

Predicting Measurements at Unobserved Locations in an Electrical Transmission System

Dirk Surmann
Uwe Ligges
Claus Weihs

January 6, 2016

Electrical Transmission Systems consist of a huge number of locations (nodes) with different types of measurements available. Our aim is to derive a subset of nodes which contains almost sufficient information to describe the whole energy network. We derive a parameter set which characterises every single measuring location or node, respectively. Via analysing the behaviour of each node with respect to its neighbours, we construct a feasible random field metamodel over the whole transmission system. The metamodel works in a discrete spatial domain to smooth the measurements across the network. In the next step we work with a subset of locations to predict the unobserved ones. We derive different graph kernels to define the missing covariance matrix from the neighbourhood structures of the network. This results in a metamodel that is able to predict unobserved locations in a spatial domain with non-isotropic distance functions.

1 Introduction

The increasing consumption and trading of electrical power puts a strain on the electrical transmission system. Getting close to operational limits of the transmission lines and other components of the electrical system, different issues are observed in the energy network like overloads on transmission lines and other components. By reducing the security margins of all components to manage the higher loads in the electrical system, it is more vulnerable against additional issues. The higher vulnerability can cause a higher risk of blackouts and the more frequent usage of cold reserve to stabilise the energy network. Applications for monitoring and protecting wide-area networks, like the European network, are developed in MÜLLER et al. (2012).

One issue is the part of energy that permanently oscillates through the electrical transmission system with a low frequency of less than 2 Hz. The so called Low Frequency Oscillation (ROGERS 2000) is an inter-area oscillation between different areas in the European energy network. The main mode of the Low Frequency Oscillation (short LFO) is stable although its damping is small. For this reason the inter-area oscillation can increase uncontrollably in case of larger disturbances. The stability of an electrical transmission system and the information received by cluster algorithms is analysed in BRATO et al. (2014). Furthermore, the energy oscillating with low frequency restricts the available transfer capacity for power transmissions in Europe. Currently, the inter-area oscillation is damped by controllers at power plants like power system stabilizers (KUNDUR 1994, ch. 12.5; ROGERS 2000). These controllers have to be adjusted carefully over a long time considering all other controllers available in the energy network. Unfortunately, it is still not possible to damp the LFO completely.

To improve the quality of the electrical transmission system it is necessary to monitor the LFO. The LFO was described in a first step by a simple harmonic oscillator in SURMANN; LIGGES and WEIHS (2014) which results in a smaller parameter set. In addition, it is possible to interpret the parameters of the differential equations. Our task is to select locations for phasor measurement units to get the important information about the LFOs. Because phasor measurement units are not that expensive, one can argue to install a unit at every busbar. However, this will result in a huge amount of data for Europe, which is not processable at all. We aim at a solution for this challenge by using a genetic algorithm which selects the optimal subset of nodes with respect to a minimal prediction error. The error is calculated between a meta-model for the parameter set (SURMANN; LIGGES and WEIHS 2014) and the data from the institute of Energy Systems, Energy Efficiency and Energy Economics at TU Dortmund.

In section 2 the database for the analysis of the New England Test System is described. The paper describes a random field model for the parameter sets in section 3 to calculate the prediction error. A simulation study over the kernel parameters is described in section 4. The results are summarised in section 5. Further work is specified in section 6 where we consider a genetic algorithm to optimise the subset of nodes with respect to the prediction error.

2 Database

Currently very few phasor measurement units are available in the European electrical transmission grid, hence data are rather incomplete. Furthermore the different operators of the European electrical transmission system mark their existing data as confidential. Finally, it is not possible to simulate the corresponding data of this huge grid. For these reasons the following analysis deals with simulated data from a smaller electrical system, the New England Test System (short NETS). The simulation software that is used to generate the data, is the well established and complex simulation system DIgSILENT PowerFactory. A precise description of the algorithm used in PowerFactory is given in KUNDUR (1994).

2.1 The New England Test System

The New England Test System (short NETS) consists of 30 busbars, which are connected over 46 transmission lines. 19 loads consume the power from ten generators. Analogous to the European high voltage transmission grid, the NETS operates at 50 Hz with a voltage of 345 kV. Exceptions are the two busbars 12 and 20 which operate with a voltage of 220 kV. The total capacity is about 5.6 GW.

Figure 1 sketches the layout of the electrical transmission network. At each node, which is

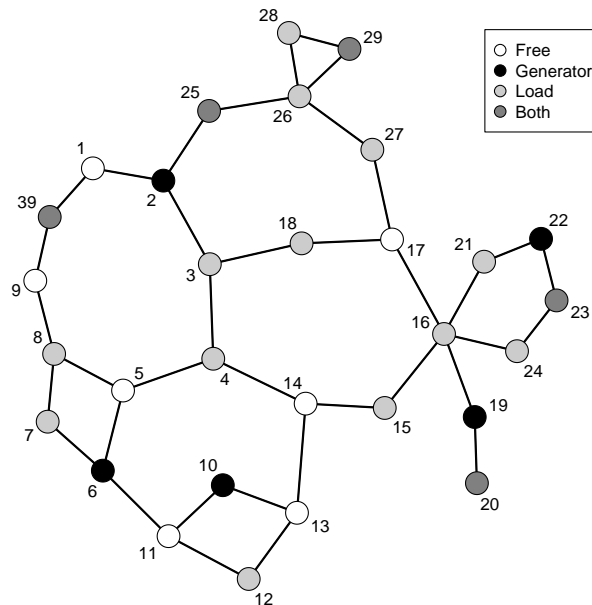


Figure 1: The transmission network New England Test System

equivalent to a busbar, we are able to measure data. The numbering corresponds to the present simulation in PowerFactory. White nodes indicate busbars which do not induct or consume power, because merely transmission lines are connected. Electrical energy is inducted into the network in black nodes with a power plant. Loads, which appear at light grey nodes, consume this power. In dark grey nodes we find both, a power plant and a load.

The graph structure of both networks is available in an symmetric adjacency matrix (WEST 2001). All main diagonal elements are zero, because an transmission network contains no self loops. An off diagonal element is set to the impedance of a transmission line between its corresponding nodes. It is set 0 if two nodes are not directly connected.

2.2 Measured Data

Starting from a static situation in PowerFactory, an event excites the network and generates dynamic data. It is necessary to choose events which do not modify the network topology. The increase of the consumed power and reduction shortly thereafter does this job well. We observe the complex valued voltage (voltage magnitude and angle) at every busbar with a time step size of 10 ms. This step size is fine enough to analyse the LFO, because its frequency

is lower than 2 Hz.

The considered event raises the loads at the busbars 4, 8, 12, 15, 18, 20, 21, 24, 25, 27, 28 and 39 by 10 % after 1 s. After an additional second the loads are restored to their original value. PowerFactory can only simulate LFOs by stimulating the network with a load event and observe the LFOs after a transient time interval. Figure 2 shows the voltage magnitude at busbars 14 and 25 as dotted lines after the transient time interval of 25 s up to 180 s. The result of the load event is an oscillation around the same level as in the static situation.

2.3 Representative Parameter Set

SURMANN; LIGGES and WEIHS (2014) use a harmonic oscillator to generate a representative parameter set for the oscillating measurements at each node. It is described by the homogeneous differential equation:

$$0 = m\ddot{x}(t) + d\dot{x}(t) + k(x(t) - r). \quad (1)$$

The position of the mass is described by the solution function $x(t)$. From a deflection $x(t)$, the spring generates the restoring force $F_k(t) = k(x(t) - r)$. Using a spring of length r generates an offset in the observed oscillation. The velocity of the mass is described by the first derivative with respect to time $\dot{x}(t) = \frac{d}{dt}(x(t) - r)$. A damper slows the motion by the damping force $F_d(t) = d\dot{x}(t)$ whereas the mass accelerates with $\ddot{x}(t) = \frac{d^2}{dt^2}(x(t) - r)$. KAKIMOTO et al. (2006) propose a similar model to simulate the LFO between Fukuoka and Shizuoka in Japan.

Estimating the parameters of the differential equation for each node of the NETS generates a parameter set of five parameters: the damping constant d_i , the spring constant k_i , the length from the ground r_i , the starting position $x_{0,i}$ and the starting velocity $v_{0,i}$. Table 1 summarise an extract of the estimated model coefficients. The corresponding graphs show the model as a grey solid line fitted to the data (dotted line) of two busbars in figure 2. All estimates of the

Table 1: Coefficients of differential equation model

busbar	d	k	r	x_0	v_0	SSE
14	0.0582	2.104	351.8	351.6	-1.503	3.419
25	0.0568	2.103	365.5	365.2	-1.565	59.591

spring constants are similar; leading to a correct frequency of about 0.163 Hz. The length from the fixed-point r , the starting position x_0 and the velocity v_0 are determined by the different data observed. Busbar 14 shows a good fit with a sum of squared errors (short: SSE) of 3.4, whereas the fit to the data of busbar 25 is the worst fit with an SSE of 59.6. As we can see in figure 2, the high deviance results from the fact, that a slower overtone appears at busbar 25. Due to the fact that one harmonic oscillator is not capable to model a superposed oscillation, the deviance rises.

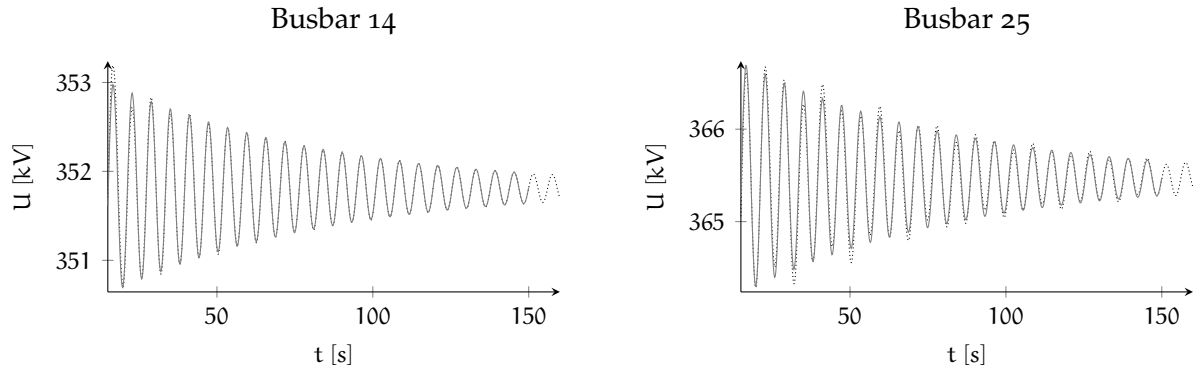


Figure 2: Fitted differential equation model (grey solid line) to the voltage magnitude data (dotted line) of two busbars

3 Spatial Modelling and Prediction

In this section we define a model to fit the data from [section 2](#). The model should be applicable to predict unobserved locations, to calculate a prediction error.

Information about the voltage in an electrical transmission grid are available at discrete locations $s \in \{1, \dots, l\}$. These are described by the nodes in [figure 1](#). Additionally, the graph shows the neighbourhood structure of the energy network. An edge in the graph indicates the neighbourhood of two nodes r and s and we use the notation $r \sim s$. Self loops are not allowed thus s cannot be neighbour of s itself. $N(s)$ describes the set of all neighbours of location s .

3.1 Spatial Smoothing across Neighbourhoods

According to FAHRMEIER; KNEIB and LANG (2007, ch. 7.2.4) we assign a coefficient $f_{\text{loc}}(s) = \gamma_s$ to each location $s \in \{1, \dots, l\}$. By defining the design matrix \mathbf{Z} in [equation \(2\)](#), we are able to write the vector of evaluated locations $\mathbf{f}_{\text{loc}} = (f_{\text{loc}}(s_1), \dots, f_{\text{loc}}(s_n))^T$ as a linear model $\mathbf{Z}\boldsymbol{\gamma}$ with

$$\mathbf{Z}[i, s] = \begin{cases} 1 & \text{if } y_i \text{ is observed in location } s = s_i \\ 0 & \text{otherwise} \end{cases}. \quad (2)$$

Obviously, this results in the linear model $\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$ for the vector of observations $\mathbf{y} = (y_1, \dots, y_n)^T$ with the common error vector $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$.

Due to the fact that we left out the neighbourhood structure we got an overfitted model because of the high number of coefficients. Hence we need to construct a criterion to obtain smooth spatial effects in which the coefficients of neighbouring regions are similar. FAHRMEIER;

KNEIB and LANG (2007) propose the penalised least squares criterion

$$\text{PeLS}(\lambda) = \sum_{i=1}^n (y_i - f_{\text{loc}}(s_i))^2 + \lambda \sum_{s=2}^l \left(\sum_{r \in \mathcal{N}(s), r < s} (\gamma_r - \gamma_s)^2 \right) \quad (3)$$

with $\lambda \geq 0$. $\text{PeLS}(\lambda)$ consists of the classical sum of squares and a penalisation term. This second term uses squared differences between coefficients of neighbouring locations and takes every combination of neighbours into account. The penalisation of higher differences between neighbouring locations support a smooth spatial effect. We can write the penalisation term as a quadratic form $\lambda \boldsymbol{\gamma}^\top \mathbf{K} \boldsymbol{\gamma}$ with

$$\mathbf{K}[s, r] = \begin{cases} -1 & s \sim r \\ |\mathcal{N}(s)| & s = r \\ 0 & \text{otherwise} \end{cases} . \quad (4)$$

This compact form of [equation \(3\)](#) is the well known Tikhonov regularization (TIKHONOV et al. 1995) with an L2-penalty. As a result, we receive the penalised least square estimator $\hat{\boldsymbol{\gamma}} = (\mathbf{Z}^\top \mathbf{Z} + \lambda \mathbf{K})^{-1} \mathbf{Z}^\top \mathbf{y}$ which minimises $\text{PeLS}(\lambda)$.

3.2 Predict Unobserved Locations

Assume we observe the data y_i at locations $s_i \in \{1, \dots, l\}$, $i \in \{1, \dots, n\}$. We fit the model $\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \varepsilon$ from [section 3.1](#) with a covariance matrix $\text{Cov}(\mathbf{y}) = \boldsymbol{\Sigma}_D$ to the data. Leverage the model for a prediction \mathbf{y}_0 at k unobserved locations we have to assume the spatial correlation between the data and the unobservables which is the $k \times n$ covariance matrix $\boldsymbol{\Sigma}_{0D}$. $\boldsymbol{\Sigma}_0$ is the $k \times k$ covariance matrix among the unobservables. With these assumptions SCHABENBERGER and GOTWAY (2005, ch. 5.3.3) define the best linear unbiased predictor in [equation \(5\)](#) which is the conditional expectation of \mathbf{y}_0 given \mathbf{y} :

$$\hat{\mathbf{y}}_0 = \mathbb{E}(\mathbf{y}_0 | \mathbf{y}) = \boldsymbol{\gamma}_0 + \boldsymbol{\Sigma}_{0D} \boldsymbol{\Sigma}_D^{-1} (\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\gamma}}) . \quad (5)$$

$\boldsymbol{\gamma}_0 = \bar{\boldsymbol{\gamma}}$ is the estimated constant mean. [Equation \(6\)](#) gives the conditional covariance, respectively:

$$\text{Cov}(\mathbf{y}_0 | \mathbf{y}) = \boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_{0D} \boldsymbol{\Sigma}_D^{-1} \boldsymbol{\Sigma}_{D0} . \quad (6)$$

3.3 Graph Kernel

The model at hand is based on a known covariance matrix of the measurements. Though, the data from [section 2](#) only provides the neighbourhood structure.

A definition of a covariance matrix is built on theoretical concepts for a graph G which consists of a set of l locations $S = \{1, \dots, l\}$ and a set of edges $E \subset S \times S$. A location $s \in S$ is a

neighbour of another location $r \in S$, if they are connected by an edge which means $(s, r) \in E$ and we denote $s \sim r$. We consider undirected graphs, that is $(s, r) \in E \Leftrightarrow (r, s) \in E$, and without self-loops ($(s, s) \notin E \Leftrightarrow s \not\sim s$ for all $s \in S$).

We define the adjacency matrix $A \in \mathbb{R}^{l \times l}$ to store the neighbourhood structure of G by $A[s, r] = 1 \Leftrightarrow s \sim r$ and $A[s, r] = 0 \Leftrightarrow s \not\sim r$ with the locations $s, r \in \{1, \dots, l\}$. The adjacency matrix is symmetric by definition of G . It is possible to generalise the matrix A by introducing weighted graphs which associate a weight $w_{sr} > 0$ to signify the edge between location s and r . If $s \not\sim r$, then $w_{sr} = 0$.

In the next theoretical step, we utilise kernel methods due to the fact that they can be applied successfully in high dimensional feature space. Let \mathcal{X} be a non-empty set. Every edge-labelled graph G is associated with a label matrix $X \in \mathcal{X}^{l \times l}$. Thereby the label of edge $(s, r) \in E$ is $X[s, r]$ and a special label ζ if $s \sim r$ respectively. A function $\kappa : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is a kernel if there exists a real Hilbert space \mathcal{H} and a corresponding feature map $\phi : \mathcal{X} \rightarrow \mathcal{H}$ that maps ζ to the zero element of \mathcal{H} . ϕ is defined such that $\forall x, x^* \in \mathcal{X}, \kappa(x, x^*) := \langle \phi(x), \phi(x^*) \rangle_{\mathcal{H}}$. Extended by a kernel and the reproducing property $\forall x \in \mathcal{X}$ and $\forall f(\cdot) \in \mathcal{H}$ it holds $\langle f(\cdot), \kappa(\cdot, x) \rangle_{\mathcal{H}} = f(x)$ the Hilbert space \mathcal{K} is denoted as reproducing kernel Hilbert space (short: RKHS).

VISHWANATHAN et al. (2010, sec. 3.2) construct a kernel on a graph G with respect to all random walks of length t between each pair of locations. A walk of length t on a graph is a sequence of locations s_1, s_2, \dots, s_{t+1} such that $s_k \sim s_{k+1}$ for all $1 \leq k \leq t$. This idea reflects the power transmission in an energy network from the statistical point of view in an adequate way. Energy flows in a transmission grid between two locations with respect to Kirchhoff's circuit laws, hence it is possible to use arbitrary walks between two locations. Admittedly, the longer a walk the less energy is transmitted by this sequence which should be taken into account in the kernel definition. Based on all random walks on the graph G we define the kernel function κ with respect to the sum of the adjacency matrices A^t :

$$\kappa = \sum_{t=0}^{\infty} \lambda(t) A^t. \quad (7)$$

Therefore, the (s, r) entry of A^t represents the number of length t random walks between location s and r . To overcome the problem that [equation \(7\)](#) does not converge, VISHWANATHAN et al. (2010) introduce a sequence of appropriately chosen non-negative coefficients $\lambda(t)$. Additionally, the authors show in theorem 3, if the coefficients $\lambda(t)$ are such that κ in [equation \(7\)](#) converges, it defines a valid positive semi-definite kernel. By the use of the RKHS properties one can show that a positive definite kernel $\kappa(x, x^*)$ is the inner product in the Hilbert space \mathcal{H} between $\phi(x)$ and $\phi(x^*)$.

With the flexible definition in [equation \(7\)](#) two special cases of interest are discussed by GÄRTNER; FLACH and WROBEL (2003). Firstly, the exponential kernel with $\lambda(t) := \frac{\beta^t}{t!}$ is used so that the sum in [equation \(7\)](#) becomes the series expansion of the matrix exponential $e^{\beta A}$. As the adjacency matrix A is symmetric and $\beta > 0$, the matrix exponential $e^{\beta A}$ is positive definite and can be interpreted as a covariance matrix. Secondly, the geometric kernel with $\lambda(t) := \beta^t$ and $|\beta| < 1$ employs the geometric series $\sum_{t=0}^{\infty} \beta^t$ in [equation \(7\)](#). GÄRTNER; FLACH and WROBEL (2003) show that the geometric series of a matrix only converge if $\beta < \frac{1}{\alpha}$ with

$\alpha := \min \{ |N(s)|, s \in \{1, \dots, l\} \}$. Hence, the geometric series of a matrix can be interpreted also as a covariance matrix.

With these two kernels we are now able to define the covariance structure in [section 3.2](#) which was assumed to be known. Therefore it is possible to predict measurements at unobserved locations in a given energy transmission graph.

4 Simulation Study

We describe properties of the different graph kernels in [section 3.3](#) on the basis of a simulation study. The study deals with the type of voltage $V \in \{\text{angle, magnitude}\}$, the type of graph kernel $GK \in \{\text{exponential, geometric}\}$, and the type of adjacency matrix $A \in \{\text{adjacency, neighbourhood}\}$. Both graph kernels are parametrised by $\xi \in (0.25, 1)$, which is equal to β for the exponential kernel and a proportion of $\frac{1}{\alpha}$ for the geometric kernel. The adjacency matrix contains the impedance between two nodes, whereas the neighbourhood matrix contains the length of their shortest paths ([WEST 2001](#)). This will give us the information about the required accuracy of the neighbourhood structure from the network. Lastly we use a parameter $\theta \in (0.05, 0.95)$ to define the ratio of omitted nodes and thereby the number of nodes to predict.

Each parameter combination is repeated 20 times in the simulation. We measure the maximal absolute residual $\max(|e_i|)$ of the complete network after predicting the omitted nodes to evaluate the simulation. The maximal residual describes the situation at hand in a more realistic way than an average statistic, because one node with a poor prediction can result in a network failure.

The results are analysed by a linear regression with interaction and quadratic terms of the continuous parameters. Five measurements with standardized residuals higher than 5 are excluded from the analysis. A Box-Cox plot identifies a square root power transformation to improve the model. After fitting the model, a stepwise model selection by BIC reduces the number of terms to estimate from 18 to 8. This results in a residual standard error of 6.779×10^{-3} on 3827 degrees of freedom. The coefficient of determination and its adjusted counterpart both have the value $R^2_{(\text{adj})} = 0.88$. [Table 2](#) describes the estimated model coefficients and its standard errors. The interactions between two parameters are characterised by a colon. The interpretation of an estimated coefficient is done exemplary for two parameters. Each estimation for discrete parameters describe the difference between the values of this parameter with one coefficient for a value set to 0. An estimation of -2.89×10^{-2} for $V_{\text{magnitude}}$ describes a lesser result for the regression model compared to $V_{\text{amplitude}}$ which is 0. An estimation for a continuous parameter, e. g. θ , multiplies the value of the parameter with its corresponding coefficient.

Firstly, the simulation study shows a better maximal residual for the voltage magnitude than for its angle. Secondly, the network immanent adjacency matrix results in lower maximal residual values than the neighbourhood matrix. This result follows from the negative interaction between the voltage magnitude and the neighbourhood matrix. In a more detailed analysis of the simulation study we figure out that an additional parameter which manages

Table 2: Regression model for $\sqrt{\max(|e_i|)}$ of NETS reduced by BIC

Term	Estimate	Std. Error
Intercept	5.03×10^{-2}	4.24×10^{-4}
$V_{\text{magnitude}}$	-2.89×10^{-2}	4.63×10^{-4}
$A_{\text{neighbourhood}}$	1.69×10^{-2}	4.63×10^{-4}
θ	4.12×10^{-2}	1.51×10^{-3}
θ^2	-1.21×10^{-2}	1.39×10^{-3}
$V_{\text{magnitude}} : A_{\text{neighbourhood}}$	-3.56×10^{-3}	4.38×10^{-4}
$V_{\text{magnitude}} : \theta$	-7.34×10^{-3}	6.89×10^{-4}
$A_{\text{neighbourhood}} : \theta$	-1.94×10^{-2}	6.89×10^{-4}

the maximal length of the shortest path available, does not improve this result. Thirdly, we observe the expected positive effect of θ on the maximal residual. Lastly, we see no significantly different performance between the exponential kernel and the geometric one. Furthermore, the parameter ξ of the graph kernel has no significant influence on the response at all.

The analysis is verified on another electrical transmission system called ‘New England Test System - New York Power System’, NETS – NYPS for short. This bigger network contains of 53 nodes and 87 transmission lines. The equivalent simulation study results in an $R^2_{(\text{adj})}$ of 0.76 and a residual standard error of 2.538×10^{-3} on 3829 degrees of freedom. [Table 3](#) shows the same significant terms in the reduced model as for the NETS system. We only observe

Table 3: Regression model for $\sqrt{\max(|e_i|)}$ of NETS – NYPS reduced by BIC

Term	Estimate	Std. Error
Intercept	1.52×10^{-2}	1.59×10^{-4}
$V_{\text{magnitude}}$	6.24×10^{-3}	1.73×10^{-4}
$A_{\text{neighbourhood}}$	5.44×10^{-3}	1.73×10^{-4}
θ	1.45×10^{-2}	5.65×10^{-4}
θ^2	-5.95×10^{-3}	5.19×10^{-4}
$V_{\text{magnitude}} : A_{\text{neighbourhood}}$	7.33×10^{-4}	1.64×10^{-4}
$V_{\text{magnitude}} : \theta$	1.73×10^{-3}	2.58×10^{-4}
$A_{\text{neighbourhood}} : \theta$	-7.76×10^{-3}	2.58×10^{-4}

differences comparing the leading signs between [table 2](#) and [table 3](#). The difference is due to the varying impact of the voltage magnitude and amplitude of both networks. All interesting kernel parameters are not affected.

5 Conclusion

The general target of this paper is the derivation of a model in a discrete spatial domain to interpolate and predict unobserved nodes. Interrogation with a subset needs less resources compared to process information from the whole electrical network. Due to the fact of miss-

ing real data we introduce a virtual test systems which is used to simulate data for every node with a frequency of 100 Hz. Each measurement at a locations is drilled down to a representative parameter set by a homogeneous differential equation. These new parameters represent the simulated measurements by interpretable coefficients. The most interesting one is the damping of the simulated oscillation.

We discuss a random field model for discrete locations which depends on the network structure in from of an adjacency matrix. It is used to obtain a smoothed spatial model across neighbourhoods. In the next step we assume to observe a subset of locations in the network and predict the unobserved ones. The theory of graph kernels is leveraged to define covariance matrices within the spatial domain and its non-isotropic distance function. This empowers us to predict locations that are not observed from a given subset of locations.

Finally, we utilise a simulation study to model the impact of kernel parameters on the residual error within the predicted locations. The study shows no significant difference between the two graph kernels and the corresponding parameters. It points out lower maximal residual values when using the network immanent adjacency matrix compared to the neighbourhood matrix. In summary, random field model with a graph kernel provide a powerful tool to predict locations in a graph with non-isotropic distance functions.

6 Further Work

Using the theory from [section 3](#) we are able to define a subset of locations as a training set s and predict the data y_0 at all remaining locations. Currently, the subset is chosen subjectively or randomly from the energy transmission network. To determine the optimal subset with respect to a minimal maximal error, a genetic algorithm will be used. One challenge will be the definition of a suitable mutation strategy. It should keep the number of locations in the subset constant. Without such a constraint the optimum will be to choose all nodes.

7 Acknowledgement

This work has been funded by the Deutsche Forschungsgemeinschaft (DFG), Forschergruppe 1511 *Schutz- und Leitsysteme zur zuverlässigen und sicheren elektrischen Energieübertragung*.

References

BRATO, Sebastian et al. (2014): *Stability Analysis and Clustering of Electrical Transmission Systems*. 2003/32864. Eldorado: Technische Universität Dortmund (cit. on p. 2).

FAHRMEIER, Ludwig; Thomas KNEIB and Stefan LANG (2007): *Regression*. 1st ed. Statistik und ihre Anwendungen. Berlin, Heidelberg: Springer. ISBN: 978-3-540-33932-8 (cit. on p. 5).

- GÄRTNER, Thomas; Peter FLACH and Stefan WROBEL (2003): 'On Graph Kernels: Hardness Results and Efficient Alternatives'. In: *Learning Theory and Kernel Machines*. Springer Berlin Heidelberg, pp. 129–143. ISBN: 978-3-540-45167-9 (cit. on p. 7).
- KAKIMOTO, Naoto et al. (2006): 'Monitoring of Interarea Oscillation Mode by Synchronized Phasor Measurement'. In: *IEEE Transactions on Power Systems* 21.1, pp. 260–268. ISSN: 0885-8950. DOI: [10.1109/TPWRS.2005.861960](https://doi.org/10.1109/TPWRS.2005.861960) (cit. on p. 4).
- KUNDUR, Prabha (1994): *Power System Stability and Control*. The EPRI Power System Engineering Series. McGraw-Hill. ISBN: 978-0-07-035958-1 (cit. on p. 2).
- MÜLLER, Sven Christian et al. (2012): 'New Applications for Wide-Area Monitoring, Protection and Control'. In: *IEEE PES International Conference on Innovative Smart Grid Technologies*. Berlin, Germany (cit. on p. 1).
- ROGERS, Graham (2000): *Power System Oscillations*. The Kluwer International Series in Engineering and Computer Science: Power Electronics and Power Systems. Kluwer Academic. ISBN: 978-0-7923-7712-2 (cit. on p. 2).
- SCHABENBERGER, Oliver and Carol A. GOTWAY (2005): *Statistical Methods for Spatial Data Analysis*. Texts in Statistical Science. Chapman & Hall/CRC. ISBN: 978-1-58488-322-7 (cit. on p. 6).
- SURMANN, Dirk; Uwe LIGGES and Claus WEIHS (2014): 'Modelling Low Frequency Oscillations in an Electrical System'. In: *Energy Conference (ENERGYCON), 2014 IEEE International*. Dubrovnik, Croatia: IEEE, pp. 565–571. DOI: [10.1109/ENERGYCON.2014.6850482](https://doi.org/10.1109/ENERGYCON.2014.6850482) (cit. on pp. 2, 4).
- TIKHONOV, A. N. et al. (1995): *Numerical Methods for the Solution of Ill-Posed Problems*. Kluwer Academic Publishers (cit. on p. 6).
- VISHWANATHAN, S. V. N. et al. (2010): 'Graph Kernels'. In: *The Journal of Machine Learning Research* 11, pp. 1201–1242 (cit. on p. 7).
- WEST, Douglas Brent (2001): *Introduction to Graph Theory*. 2nd ed. Prentice Hall. ISBN: 978-0-13-014400-3 (cit. on pp. 3, 8).