several sizes. In this figure the original grid is represented by the top left picture, whereas the following pictures show

the 100%, 50%, and 0% zone sizes, in which lines are removed from the grid and nodes aggregated to clusters. This results in grids with 12, 6, and 1 clusters respectively. The red lines in the grids represent the power lines that are clustered in the next reduction step, whereas the blue lines represent the inter-zonal power lines that cannot be clustered.

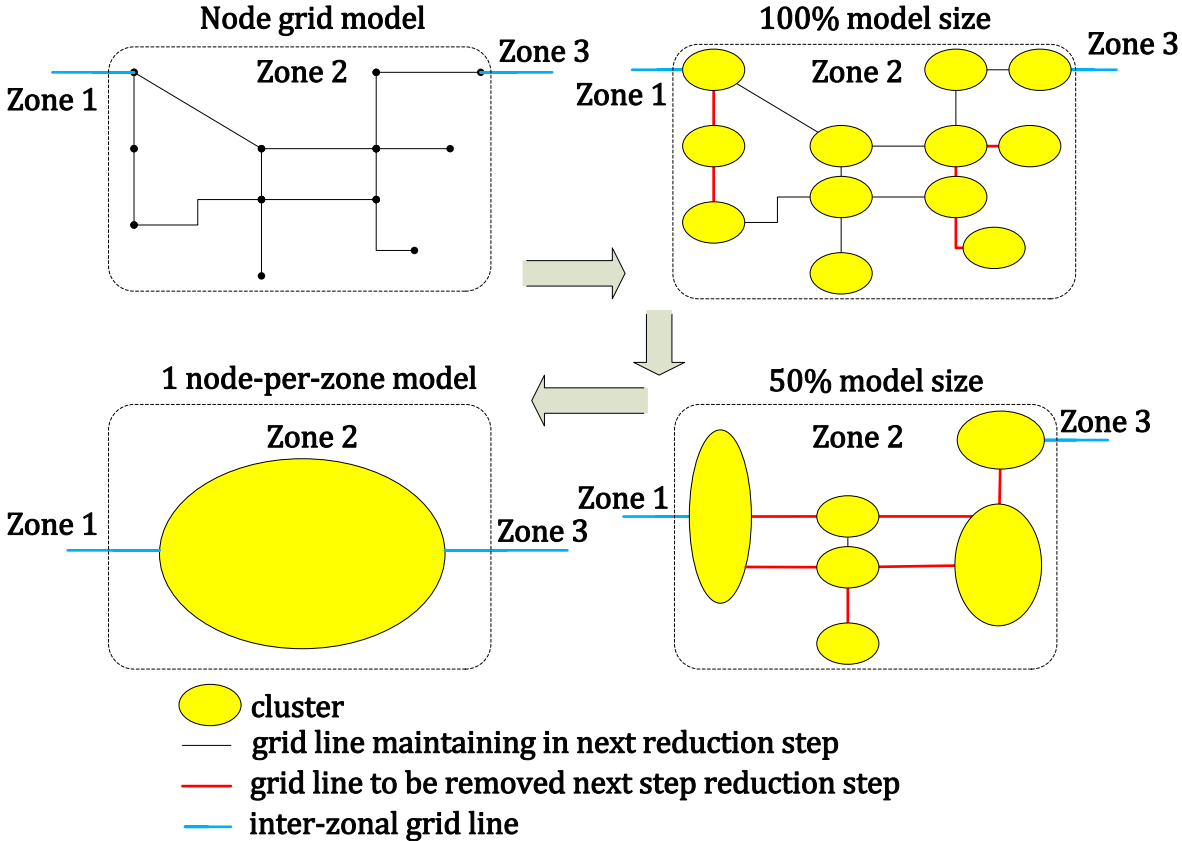


Figure 1. Schematic overview of the cluster process. An imaginative zone is represented in a 100%, 50% and 0% (1 cluster-per-zone) model size.

To identify which specific lines in the zone are clustered, the condition value c_n is introduced. The condition value c_n indicates whether the particular power line is likely to form a bottleneck in the grid. c_n is a binary indicator, consisting of two parts which indicate bottleneck likeliness. The first part of the binary indicator consists of the reactance X_n . The reactance is the decisive factor on the direction of power flows from a node. The reactances of power lines are dependent on the line length and the voltage.

The second part of the binary indicator is a scaling factor, indicating whether power flows are critical or not. Power flows in a line are directly dependent on the phase angle difference between two nodes. This can be represented by the absolute difference in power injections

in node i and node j connected by power line L_n , denoted as $\Delta P_{inj\ i,j}$. A large difference in power injections will lead to a large power flow between two connected nodes. As different time steps will be implemented, there are as many values for $\Delta P_{inj\ i,j}$ as there are time steps. The largest value for $\Delta P_{inj\ i,j}$ between two nodes is selected as the input for the calculation, denoted as $\max(\Delta P_{inj\ i,j})$. Whether a power flow is critical for a line, the capacity of the line TC_n is decisive. Therefore the second part of the binary indicator is the division of $\max(\Delta P_{inj\ i,j})$ and TC_n , resulting in a unit-less scaling factor of the reactance.

For every line the condition value c_n is defined as:

$$c_n = X_n \cdot \frac{\max(\Delta P_{inj\ i,j})}{TC_n} \quad (2)$$

A high condition value indicates that the line is likely to become a bottleneck in the grid.

For every zone, the clustering process starts with a certain threshold value γ . Only lines within the zone are being removed of which the condition value c_n is smaller than γ . As the initial starting value of γ is very low, no lines will be aggregated during the first loop. Next loop γ is increased by 10%, and another clustering loop is performed. This process is continued until finally a value for γ has been reached for which enough nodes have been aggregated, and the targeted number of clusters in the zone has been reached. The schematic overview of this process is depicted in figure 2.

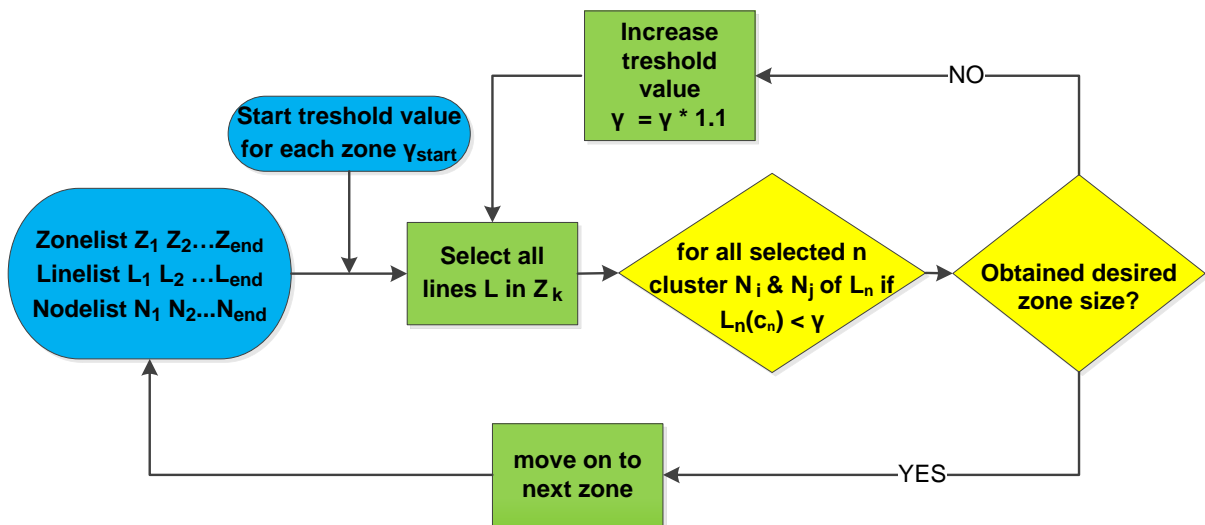


Figure 2. Process overview of the clustering process.

1. Select zone Z_k
2. Initiate starting threshold value γ
3. Select all lines in Z_k
4. Cluster all lines in Z_k with $c_n \geq \gamma$
5. Has ξ been reached? If no, increase γ with 10% and return to step 4. If yes, $k = k + 1$ and return to step 1.

Appendix 1 contains the MATLAB script that has been developed for the cluster process. The result of this process is a set of all nodes, each designated to a particular cluster. As every zone has been reduced to the same value of ξ , the final size of the complete reduced grid is also ξ . For example, if the original system consisted of 1000 nodes, and the targeted granularity would be 50%, then there would be 500 clusters present in the reduced system. The number of clusters is denoted as N_{R+1} .

2.2 Calculation of PTDF matrices

The input information for the calculation of the PTDF consists of a list of nodes, and the lines that connect those nodes. Typically, the grid system consists of $N+1$ nodes, where the $+1$ represents the designated slack node. The slack node is excluded to make the system of equations non-singular. The number of power lines in the system is denoted by L .

Calculating the PTDF matrix starts with the incidence matrix $C(i,j)$, which is a $L \times N$ matrix describing the line connections between the system nodes:

- It solely consists of 0's, 1's and -1's
- If $C(i,j)$ is 0 if node i is not connected to line j
- If $C(i,j)$ is 1 if node i is connected to line j , and the positive flow direction in line j is away from i
- If $C(i,j)$ is -1 if node i is connected to line j , and the positive flow direction in line j is towards i

From the incidence matrix and the reactance vector X the line susceptance matrix B_{branch} ($L \times N$) can be derived:

$$B_{branch} = X^{-1} \cdot C \quad (3)$$

The line susceptance matrix links the inverse of the reactances of the power lines to the respective connected from- and to-node:

- Every row $B_{branch}(i,:)$ represents a power line containing two connected nodes. It contains either one positive and one negative value, or solely zero values
- $B_{branch}(i,j)$ contains the positive inverse value of the reactance if node i is the from-node of line j
- $B_{branch}(i,j)$ contains the negative inverse value of the reactance if node i is the to-node of line j

This is followed by the calculation of the bus susceptance matrix $B_{bus}(N \times N)$. The bus susceptance matrix represents the interconnection topology of busses:

$$B_{bus} = C^T \cdot B_{branch} \quad (4)$$

- $B_{bus}(i,j)$ contains the positive reactance value if node i is the from-node connected to to-node node j
- $B_{bus}(i,j)$ contains the negative reactance value if node i is the to-node connected to from-node node j

In the DC formulation of the power grid, the power injection is linearly dependent on the bus angle. P_{inj} represents the $(N \times 1)$ vector containing the injection in each node, B_{bus} the bus susceptance matrix and ϑ the $(N \times 1)$ vector with bus voltage angles:

$$P_{flow} = B_{branch} \cdot \theta \quad (5)$$

With P_{inj} and ϑ defined as:

$$P_{inj} = \begin{bmatrix} P_{inj,1} \\ P_{inj,2} \\ \vdots \\ P_{inj,N} \end{bmatrix} \quad \text{and} \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_N \end{bmatrix} \quad (6)$$

A similar relation can be retrieved for the power flow P_{flow} , using the line susceptance matrix and the bus voltage angle:

$$P_{flow} = B_{branch} \cdot \theta \quad (7)$$

By putting together equation (5) and (7), the following relation can be obtained:

$$P_{flow} = B_{branch} \cdot B_{bus}^{-1} \cdot P_{inj} \quad (8)$$

A relation has been found between the line flows and the bus injections. When large model sizes are applied, calculating the PTDF matrix can still be computationally demanding. Taking the inverse of B_{bus} becomes exponentially more demanding as the model size grows. Therefore an alternative approach has been applied to obtain the bus susceptance matrix. By using the singular value decomposition method MATLAB® offers, a matrix results that accurately approximates the bus susceptance matrix, but is computationally a lot more efficient. The MATLAB® script for this operation is displayed in equation (9).

$$[U, S, V] = svd(B_{bus}) \quad (9)$$

$$s = diag(S)$$

$$k = sum(s > 1e-9)$$

$$B_{bus_inv} = (U(:, 1:k) * diag(1./s(1:k)) * V(:, 1:k))'$$

The PTDF matrix Φ of the original grid can be defined as:

$$\Phi = B_{branch} \cdot B_{bus}^{-1} \quad (10)$$

And equation (8) can be rewritten to the following equation:

$$P_{flow} = \Phi \cdot P_{inj} \quad (11)$$

Appendix 2 contains the MATLAB® script that develops the PTDF matrix.

2.3 PTDF calculation of the reduced grid

As the clusters have been established, the PTDF for the reduced grid can be derived. This operation will be conducted with the method proposed by [1]. The reduced grid consists of N_R+1 clusters, and L_R equivalent lines.

The first step is to construct a matrix Π_{bus} ($N_R \times N$). This matrix describes the relationship between clusters and busses. The following properties belong to Π_{bus} :

- It consists of solely 0's and 1's
- $\Pi_{bus}(i,j)$ is 0 if bus j does not belong to cluster i
- $\Pi_{bus}(i,j)$ is 1 if bus j does belong to cluster i
- Each row only contains one 1

By using Π_{bus} the cluster injection vector for the reduced grid can be retrieved:

$$(P_{inj})_R = \Pi_{bus} \cdot P_{inj} \quad (12)$$

The reduced PTDF will describe the relation between the cluster injections and the inter-cluster power flows. First of all, the matrix ψ ($L_R \times N_R$) is derived, which describes the contribution of the cluster injections within the cluster to the inter-cluster power flows. So the element $\psi(i,j)$ represents the contribution of the i^{th} cluster's injection to the j^{th} inter-cluster power line.

$$\Psi = \Pi_{flow} \cdot \Phi \cdot diag(P_{inj}) \cdot \Pi_{bus}^T \quad (13)$$

Π_{flow} ($L_R \times L$) is the matrix translating the power flows of the full grid to power flows in the reduced grid.

- $\Pi_{flow}(i,j)$ is 1 when line j in the full grid system belongs to line i in the reduced system, and have the same orientation
- $\Pi_{flow}(i,j)$ is -1 when line j in the full grid system belongs to line i in the reduced system, and have opposing orientation
- All other values are 0

The sum of each row of the matrix $\psi(:,j)$ equals an inter-cluster power flow $P_{flow}^{inter-clusteral}$:

$$P_{flow}^{inter-cluster} = \begin{bmatrix} P_{flow,1}^{inter-cluster} \\ P_{flow,2}^{inter-cluster} \\ \vdots \\ P_{flow,L_R}^{inter-cluster} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{N_R} \Psi^{(1,j)} \\ \sum_{j=1}^{N_R} \Psi^{(2,j)} \\ \vdots \\ \sum_{j=1}^{N_R} \Psi^{(L_R,j)} \end{bmatrix} \quad (14)$$

This expression for the inter-cluster power flow can also be written differently. Borrowing from equation (11):

$$\Psi = \Phi_R \cdot \text{diag}(P_{inj})_R \quad (15)$$

And when combining the equations (12), (13) and (15), ψ can be rewritten as:

$$\Psi = \Phi_R \cdot \text{diag}(\Pi_{bus} \cdot P_{inj}) = \Pi_{flow} \cdot \Phi \cdot \text{diag}(P_{inj}) \cdot \Pi_{bus}^T \quad (16)$$

From which the reduced PTDF matrix can be derived:

$$\Phi_R = \Pi_{flow} \cdot \Phi \cdot \text{diag}(P_{inj}) \cdot \Pi_{bus}^{-1} \cdot \text{diag}(\Pi_{bus} \cdot P_{inj})^{-1} \quad (17)$$

The new formula for calculating the reduced power flows becomes:

$$(P_{flow})_R = \Pi_{flow} \cdot \Phi \cdot \text{diag}(P_{inj}) \cdot \Pi_{bus}^{-1} \cdot \text{diag}(\Pi_{bus} \cdot P_{inj})^{-1} \cdot \Pi_{bus} \cdot P_{inj} \quad (18)$$

With this formula the results in power flows can be calculated for all grid sizes.

The importance of the set of reference injections in equation (17) needs to be stressed. The reduced PTDF is an operating point dependent matrix. The further the deviation from the operating point, the larger the error in power flow will be. Next section will describe the error calculation.

2.4 Power flow error calculation

In this section we introduce an error metric to describe the inter-zonal error that results from the grid simplification. First of all, the absolute error over a power line is defined. After

this, the relative error over a power line is calculated as a share of its respective thermal limit. Then, the average of the relative error is taken over all time steps. From the average relative errors, again an average over all inter-zonal power flows from a particular zone are taken as representative for the error behavior in that respective zone.

The absolute error of a power flow is defined as the absolute differences between the power flow in the reduced model $(P_{flow})_R$ and the full-grid model. Clusters in the reduced grid can be connected by more than one line in the full grid. This is illustrated in Figure 3. The full grid representation on the left side in the picture shows an example system before any clustering has been performed. In order to calculate the equivalent flow in the full-grid model, the flows n in the full grid model belonging to the reduced flow i are summed up in $\sum_{n \in i}(P_{flow})$. In the right hand picture, the nodes have been aggregated to two clusters, which are connected by a singular power line i .

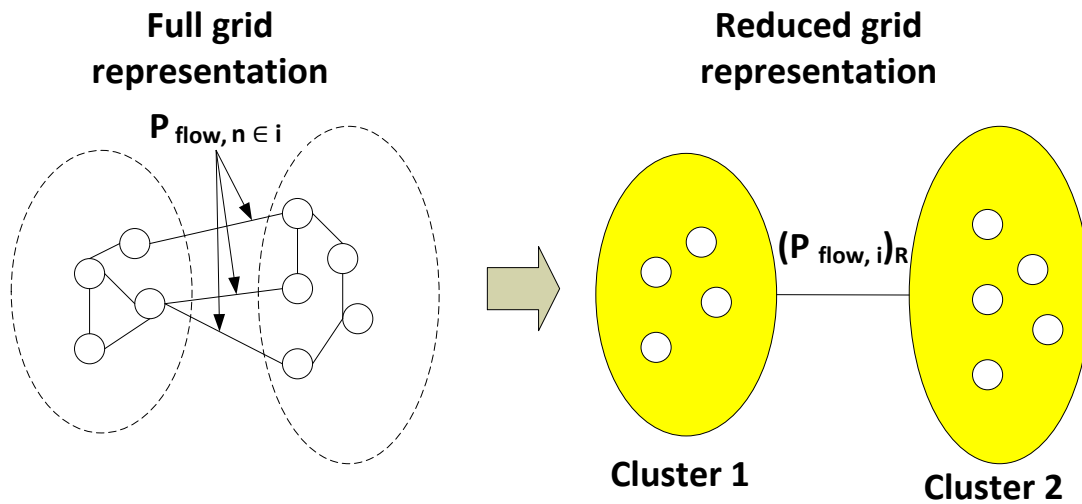


Figure 3. Multiple lines in the full grid representation are combined to a singular line in the reduced grid representation.

Equation (19) describes the calculation of the absolute error in reduced flow i .

$$E_{abs,i} = \left| \sum_{n \in i} (P_{flow}) - (P_{flow,i})_R \right| \quad (19)$$

Often, relative errors are measured as the absolute error divided by the original flow $\sum_{n \in i}(P_{flow})$. However, results show that for very small values of $\sum_{n \in i}(P_{flow})$, the relative error will become very large even when the absolute error might be very small. As this might

lead to a distorted error metric, there has been chosen for a more stable variable to measure the relative error. Therefore, the relative error is calculated as the share of the net available transfer capacity (ATC) of a reduced power line. The net available transfer capacity $ATC_{R,i}$ is the total aggregated capacity of the lines n belonging to reduced line i .

$$ATC_{R,i} = \sum_{n \in i} ATC_n \quad (20)$$

The relative error then becomes:

$$E_{rel,i} = \frac{E_{abs,i}}{ATC_{R,i}} \quad (21)$$

960 different time steps are applied to the power system model. Therefore, 960 error results are obtained per line flow. From these results, an average $\bar{E}_{rel,i}$ is taken for every power line i .

$$\bar{E}_{rel,i} = \frac{\sum E_{rel,i}}{960} \quad (22)$$

It should be mentioned that other metrics might be more suitable for error analysis, depending on the nature of the analysis. For example, for certain types of analysis the maximum error or the geometric average error could be of more interest.

For every zone k an error metric is introduced. For this analysis, only the inter-zonal flows are being analyzed on error size. This has the benefit that the same set of lines is considered for all model sizes. The inter-zonal lines are the only lines that remain un-clustered in all situations, including in the 0% grid size situation. By only including inter-zonal lines a consistent set of lines is being analyzed, even when aggregation levels differ. By doing so, comparing the results for different reduction levels is more valid.

The zonal error \bar{E}_k is based on the errors in the inter-zonal flows of that particular zone. Specifically, it is calculated by dividing the sum of the average error for each inter-zonal flow $\bar{E}_{rel,i}$ by the number of inter-zonal flows in zone k .

$$\bar{E}_k = \frac{\sum \bar{E}_{rel,i}}{number_{linesR,k}} \quad i \in k, interzonal \quad (23)$$

The expression \bar{E}_k is used as a measurement for a zone's performance when reduced to a smaller size.

2.5 Calculating computation time

The third research question inquires the relation between required computation time and the occurring error in the reduced grid. However, the approach of using PTDF matrices for the calculation of resulting power flows does not lead to results for computation time. An optimal flow model would lead to results in which the computation time can be measured for different model sizes.

An optimal power flow optimization model is implemented in the software package AMPL[®], in which the transmission grid and node capacities of the case study of this particular report have been incorporated. This model has been constructed for the purpose of the EU project BETTER [9]. As at the time of this report not all results were available from the simulations, the required computation time is predicted based on the mechanics of the software. To find the optimum solution in AMPL[®], the Karmarkar's algorithm for the interior-point method is applied. This is a method commonly applied in solving large optimization problems, chosen for its efficiency. The required computation time CT for this method is proportional to the number of variables n_{var} to the power of 3.5 [10].

$$CT \propto (n_{var})^{3.5} \quad (24)$$

Although the power lines in the model bring forward additional variables for the optimization model, as input for the number of variables n_{var} only the number of nodes/clusters in a grid is taken. However, the results from the clustering process for the case study show that the number of lines also proportionally reduces with the number of nodes. Taking only the number of clusters in a reduced model to indicate the number of variables in a model is therefore a valid assumption.

The error metric for this particular part of the analysis takes the average error over all inter-zonal power lines i in the complete transmission grid system, and over all 960 time steps t , for a particular grid size.

$$\bar{E} = \frac{\sum_t \sum_i E_i}{960 * n_i} \quad i, \text{ inter-zonal} \quad (25)$$

In section 4.4 the required computation time is plotted versus the error metric.

3 Application of the simplification algorithm to the ENTSO-E transmission grid model

The case study at hand consists of the high voltage grid of 39 countries, constructed based on the power lines of the ENTSO-E grid map of 2012. Section 3.1 will list how the different countries of the model are divided over zones. Section 3.2 will continue to describe how the power grid is modeled, after which transformers are integrated into the transmission grid in section 3.3. Finally, section 3.4 will explain how the injections in the nodes are calculated.

3.1 Grouping countries into zones

As some countries contain very few nodes, they are combined in order to keep the total number of zones limited. By doing so, it will be easier to reduce the grid model to fewer clusters, as the minimum number of clusters is dependent on the number of zones for the 0% grid size situation. If more than one country has been designated in one zone, it is performed in such way that the respective countries have an extensive number of power lines connecting them.

The 39 countries from the grid map have been included in the analysis, and have been categorized in the following 26 zones. Figure 4 shows the division on a map.

1. Turkey	7. Austria Slovenia	15. Denmark	24. Algeria
2. Greece		16. Germany	25. Tunisia Libya
3. Bulgaria	8. Hungary Slovakia	17. Belgium Netherlands Luxembourg	26. UK Ireland
4. Croatia Bosnia	9. Czech Republic		
5. Albania Montenegro Kosovo Serbia Macedonia	10. Poland	18. France	
	11. Estonia Latvia Lithuania	19. Spain	
			20. Portugal
		21. Switzerland	
	12. Finland	22. Italy	
	13. Sweden		
6. Romania	14. Norway	23. Morocco	

Table 1. The countries of the grid model designated to zones.

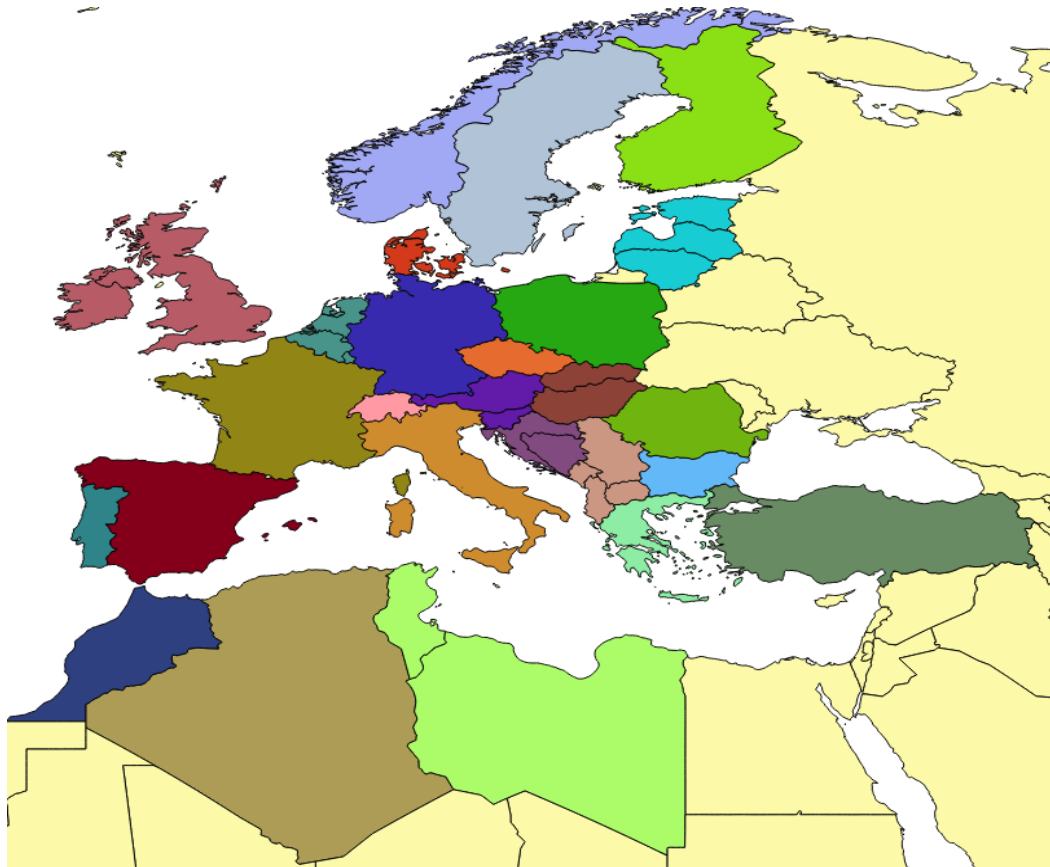


Figure 4. Map of zone division over countries.

3.2 Modeling the power grid

From the countries listed in the previous section, the transmission grid network is modeled. Based on the ENTSO-E grid map, the following voltage levels have been included in the grid model:

- 220 kV
- 300-330 kV
- 380-400 kV
- 750 kV

In addition to this, all DC lines represented within the ENTSO-E grid map have been included. The DC lines have been modeled to behave as AC 380 kV lines. In reality, DC lines behave like gateways that can be independently operated, rather than phase angle dependent variables. But again, there should be stressed that the aim of this report is to measure the accuracy of power system reduction, not to obtain accurate results on power flows. The 110 kV

transmission lines are not fully illustrated in the ENTSO-E grid map. The only 110 kV lines that have been integrated are the intra-country power lines in Denmark and Sweden. In particular in Denmark, the 110 kV lines form the densely constructed backbone of its transmission network. Therefore they should not be excluded from the grid model. Additionally, any inter-zonal line that is at either side of the border not connected to its zone's power grid is excluded. When a node is not connected to the rest of the zone's power grid, it would form an island during the clustering process. This would prevent the model from eventually reaching the one-node-per-zone situation.

The grid map is transformed in numerical data using the software QGIS 2.4.0-Chugiak. In this software package the ENTSO-E grid map is projected onto a vector map of the world including GIS coordinates. By doing so, several important characteristics of the power lines can be retrieved from the grid map that are key to grid analysis. First of all, the coordinates of the beginning- and end point of a power line can be obtained, indicating the location of a system node. Secondly, the software package enables the calculation of power line lengths. From the power line length the respective reactance can be calculated, as this is a length dependent property. If l_i represents the length of a line i , p the voltage level of the respective line, and x_p the inductive reactance per meter, the reactance X_i for a power line becomes:

$$X_i = x_p \cdot l_i \quad (26)$$

And finally, other important characteristics can be coupled to the transmission lines, like the number of circuits or the capacity of the line.

3.3 Introducing transformers in the model

As the model involves several voltage levels, the transmission grid as presented in the model would in reality involve a transformer in every node where more than one voltage level is included. These transformers – just like transmission lines – carry a certain reactance and capacity, hence influencing power flows. A deeper analysis into the values for these characteristics shows that compared to transmission lines, the average transformer carries a reactance that is larger than the average line in the grid model. Consequently, it would be a major simplification not to involve transformers in the grid model.

The transformers have been modeled to act like power lines in the grid model. For further analysis those transformers could act as phase shifters or to maintain certain offset flows. Any node to which a number of voltage levels are connected is split in just as many nodes as there are voltage levels. Although these nodes share the same coordinates, in the grid model they are connected to each other by transformers. This process is illustrated in Figure 5.

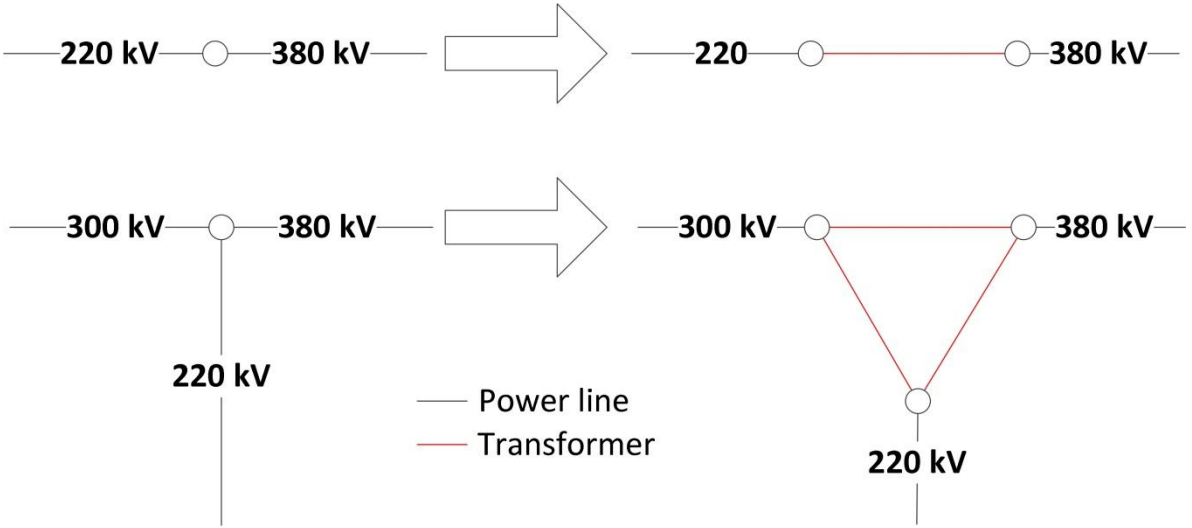


Figure 5. Schematic overview of the process reforming a node with several voltages connect to it, into the equal number on nodes.

These imaginative power lines represent transformers, carrying the properties of this particular transformer in this country. The data used for the transformer properties is from an internal database of the Energy Economics Group [11]. As not all data for transformation between a certain voltage level and every other voltage level is available, certain assumptions have been made. First of all, the properties of a transformer are determined by the highest voltage level in a node. So a transformer between a 220 kV and a 380 kV level receive the properties assigned of a 380 transformer. Secondly, no transformer data is available for voltages higher than 380 kV. There will be assumed that transformers with a maximum voltage higher than 380 kV carry the properties of a transformer with a maximum voltage of 380 kV.

3.4 Calculation of node injections

In order to make an estimation of the injections in the system busses, both demand and generation per country have been derived via the construction of generic time series. Generation patterns as well as the demand time series in nodes and countries are estimated based on the respective production capacity at each node.

First of all, we assign power plants from a locational power plant database to the closest node within the grid. The distance between a system node and the power plant is computed by inserting both of their global coordinates in a Vincenty's algorithm for a WGS-84 ellipsoidal Earth. This results into a list with the distance of a country's nodes to the respective power plant. The assumption is made that a power plant can only be assigned to a node that is located in the same country, as the local power grid is likely to be constructed in such way that the plant is to its national grid, rather than another country's grid. The node which fulfils this requirement, and which has the shortest distance to the plant is selected. The power plant will be assigned to this particular node.

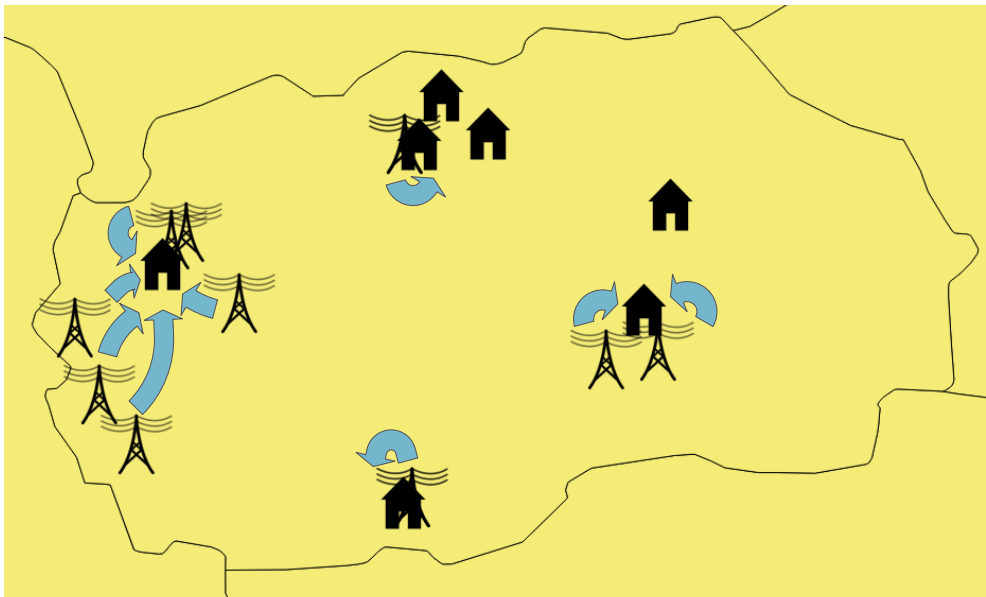


Figure 6. Example of the assignment process of plants to nodes for Macedonia. The nodes are illustrated by the houses, whereas the electricity lines represent power plants.

The time dimension of the model will be formed by the aforementioned 960 time steps, for each of which a snap shot injection pattern is calculated for each node in the grid. The 960 time steps represent 960 hours, fitting into 40 days of 24 hours each. The 40 days are evenly distributed over the year, with 9 days intervals. From the available plant capacity the

generation and demand profiles are determined for each time step according to the same algorithm for all countries.

The generation profiles are created using a merit order method on a country level. The following generation power sources are included:

- PV power
- Hydro power (only run-of-river, storages are neglected)
- Wind power
- Gas power
- Oil power
- Coal power
- Lignite power
- Nuclear power
- Geothermal power
- Biomass power
- Waste power

Within these sources no differentiation is made between distinct technologies. For example, a gas combined cycle plant and a gas internal combustion cycle plant are considered to be equal. Every technology has the same marginal costs based on historic fuel prices and default efficiencies and a carbon price of 11.9 EUR/tCO₂ [11]. For simplicity we do not consider pump storages in this analysis.

Gas, oil, coal, lignite, nuclear, biomass, and waste plants are being considered as dispatchable generation capacity. They are ranked after PV, wind and hydro generation in the merit order. The generation of these three non-dispatchable sources is variable and depending on natural circumstances. For each non-dispatchable source per country a performance curve is produced. The performance curve indicates the production of the technology per country, so generation of each non-dispatchable source is expected to perform uniformly over the whole country. So at every time step, the generation per MW installed capacity is thus equal for all plants of a non-dispatchable source in a country.

PV performance curve

PV generation is subject to the following pattern:

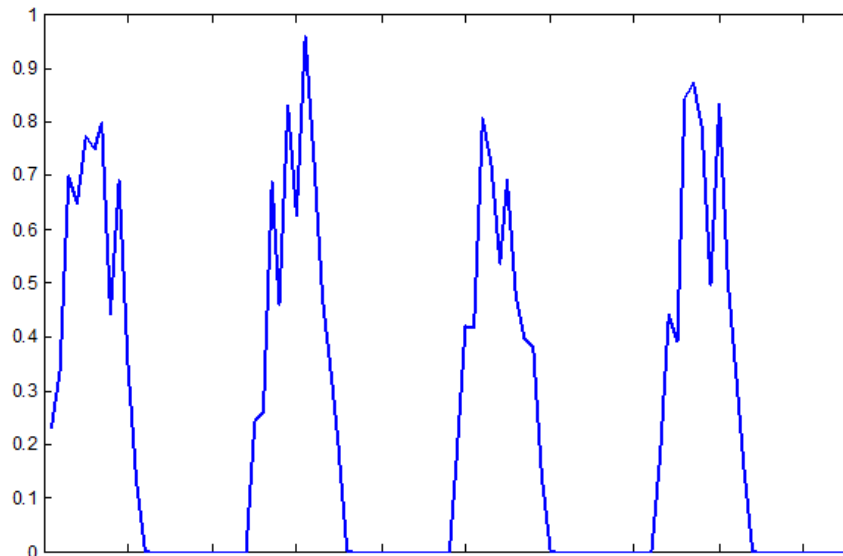


Figure 7. Diagram exemplifying node PV generation, as a share of the installed capacity per time step.

Figure 7 represents an example of the generation pattern, for four consecutive days. The Y-axis represents the solar generation as a share of total capacity. A sinus pattern describes the generation profile by day, and by night the generation reduces to zero. An extra noise layer scaling down the sinus shape resembles sun blockages due to clouds.

Wind performance curve

Wind power is assumed to be producing electricity according to a Gaussian distribution. Each wind mill is producing at an average of 50% of its generation capacity, with a standard variation of 20% of the capacity over the time steps.

Hydro performance curve

For hydro generation, actual data from river flows from all over Europe is subtracted from SMHI [9] for the year 2009 (most recent available data). This website provides flow information for a great number of rivers over Europe. Per country one river is selected, this river is assumed to be representative for the flow behavior of the entire country. The generation of hydro run-of-river is assumed to be directly related to the flow speed of the water. This share indicates the generation of a hydro power plant as a share of installed capacity.

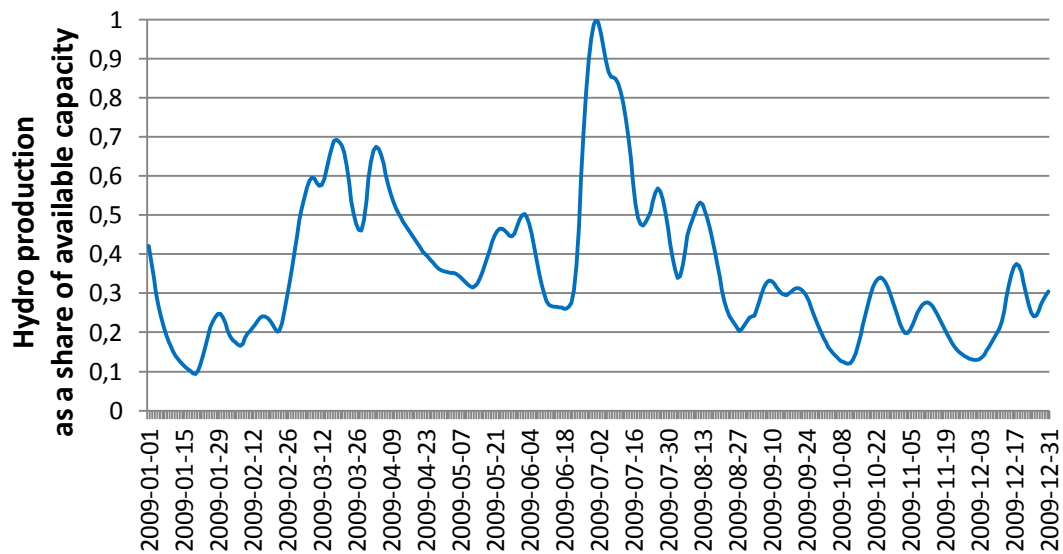


Figure 8. Example (Austria) of the variability of the production of hydro run-of-river in a country, as a share of installed capacity per time step.

As only 40 days a year are modeled, data is spread out over the 40 days in order to integrate the water flow behavior of the whole year in the hydro generation injections. For every country the water flow data is normalized, taking the highest water flow per year as 1, and any lower number as a share of this, ranging from 0 to 1.

Node demand

As direct node demand data is not available, the node demand is calculated based on the dispatchable capacity in a node. Unlike renewable sources like wind or hydro – dispatchable capacity is likely to be installed in those regions where demand is high, making this a reasonable assumption. This node demand value is extended with 50% of the hydro run of river capacity in a country. Hydro capacity is added to correct for the fact that in some regions (e.g. Scandinavia, Switzerland) the majority of installed capacity is hydro. Taking only dispatchable capacity in those countries as a reference would lead to very unrealistic demand values. The maximum node demand is set at 90% of the combined dispatchable+50%-hydro-capacity present in a node. The remaining 10% of all plant capacity is randomly distributed over the maximum node demand of all nodes in the country.

The node demand is randomized over the time steps, ranging from 50% to 100% of its maximum value. This percentage is set on a country scale, for example, in time step 453 the demand percentage in all nodes located in the France is 84% of its maximum.

Dispatchable generation

By subtracting the non-dispatchable power production from the demand, the demand part remains that will be covered by dispatchable capacity. Which dispatchable source is active at what moment, is established based on a merit order model. The order is based on variable costs of power production for Germany [11], and is uniformly set for all countries as:

1. Waste power
2. Geothermal power
3. Nuclear power
4. Lignite
5. Coal
6. Biomass
7. Gas
8. Biogas
9. Oil

This merit order model is supplemented with data of the available net transfer capacities between countries, as provided by the ENTSO-E [13]. The values used are displayed in Appendix 3. The values for capacities are empirically measured for peak hours during working days, in the summer of 2010. The net transfer capacities enable power trade between countries. In this situation not every country needs to sustain its own power supply. They can also depend on their neighboring countries in case the power sources are cheaper there, or in case there is a capacity shortage to provide for the inland demand.

This merit order model finds the power production for every country. For every time step, the plants coming first in the merit order are producing power to fulfill demand, supplemented by possible import from- or export to other countries.

Final injections

The final injections are obtained by subtracting the resulting generation from each present plant by the local demand in each node. This results in a matrix of the size $N \times 960$, describing the injections in all system nodes for each time step. Although the algorithm will not precisely imitate reality, it will give a proper insight in the resulting power flows in the power grid. Besides, the purpose of this paper is not to precisely measure power flows, but to measure the accuracy of the reduction method. Slightly realistic injection scenarios are desired though, as the injections directly influence the power flow. This will provide a clearer insight into the source of differing error patterns between zones. The reference set of injections on which the reduced PTDF matrix is designed, is an average over time for all injections.

This is the input for the analysis, for which the results are presented in chapter 4.

4 Results

In this chapter the results will be described. Section 4.1 will test the clustering algorithm on a small example grid. Section 4.2 then continues to show the results of the cluster algorithm on the case of the ENTSO-E grid map. The errors will be illustrated in diagrams and discussed in section 4.3. Finally, section 4.4 shows the relation between the required computation time and accuracy.

4.1 Validation cluster algorithm

To check whether the clustering algorithm works as described, it is first tested on a smaller test grid before applying it to the large case. For this purpose, a 12-node model is designed connected by 12 lines. This grid is illustrated in figure 9, with the nodes marked from A to L, and the lines numbered from 1 to 12. A simulation on this test-grid will show whether the clustering process aggregates the nodes together as proposed in section 2.1.

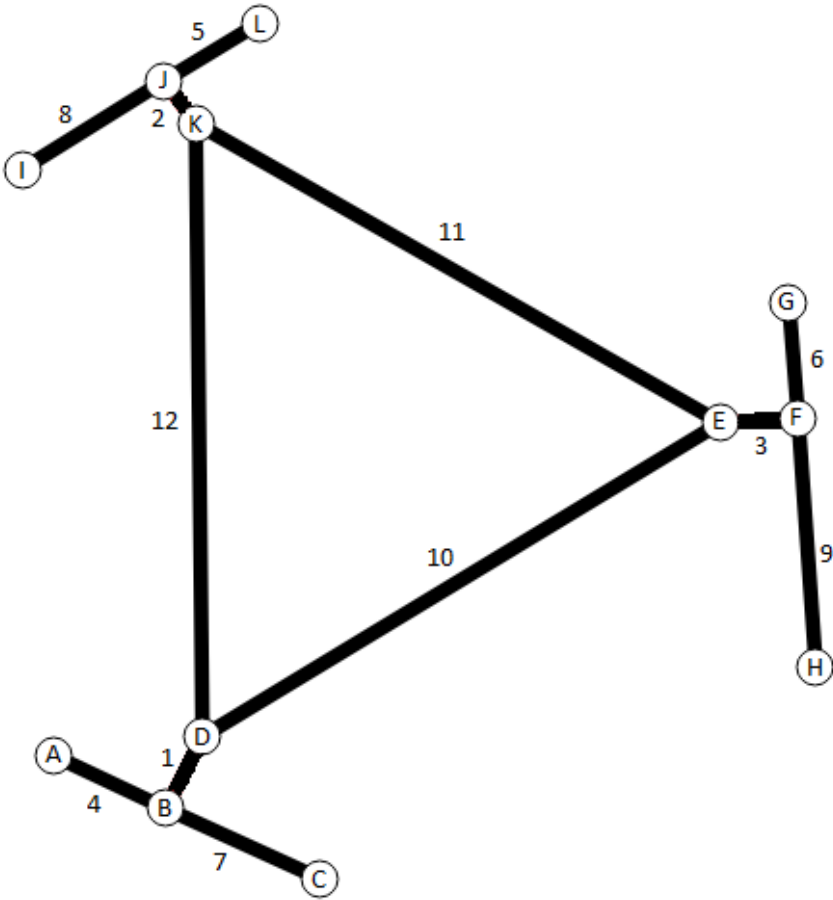


Figure 9. The 12-node test grid for which the clustering algorithm is tested.

Like in the large grid of the case study, all lines carry properties like reactance and capacity, for which the values for this example case are based on a 380 kV power line. Each line is carrying a certain condition value c_n , which is decisive in the order of which the lines shall be clustered.

The node properties are listed in table 2. The X- and Y coordinates represent the location of the node. Each node has been assigned a random injection for this example.

Node	X-co (°)	Y-co (°)	Injection (MW)
A	-117.69	54.80	-300
B	-116.61	54.29	-400
C	-115.11	53.61	-200
D	-116.25	54.99	-500
E	-111.24	58.02	0
F	-110.49	58.06	100
G	-110.58	59.19	200
H	-110.32	55.66	-100
I	-118.00	60.47	200
J	-116.62	61.33	400
K	-116.32	60.92	400
L	-115.70	61.89	300

Table 2. The node properties of the example grid

Only one time step is used for simplicity. This means that the value for $\max(\Delta P_{inj\ i,j})$ in equation (2) from section 2.1 becomes $P_{inj\ i,j}$. The formula for c_n becomes:

$$c_n = X_n \cdot \frac{\Delta P_{inj\ i,j}}{TC_n} \quad (27)$$

The values for c_n are normalized to the maximum value of all power lines. The results are listed in Table 3.

Line ID	From node	To node	Length (km)	X (p.u)	TC (MW)	$\Delta P_{inj,i,j}$ (MW)	Normalized c_n
1	B	D	86.8832	0.2578	1700	100	0.0146
2	J	K	56.9351	0.1690	1700	0	0.0000
3	E	F	83.6994	0.2484	1700	100	0.0141
4	A	B	133.1529	0.3952	1700	100	0.0224
5	J	L	119.9329	0.3559	1700	100	0.0202
6	F	G	125.4485	0.3723	1700	100	0.0211
7	B	C	182.3115	0.5411	1700	200	0.0614
8	I	J	180.3616	0.5353	1700	200	0.0607
9	F	H	267.1199	0.7927	1700	200	0.0900
10	D	E	651.3024	1.9329	1700	500	0.5484
11	E	K	649.7382	1.9283	1700	400	0.4377
12	D	K	659.7762	1.9581	1700	900	1.0000

Table 3. The line properties of the example grid

Now that the value for c_n is calculated, the order in which the lines should be removed is also known. First the lines with the lower c_n values are removed, until finally the last line is removed in the 0% grid size. Table 4 lists all the lines which should be removed compared to the previous grid size, based on the c_n values from table 3.

Grid size	Lines to be removed
100%	
75%	1, 2, 3
50%	1,2,3,4,5,6
25%	1,2,3,4,5,6,7,8,9
0%	1,2,3,4,5,6,7,8,9,10,11,12

Table 4. The lines removed for each grid size

The data resulting from the clustering process is visualized into equivalent grids. If nodes are aggregated together by removing the existing power line between them, the resulting coordinates of the cluster are based on the average coordinates of the nodes it contains. Figure 10 illustrates the result of the clustering processes.

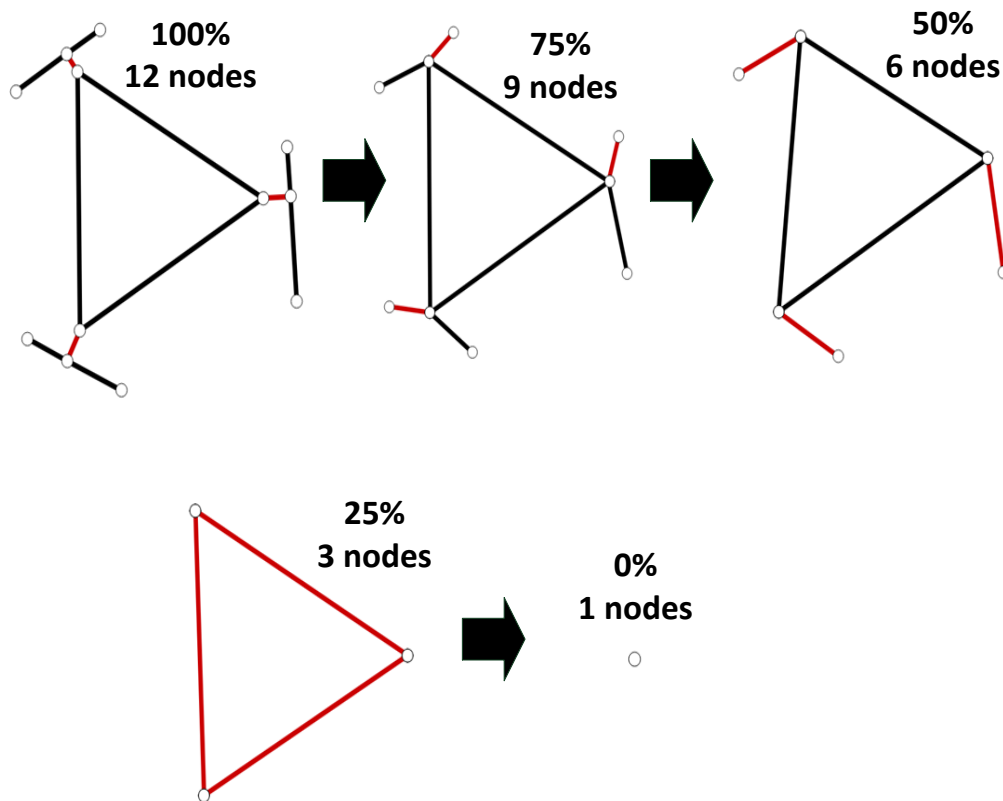


Figure 10. Reduction process of a 12 node example grid. The red lines illustrate the lines that are supposed to be clustered in the next reduction step.

Five pictures illustrate the topology for different targeted grid sizes. In the first picture, the full grid (100%) is illustrated, and the following pictures are a reduced version of it (75%, 50%, 25%, and 0%). The red lines in this illustration represent the power lines for which the nodes it connects are not clustered in the current reduction step, but which are expected to be clustered in the next reduction step based on their condition values. If the algorithm works as predicted, the red lines are removed in the next step, aggregating their two connected clusters in the process.

As can be deduced from this example, the cluster progress performs as expected for every reduction step. In each step, the lines marked red in the previous step have been removed, clustering the respective beginning and end points of those lines. As the clustering process has been validated, it is now applied on the case study of the ENTSOE-E grid.

4.2 Reduction process of the case study

The model has initially been computed at the sizes 100%, 75%, 50%, 25%, and finally 0%. The clustering algorithm is able to reduce every zone to one cluster. This means the grid has been properly modeled, and the clustering process is not leading to island-formations within zones. The number of clusters and equivalent lines in the reduced grid are listed in Table 5.

Modelsize	0%	25%	50%	75%	100%
Number nodes	26	1210	2461	3915	5263
Number lines	46	1637	3367	4930	6496

Table 5. The grid characteristics per model size

The full grid is illustrated in Figure 11. The different colors of the power lines illustrate the different voltage levels in the grid. Figure 12 and Figure 13 illustrate the result after reduction to 50% and 0%. As parallel lines of different voltage levels are combined, no distinction is made in this picture to indicate different voltage levels.

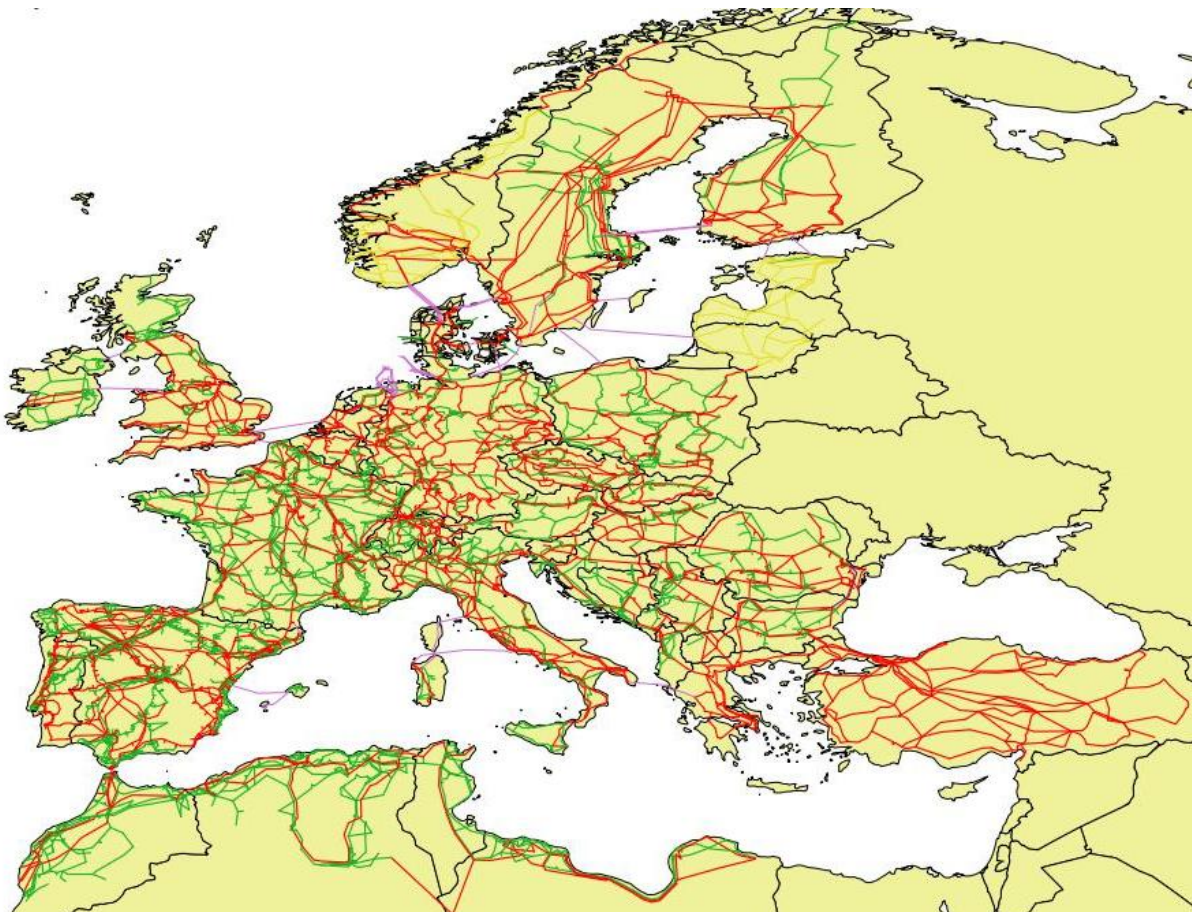


Figure 11. The full grid at 100%, before any reduction has occurred.

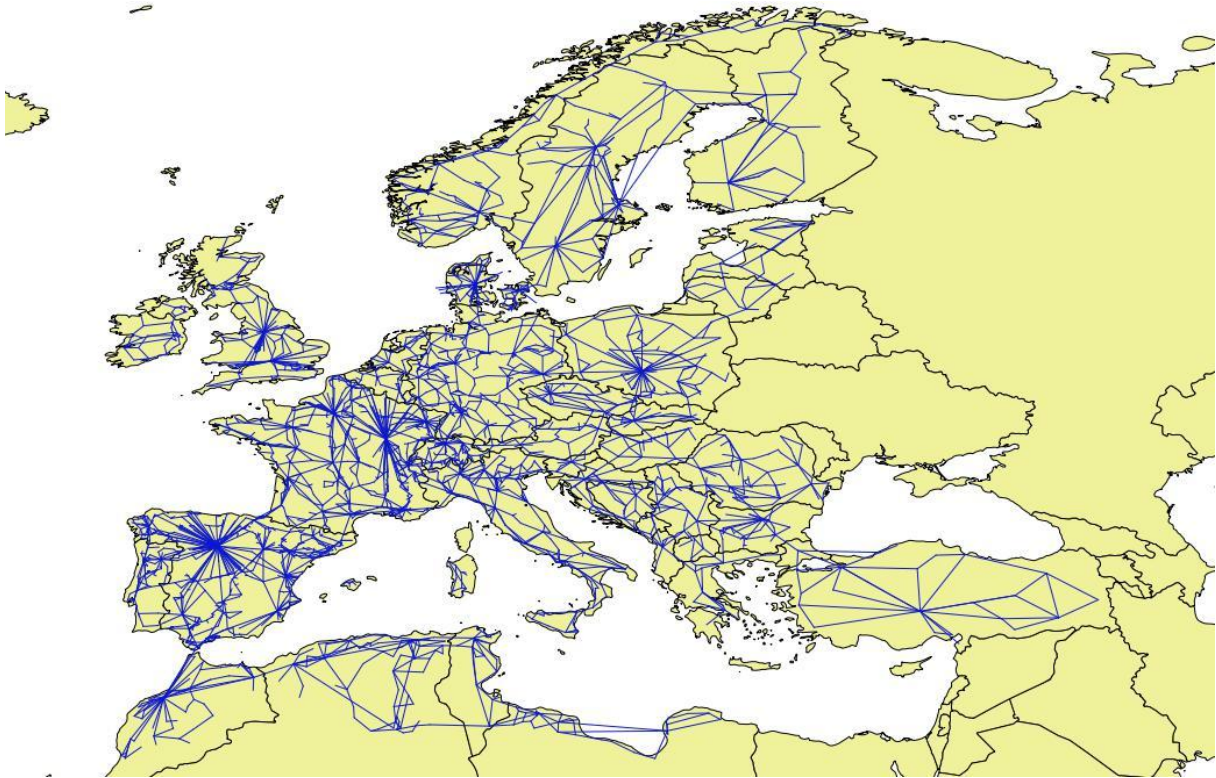


Figure 12. The grid reduced to 50% of its original size.

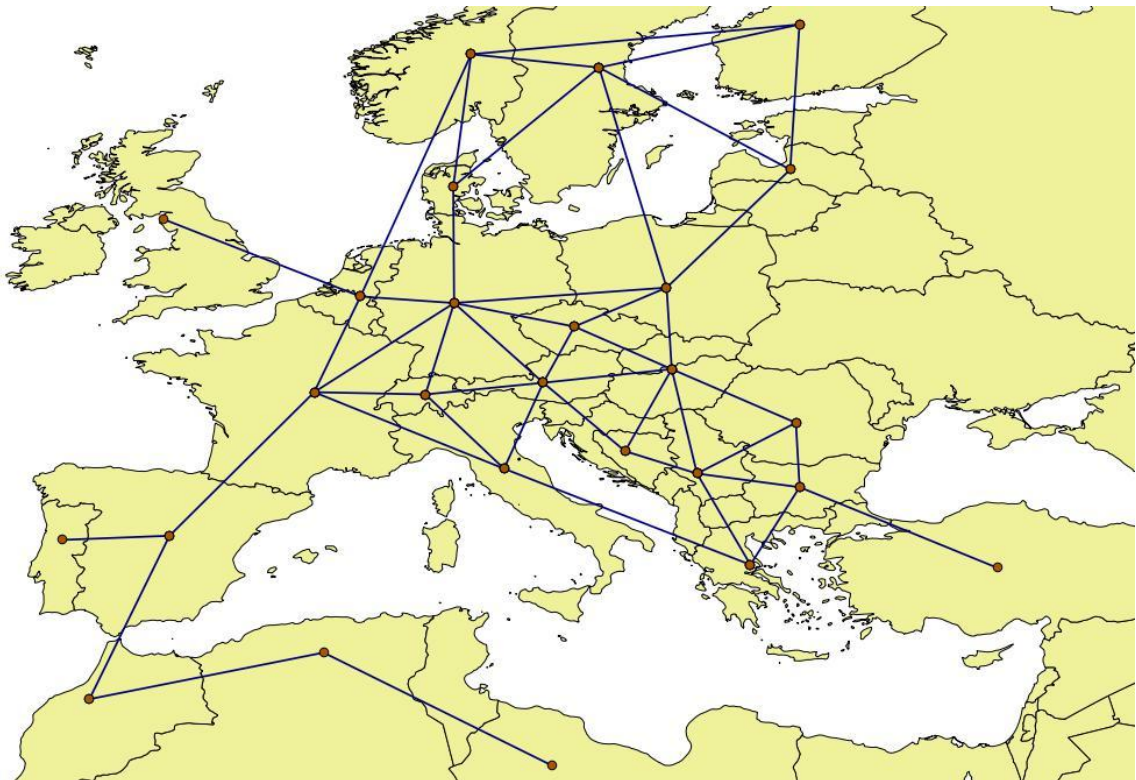


Figure 13. The grid reduced to 0%, thus becoming a one cluster per zone situation.

The smallest grid of 0% only contains one node per zone. In this case, all the plant capacity is directed to one point in the zone, as is the demand. The power lines represent the available

transfer capacity between countries, whereas the power flows over a line represent the power trade between the countries.

4.3 Error analysis

With the method described in section 2.4, the error metric is calculated for each zone. The error results of this process are illustrated in Figure 14. Results are obtained for the 100%, 75%, 50%, 25%, and 0% grid size situation. Each point in the graph describes the error for a particular zone for a particular grid size.

The red line in the graph represents a cutoff value for which the occurring error in a zone is allowable or not. An error limit in inter-zonal power flow is established of 20% of the ATC-value. This value has been considered as allowable in the context of predicting power flows in the project BETTER [9], and is also used as a benchmark in this thesis. With an allowable error benchmark of 20%, the model size can be found for each zone for which the error for this zone remains under the set benchmark.

As can be read from the figure, most zones show rather constant and low error behavior up until the point where each zone is reduced to 25%. This means that most zones can be reduced to 25% without the occurrence of very large errors.

To gain insight in what happens precisely in the interval after 25% more simulations are performed. With ascending steps of 2.5% the model size is reduced from 25% to 0%. The results of this process are illustrated in Figure 15. This diagram shows surprising results. Whereas some zones gradually increase the average error as the grid size reduced, others seem to have a more unpredictable character.

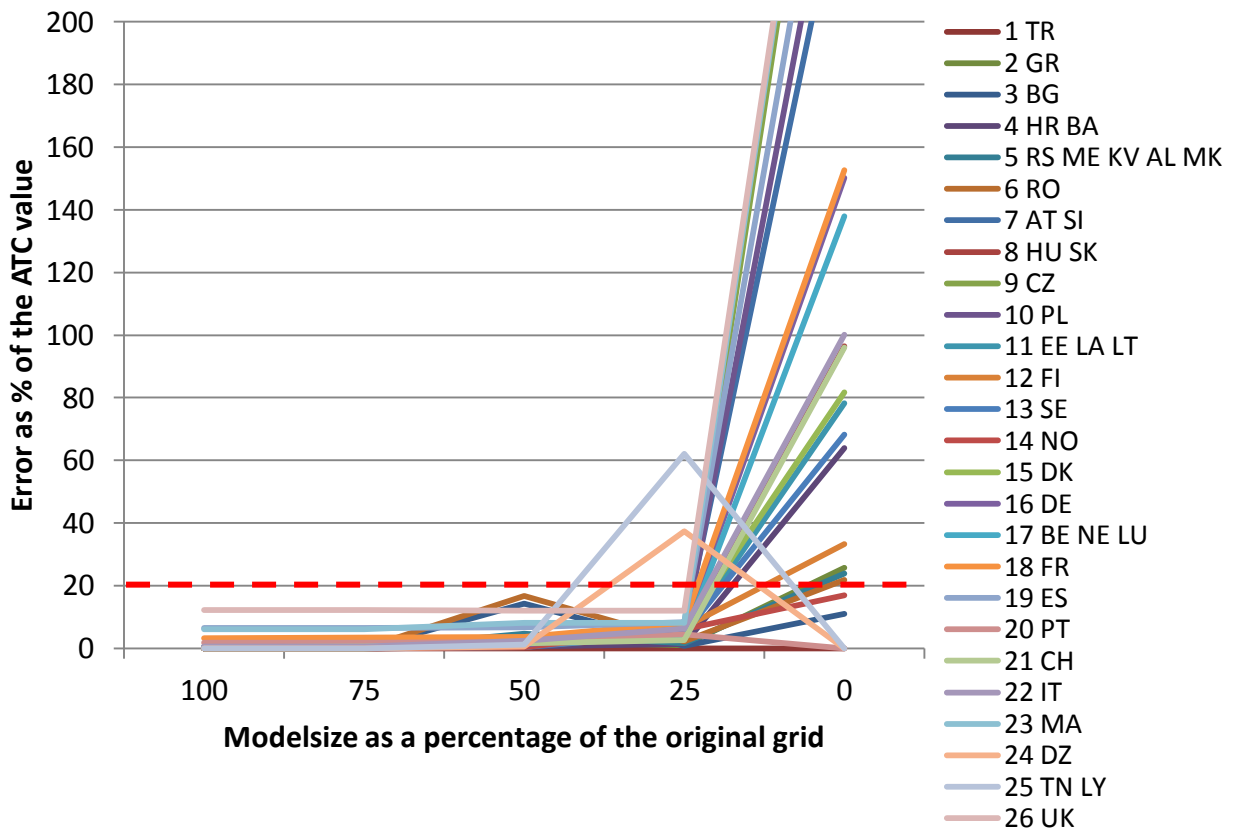


Figure 14. The error diagram per zone, for reduction from 100% to 0%. The red dotted line illustrates the allowable error for each zone.

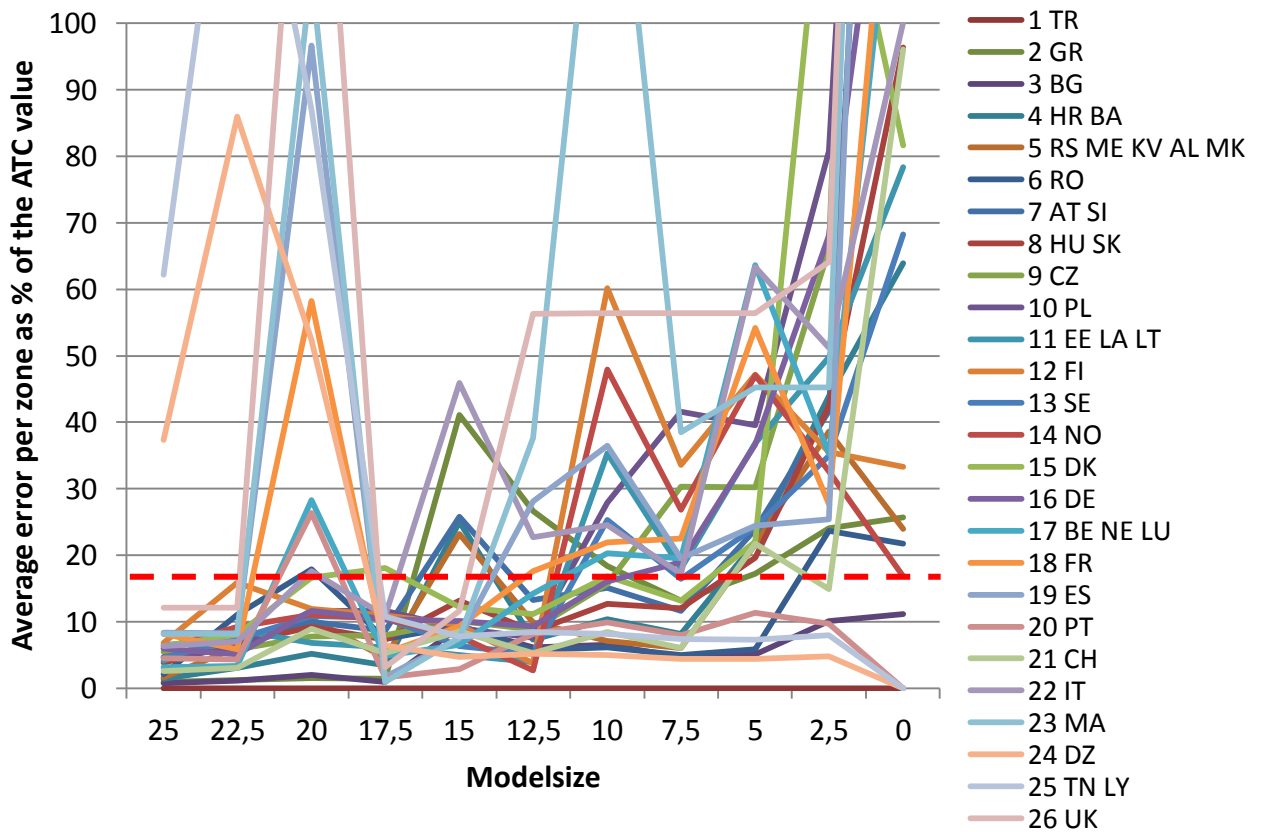


Figure 15. The error diagram per zone, for reduction from 25% to 0%. The red dotted line illustrates the allowable error for each zone.

The decrease of errors in some zones, as seen in figure 15, can be explained by Figure 16. The first situation in this figure illustrates two inter-zonal flows opposing each other. In the second step, the clusters within each zone are aggregated, and the previous two inter-zonal flows are joined into one flow. According to equation (19) the absolute error is directly dependent on the absolute power flow over a power line. As two power lines with opposing direction cancel out each other's flow when they are combined, they also cancel out each other's error. On top of this, according to equation (21), the relative error is negatively dependent on the net available capacity. As the ATC-value increases when the two lines are combined, this also contributes to a lower relative error.

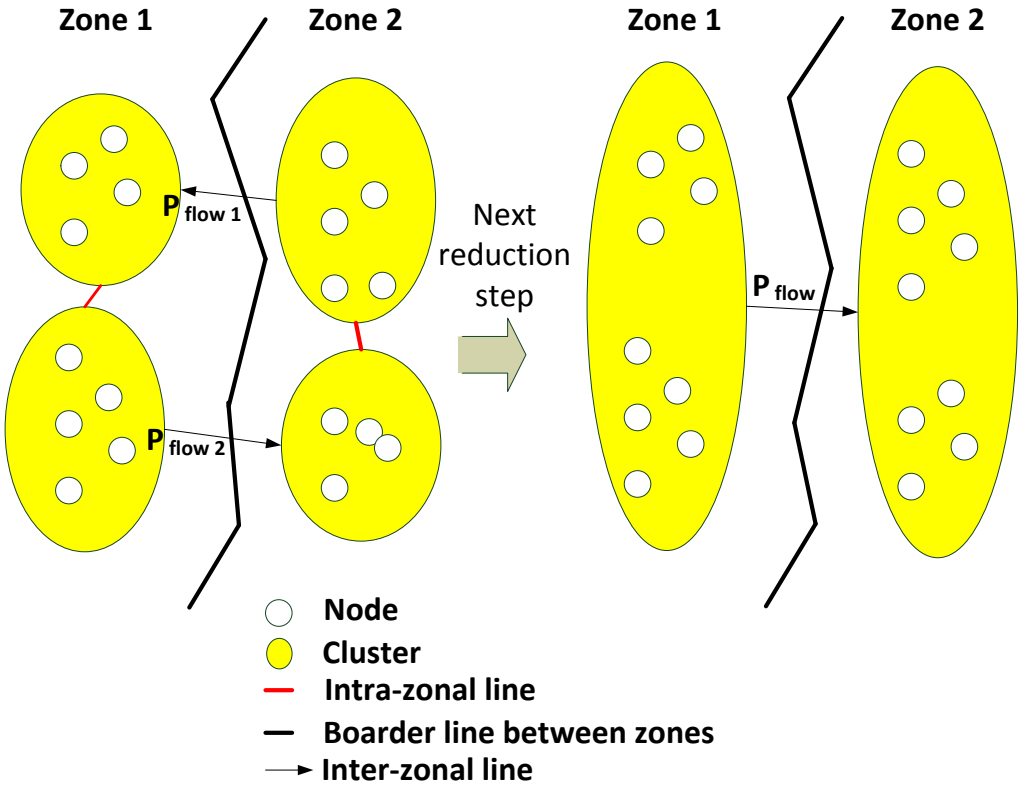


Figure 16. Two opposing cross boarder lines are clustered to one line.

The results from Figure 14 and Figure 15 can be used to obtain the allowable reduction for each zone. For each zone is identified at what grid size it crosses the allowable error of 20% ATC-value for the first time. The grid size right before this step is taken as the allowable reduction for the zone. The results for each zone are illustrated in Figure 17.

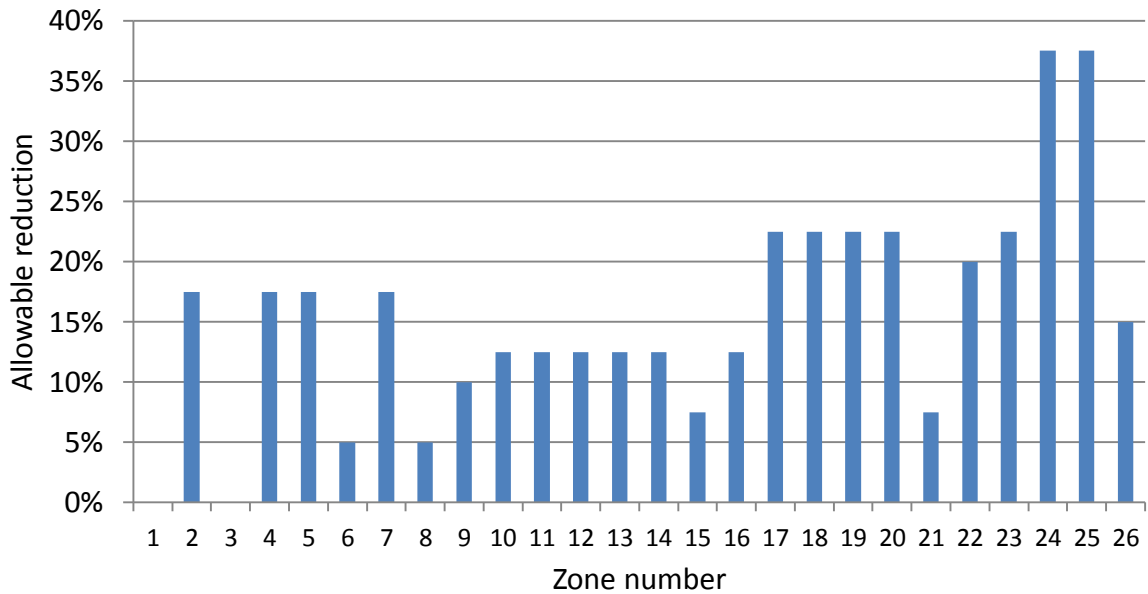


Figure 17. Allowable reduction per zone.

With the allowable reduction per zone identified, now a deeper analysis is performed as to whether grid characteristics can be identified for zones that improve the allowable grid reduction. The reactance has been chosen as the most suitable indicator for the nature of a grid. It contains both the length of a power line, as well as its voltage characteristic in one indicator. The power flows from a node are directly dependent on the reactances of the connected power lines.

To identify whether a relation exists between reactance and the allowable reduction, first of all the reactances of all power lines of each zone are ordered according to size. This is exemplified for the case of UK in Figure 18. For each zone the allowable reduction is now known. From this, there can be derived how many lines have been removed in the clustering process to get to the point of this allowable reduction of this zone k . The number of lines removed is denoted as RL_k . The cutoff reactance for each zone, is taken as the $(RL_k)^{th}$ reactance in the zone.

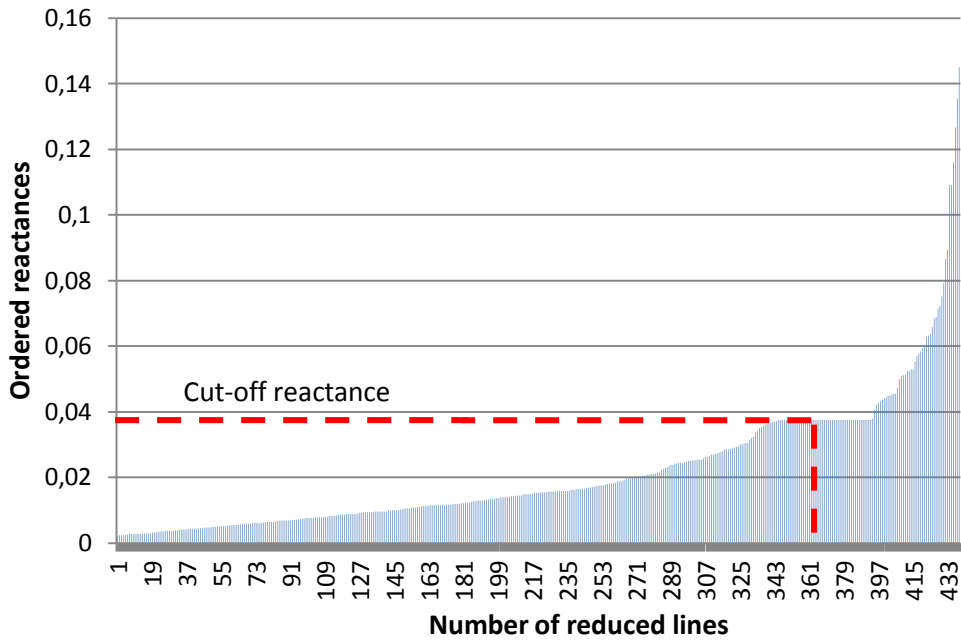


Figure 18. Reactances of power lines of the UK ordered according to their values. The number of removed lines in the clustering process, selects the cutoff reactance.

The cutoff reactance is plotted against the allowable reduction for each zone. If a relation exists between the reactances in each zone and the allowable reduction, a linear pattern between the two indicators could be identified. As can be seen in Figure 19 there is not a clear relation that can be observed. The allowable reduction of a zone's grid is thus not directly dependent on the reactances.

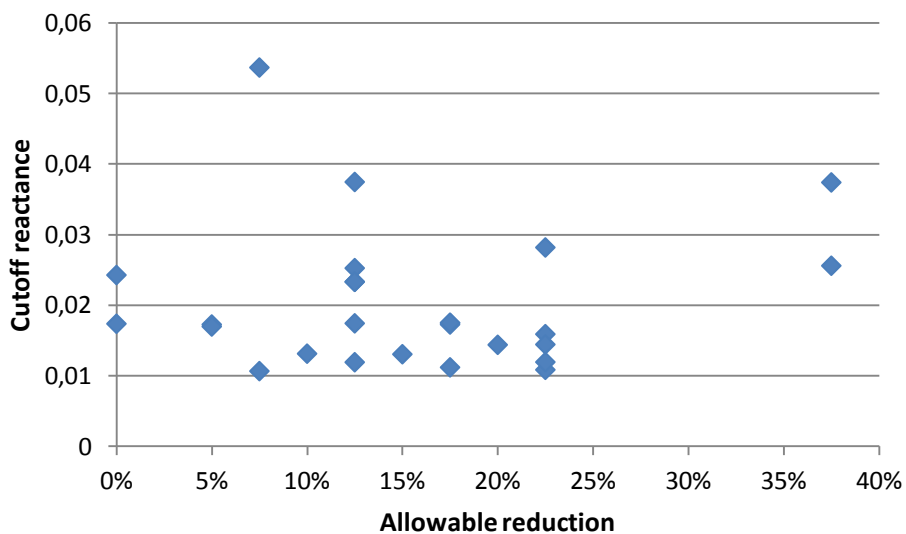


Figure 19. The cutoff reactance vs. the allowable reduction in each zone.

The absence of a noticeable relation between these two factors can be explained by the strong influence of the power injections. Obviously, the power injections directly influence power flow. When using a reduced PTDF, the relation between power flow and node injection is assumed to be a linear simplification. The accuracy of the power flows in the reduced grid is strongly dependent on the deviation from the reference set of injections for which the reduced PTDF is established. The error of the calculated power flow becomes larger as the node injection deviates further from the reference injections. That being said, the installed plant capacity in each node is arbitrary. On top of that, the demand and generation patterns are also randomly distributed over time. Therefore, the error in power flows will also show an arbitrary character.

An important factor that can be identified however is how a zone is interconnected to the rest of the grid. For example in the case of Turkey, that could be reduced all the way to the minimal size of one cluster per zone. The fact that Turkey is connected to the European power grid by just one power line ensures that the inter-zonal power flow remains equal for all levels of reduction. Whether the Turkish power grid is fully modeled or reduced to one cluster, the power flow in the respective line will remain equal.

4.4 Time vs. Accuracy balance

This section serves to gain insight in the trade-off between accuracy and computation time. The required computation time and error metric are calculated according to the method proposed in section 2.5.

The error metric and the required computation time are plotted against each other in Figure 20. This is done for all model sizes for which a simulation has been performed, ranging from 0%, 2.5%, 5% ... to 75% and finally 100%. The required computation time is taken as a share of time relative to the full grid model. Note that the required computation time on the X-axis is displayed in a logarithmic scale. The point in the far left bottom represents the full grid model for which no error occurs while the point in the far upper right corner represents the 0% model size.

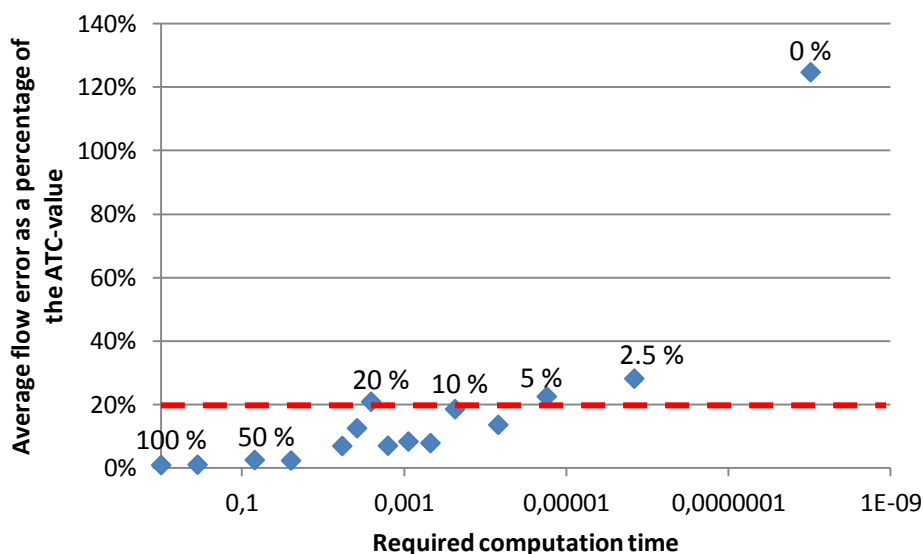


Figure 20. The relation between reduction and the average error of the line flows, for several reduction scales. The required computation time is represented as a share of computation time of the full grid, on a logarithmic scale.

This relation can serve researchers to make an estimation of how far they will reduce their transmission grid models. The allowable inaccuracy is of course dependent on the nature of one's analysis. Different purposes will require different levels of accuracy, and the same goes for computation time. By using this graph insight is gained in the balance between the simulation time that could be gained from simplifying a transmission grid into a smaller

number of clusters, and the accuracy that is sacrificed because of this. The graph clearly shows that as the relation between the number of variables and computation time is exponential. Significant gains can be made in computation time without reducing the accuracy of the model very much. As an example, we can read from the figure that when maintaining an allowable error of 20% of the ATC-value, the first time to cross this barrier is at a 20% grid size. This would mean the grid still performs within the limits of accuracy of this thesis at a 22.5% reduction. The required computation for this grid size is a factor 261 times smaller than for the original grid, when applied in an optimization model.

5 Critical reflection

An important point where the study could be complemented is on the data used to calculate generation capacity. The database of power plants only consists of large-capacity plants. The smaller scale capacity – e.g. private PV and wind – are not included. The injection patterns would likely be different if small-scale power production would also be involved. Additionally, there is no correlation between the time steps in the current approach. Because of the time-independent nature, hydro pump storage and hydro dams have not been involved in the analysis. In certain regions in particular, these plants are responsible for the main share of power production. Additionally, they might function as a battery for surrounding regions. This will be of significant influence on the power flows.

In addition to this, the method applied calculated the error for each zone based on the average error over time for all inter-zonal flows. For certain applications, it would be valuable to have insight in the maximum error occurring rather than the average error. For strategic net planning for example, it would be essential to obtain data on the maximum error occurring for a line over time when reduction is performed on a grid.

Also, it can be questioned to what extent the results from the PTDF-approach are extendable to models with a different approach. Future research would have to show whether the reduction method is accurate for alternative approaches as well. One of those approaches is using an optimal flow model rather than a PTDF-based approach for calculating power flows. Shi et al [1] suggests an accurate method for deriving the reactances of the reduced grid. The capacity of an equivalent line can be found by adding up the capacity of the lines it represents in the full grid. By doing so, the properties of the equivalent lines of the reduced grid are established. This can be used to simulate the reduced transmission grid in an optimization model. An optimal flow model offers the possibility to avoid a number of inaccuracies in the current approach by adding additional constraints. For example, the occurrence of opposing power flows between countries could be prevented. In the current model it might occur that country A both imports from - and exports to country B over different power lines. This would not occur normally in realistic power markets.

6 Conclusion

In this section, the research questions stated in the introduction will be answered.

1. Power grid models can be clustered (reduced) to a certain degree, ranging from full-scale, unreduced models to models with just few nodes. What is the relation between granularity and error size?

The allowable reduction degree is defined as the reduction level for which an average inaccuracy no greater than 20% of the ATC-value of a power line is maintained. This analysis is conducted for every zone separately, only reviewing the inter-zonal flows of a zone. The results show that every zone can be reduced to 37.5% of its original size, without leading to inaccuracy greater than the set benchmark. Most other countries can be reduced to even smaller sizes before the analyzed set of cross-border lines show major inaccuracies. The second research question will answer whether or not characteristics exist that explain these differences.

2. Topologies of grid models vary in several characteristics. Can we reduce the errors of reduction algorithms when we make use of certain characteristics of the particular power grid?

A deeper analysis of the results (Figure 19) shows that the allowable error per zone is not directly correlated with the reactances of power lines in a zone. They are dependent on case specific situations in which especially node injections play an important role. The node capacity, as well as the node injections, is distributed randomly however. This finding emphasizes the importance of accurate node injections. An important factor which is decisive in how far a zone can be reduced however is the way a zone is connected to the rest of the grid. The example of Turkey shows that when a zone is connected by only one singular line, the cross-border power flow matches the full grid for all reduction cases.

3. The reduction of power grid models will lead to improved computation times. What is the relation between error size and required calculation time?

Of course it depends on the purpose of the analysis as to what level of inaccuracy is acceptable. Therefore, a relation has been identified between the model size and the

accuracy. This relation can function as a guide line for transmission grid modelers that desire to reduce their grid model size without allowing the inaccuracy to increase above certain levels. The results show that the models can be reduced to a large extent, without the losing too much accuracy. A significant gain in computation time is therefore possible. For the set benchmark of accuracy of this thesis, the model could be reduced to 22.5% of its original size. When applied in an optimization model, this would lead to a computation time 261 times smaller than the full grid model.

7 Literature overview

- [1] D. J. T. Di Shi, "An Improved Bus Aggregation Technique for Generating Network Equivalents," in *Power and Energy Society General Meeting*, San Diego, USA, 2012.
- [2] G. A. Antonis Papaemmanouil, "On the reduction of large power system models for power market simulations," ETH Zurich, Zurich, 2011.
- [3] A. Ortner, "Transmission grid representations in power system models - a PDDF-based approach," in *NYC IAEE 2014*, New York, USA, 2014.
- [4] J. Ward, "Equivalent circuits for power flow studies," *AIEE Trans. Power App. Syst*, vol. February, no. 8, pp. 373-380, 1949.
- [5] P. Digo, *Nodal Analysis of Power Systems*, Abacus Press, 1975.
- [6] S. S. H.K. Singh, "A sensitivity based network reduction technique for power transfer assessment in deregulated electricity environment," in *Transmission and Distribution Conference and Exhibition*, Asia Pacific, 2002.
- [7] T. J. O. Xu Sheng, "PDDF-Based Power System Equivalents," *Transactions on Power Systems*, vol. 20, no. 4, pp. 1868-1876, 2005.
- [8] M. Liu, "Role of distribution factors in congestion revenue rights management," *IEEE Transactions on*, pp. 802-810, 04 05 2004.
- [9] BETTER, "BETTER project," BETTER, 2014. [Online]. Available: <http://better-project.net/>. [Accessed 21 10 2014].
- [10] N. Karmarkar, "A new polynomial-time algorithm for linear programming," in *STOC '84 Proceedings of the sixteenth annual ACM symposium on Theory of computing*, New York, 1984.
- [11] E. E. Group, "Energy Economics Group," Energy Economics Group, 01 01 2014. [Online]. Available: <http://www.eeg.tuwien.ac.at/>. [Accessed 24 08 2014].
- [12] S. E.-h. web, "SMHI E-hype web," SMHI, 13 03 2013. [Online]. Available: <http://e-hypeweb.smhi.se/timeseries/>. [Accessed 09 09 2014].
- [13] ENTSOE-E, "ENTSOE-E," ENTSOE-E, 2010. [Online]. Available: <https://www.entsoe.net/transmission-domain/ntcDay/show>. [Accessed 24 08 2014].

Appendix 1 – MATLAB script clustering process

```
number_zones = max(mpc.bus(:,6)); %count number of zones
num_nodes_in_zone = zeros(number_zones,1); %count number nodes in each zone
num_clusters_in_zone = zeros(number_zones,1); %count number clusters in each zone

clusterintensity = 0.5; %targeted granularity
number_nodes = length(mpc.bus(:,1)); %total number of nodes
num_lines = length(mpc.branch(:,1)); %total number of lines
mpc.bus(:,7) = 0; %column with cluster number, initially all 0

%calculate clustering condition value = resistance*max(dgen)/thermal capacity
for i=1:num_lines
mpc.branch(i,20) = abs(mpc.branch(i,11))*abs(mpc.bus(mpc.branch(i,18),12)-
mpc.bus(mpc.branch(i,19),12))/mpc.branch(i,13));
end

%normalize condition index from 0 to 1
max_threshold = max(mpc.branch(:,20));
mpc.branch(:,20) = mpc.branch(:,20)/max_threshold;

g=0;

% assign cluster ID's based on clustering condition, clustering process gets repeated while increasing the
condition threshold
% until desired zone size is reached

for j=1:number_zones %clustering process repeated for all zones
    num_nodes_in_zone(j,1) = length(mpc.bus(mpc.bus(:,6) ==j));
    num_clusters_in_zone(j,1) = num_nodes_in_zone(j,1);
% number of clusters in zone is initially equal to the number of nodes

    threshold=0.00000000001; % very low starting value of the threshold for each country

%selection of only nodes of the current zone
    zone_nodes_IND = find(mpc.bus(:,6) ==j);
    zone_nodes = mpc.bus(zone_nodes_IND,:);

% selection of only branches IN the current zone
    zone_branches_IND = find(mpc.branch(:,21)==j & mpc.branch(:,22)==j);
    zone_branches = mpc.branch(zone_branches_IND,:);

%count number of branches in zone
    num_branches_zone = length(zone_branches(:,1));

%exclude zone numbers that do not contain a grid
    if length(zone_nodes(:,1)) > 0

        t = 0; % counting variable to prevent infinite looping.

while num_clusters_in_zone(j,1)>=(clusterintensity*num_nodes_in_zone(j,1)) & t<=300
%condition determining whether the cluster process for this zone is ready or not, by comparing to the targeted
granularity. At t = 300, threshold value is 1 and all lines should be clustered.
```

```

for i=1:num_branches_zone %check all lines in zone

% condition value comparison to threshold for line aggregation
if (zone_branches(i,20)<threshold)

%create clustersID's
%find the from_node of a branch in the zone nodes
    from_IND = find(mpc.bus(:,11) == zone_branches(i,18));

%find the to_node of a branch in the zone nodes
    to_IND = find(mpc.bus(:,11) == zone_branches(i,19));

% if both nodes from a branch already have a cluster number, the to_node gets the number of the from_node
if mpc.bus(from_IND,7) ~= 0 && mpc.bus(to_IND,7) ~= 0

    replace_cluster = mpc.bus(to_IND,7);
    for l=1:number_nodes
        if mpc.bus(l,7) == replace_cluster;
            mpc.bus(l,7) = mpc.bus(from_IND,7);
        end
    end

% if only the from_node already has a cluster number, the to_node gets the number of the from_node
elseif mpc.bus(from_IND,7) ~= 0
    mpc.bus(to_IND,7) = mpc.bus(from_IND,7);

% if only the to_node already has a cluster number, the from_node gets the number of the to_node
elseif mpc.bus(to_IND,7) ~= 0
    mpc.bus(from_IND,7) = mpc.bus(to_IND,7);

% if none of the nodes of a branch has a cluster number, the cluster counter is going up 1, and both from_node
and to_node get this cluster number assigned
else
    g=g+1;
    mpc.bus(from_IND,7) =g;
    mpc.bus(to_IND,7) =g;
end
end

%number all unnumbered nodes in the zone
for b=1:number_nodes
if mpc.bus(b,6) == j
if mpc.bus(b,7) == 0
    g = g+1;
    mpc.bus(b,7) = g;
end

end
end

%create again a matrix of the busses in the zone
zone_nodes = mpc.bus(zone_nodes_IND,:);
%count the clusters in the zone
num_clusters_in_zone(j,1) = length(unique(zone_nodes(:,7)));

```

```
% increase the threshold
threshold = threshold*1.1;

t= t +1; % counting variable goes up

% if the number of clusters is not low enough yet, the while loop is repeated with a higher threshold value
end
end
end
```

Appendix 2 – Matlab script PTDF calculation

```
% Build the incidence matrix C_inc
C_Inc = (kron(ones(1,length(mpc.bus(:,1))),mpc.branch(:,18)) == ones(length(mpc.branch(:,1)),1) *
[1:length(mpc.bus(:,1))] + ... (-1).*(kron(ones(1,length(mpc.bus(:,1))),mpc.branch(:,19)) ==
ones(length(mpc.branch(:,1)),1) * [1:length(mpc.bus(:,1))]);

% Build the reduced incidence matrix (without slack node)
C_IncWoS = C_Inc;
C_IncWoS(:,mpc.bus(:,8)==3) = [];
% Build the reduced node (bus) matrix (without slack node)
mpc.busWoS = mpc.bus;
mpc.busWoS(mpc.busWoS(:,8)==3,:) = [];

B_flow = diag(mpc.branch(:,12).^(-1)) * C_IncWoS;
B_bus = C_IncWoS.' * B_flow;

[U, S, V]= svd(B_bus);
s = diag(S);
k = sum(s> 1e-9); % simple thresholding based decision
B_bus_inv = (U(:, 1:k)* diag(1./ s(1:k))* V(:, 1:k))';

PTDF = B_flow * B_bus_inv;

% Results for the power flows of the original grid
P_flow_lines = PTDF * mpc.busWoS(:,12);
```


Appendix 3 – Available net transfer capacities

	AT	BE	DK	FI	FR	DE	GR	IE/NI	IT	LU	NL	PT	ES
AT						1600			200				
BE					1300						2300		
DK						2050							
FI													
FR		2900				2600			2400				1200
DE	1600		1500		3200					980	4000		
GR									500				
IE/NI													
IT	70				870		500						
LU													
NL		2200				3900							
PT													1200
ES					500							1200	
SE			1640	2050		600							
GB					2000			410			1016		
CZ	800												
EE				350									
HU	350												
LV													
LT													
PL						1200							
SK													
SI	900								330				
BG							550						
RO													
TK							92						
HR													
RS													
BA													
MK							300						
AL							100						
NO			950								700		
CH	1000				1100	4400			3460				
ME													
KO													
MA													600

	SE	GB	CZ	EE	HU	LV	LT	PL	SK	SI	BG	RO	TK	HR
AT			600		500					900				
BE														
DK	2070													
FI	1650			658										
FR		2000												
DE	600		800					800						
GR											500		67	
IE/N I		80												
IT										120				
LU														
NL														
PT														
ES														
SE								600						
GB														
CZ								800	2100					
EE						500								
HU									500			600		1000
LV				500			1100							
LT						1250								
PL	600		1900						600					
SK			1100		1150			500						
SI														1000
BG												600	133	700
RO					700						600			
TK											183			
HR					800					1000				
RS					700						300	500		450
BA														600
MK											200			
AL														
NO	3545													
CH														
ME														
KO														
MA														

	MK	AL	NO	CH	ME	KO	MA
AT				1000			
BE							
DK			950				
FI							
FR				3000			
DE				2060			
GR	400	150					
IE/NI							
IT				1440			
LU							
NL			700				
PT							
ES							900
SE			3700				
GB							
CZ							
EE							
HU							
LV							
LT							
PL							
SK							
SI							
BG	400						
RO							
TK							
HR							
RS	500	210			450		
BA					400		
MK							
AL					200		
NO							
CH							
ME		200					
KO							
MA							