A Data-Driven Approach for Constructing a Zonal Network Model for Long-Term Dispatch Planning

Phen Chiak See

Abstract

The most important aspect of the planning of power generation dispatch is its complementary relationship with the power flows on the interconnectors of the transmission network. A plan could become invalid if the power flow created violates the transfer limit of any interconnectors. The long-term dispatch planning is more affected by this because of the relative difficulty in predicting the state of the transmission network in advance. It is generally planned in a trade-based manner, without a means to explicitly compute the power flows. Deviations in the plans are then corrected near the time of dispatch, in the expense of opportunity costs. In Europe, the flow-based market coupling is proposed in the Central Western Europe, which is an effective means for modeling the inter-zonal power flows and transfer capacity allocations. However, its usefulness for long-term planning is still limited. Especially, the Power Transfer Distribution Factors (PTDFs) of its model (key parameter for integrating power flow in dispatch planning) is stochastic and unpredictable. This paper introduces a data-driven approach to construct a zonal model for closing the gap between the long-term and the actual dispatch plan. It is able to reconstruct all zonal PTDFs existed in the system, by inverse modeling the ex-post power flow data. The paper as well presented the validity of decomposing a large zonal model into substantially smaller sub-problems, and the existence of clusters of ex-post power flow cases which share the same zonal PTDFs. These features have greatly simplified the implementation of the method.

Keywords: Power generation dispatch, flow-based market model, zonal modeling, inverse modeling, *k*-nearest neighborhood classification, genetic algorithm

I Introduction

The European electricity markets are undergoing a steady transformation towards a single market in electricity [1] through the adoption of Flow-Based Market Coupling (FBMC). The FBMC is practical, as it represents each of the existing market zones akin to a nodal element and performs electric power dispatch calculations in a way that resembles the nodal pricing approach. One of the key activities in FBMC is the creation of a common (zonal) network model which ideally represents the real inter-zonal power flows. A stylized example of the zonal modeling is shown in Figure 1. This enables the forward planning of the allocation of inter-zonal transfer capacity, and an optimal power dispatch plan for serving the regional electricity demand.

Yet, the zonal modeling in is still imperfect. It has limited usefulness in determining the volume of the inter-zonal power transfers at time much earlier than the actual hour of dispatch (i.e., weeks or months ahead). The values in its zonal Power Transfer Distribution Factors (PTDFs) matrix, Γ^{A} is dependent on the state of the system's operation. Specifically, they are determined by (i) the reactance of the transmission lines of the detailed network, and (ii) the share of power generation committed by each of the generators within a zone for the zonal production of the next unit of power. The latter is called the Generation Shift Keys (GSKs), Π_z [2]. Although the reactance of the transmission lines are constant, the Π_z is stochastic (as the Π_z is different from time to time). As a consequence, the Γ^{A} is not firm and several of them could exist in a zonal model. The Γ^{A} too is stochastic and indeterminate in value [3].

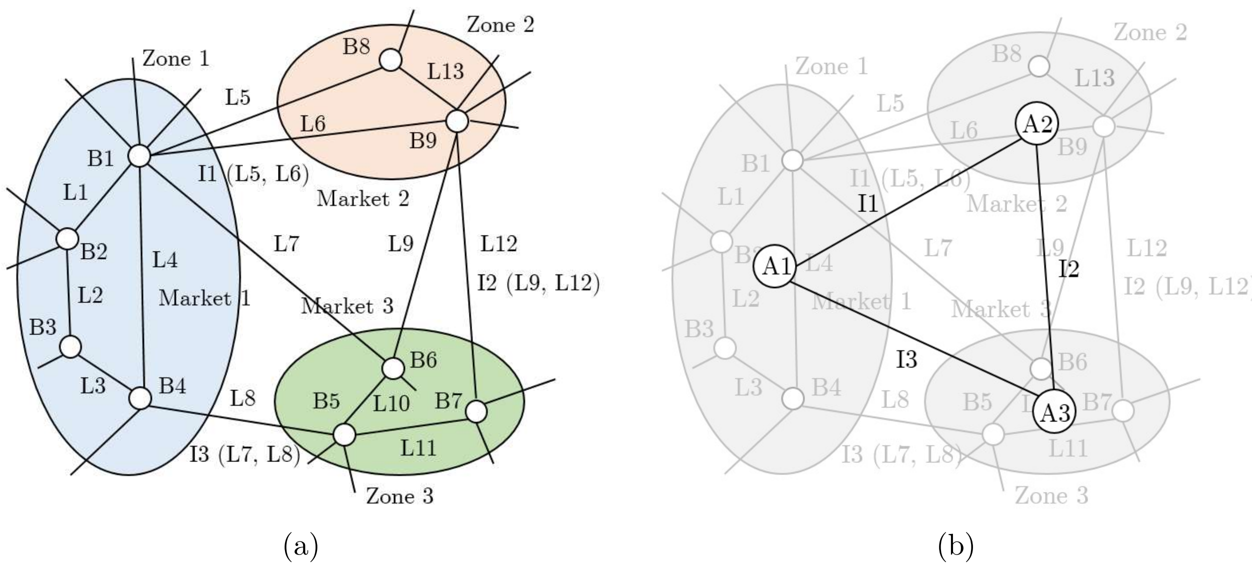


Figure 1. A stylized example of zonal modeling in FBMC: (a) a 9-bus transmission network model, (b) a 3-bus zonal model of the 9-bus transmission network

The actual inter-zonal power flows would deviate from those presumed by the market players. Often, the committed power generation plans has to be corrected (through re-dispatching) to ensure balancing in the supply and demand while not violating the transmission constraints. The zonal model is more firm at times closer to the dispatch, but imprecise for long-term dispatch planning. As pointed out by [4], trading long-term contracts in FBMC is difficult because of the inability to forecast future transmission capacity in the zonal model. Also, the FBMC currently implemented in CWE is only able to publish day-ahead Γ^{A} for the reference for the market players (the traders cannot anticipate the future Γ^{A}) [4].

In the authors' opinion, the restriction of having a proper zonal Γ^{A} information is caused by the model-based (bottom-up) calculations of zonal modeling. Although the details of the transmission network (the impedance and transfer limit of the transmission lines) are known to the Transmission System Operators (TSOs), they cannot regulate a fixed Π_z for all of the zone in advance. Also, the model-based approach is unable to consider all of the Π_z because they could theoretically exist in an infinite number of combinations.

This paper describes a zonal modeling alternative [5] that could close the gap between the long-term planning and the real-time market performance. It is a data-driven (top-down) inverse modeling approach which construct an equivalent zonal model based on the ex-post power flow data (the historical inter-zonal power flows data e.g., published by the TSOs). The most important steps in the inverse modeling is the discovery of the virtual reactance of the inter-zonal interconnectors. The method then compute all of the Γ^{A} existed in the ex-post data using the virtual reactance values. Each of the Γ^{A} is implicitly related to a Π_z at the time of their presence. As the elements in Γ^{A}, $\beta^{\mathrm{A}}_{z\longrightarrow z_0,m}$ are sensitivity values associated to the change in an inter-zonal interconnector created by the power generation at a zone, they could be directly used for computing the inter-zonal transfer scenarios associated to a long-term dispatch plan. For instance, one could assess the best and worst case impact of a power generation plan on the inter-zonal bottlenecks using the Γ^{A} found, by computing the lower and upper bound of power flow on them.

The paper is organized as follows. Section II introduces the theory of the inverse modeling, and two unique characteristics of the linear power flow formulations that lead to the substantial enhancement of the effectiveness of the method: (i) the valid decomposition of a larger network model into smaller sub-problems; and (ii) the existence of groups of power flow cases that could be discovered by a simple heuristic approach, the cases in each of the groups share an approximately identical Γ^{A}. Section III describes the complete implementation of the inverse modeling. Section IV demonstrates the test of the method on a sub-problem of the Norwegian transmission system comprises of three market zones. Finally, Section V presents the discussions and conclusions.

II The inverse modeling

Any real-world problems possess (i) a set of observed data, d and; (ii) a set of model parameters, M. A forward operator, G maps the M to d, forming a forward problem [6] [7]:

$$d = G(M) \tag{1}$$

For an inverse problem, the d is known, while M is unknown [8]. Thus, a solution to the inverse problem is achieved by reconstructing the M based on the known d:

$$M = G^{-1}(d) \tag{2}$$

where, G^{-1} is the inverse operator of the problem. In the inverse modeling of a transmission network, the d is the ex-post power flow on the transmission line k, F_k; and M is the set of reactance values of the transmission lines, X where $x_k \in X$ (x_k is the reactance of transmission like k).

The G is the linear power flow theory [9] that relates the X to the F_k as follows

$$F_k = \mathrm{B}_{ij}(\theta_i - \theta_j) \tag{3}$$

where, the i and j are the indices of the buses connected by the transmission line k; the θ_i and θ_j are the voltage angle at bus i and j; the B_{ij} is the admittance of the transmission line k. The θ_i and θ_j are derived by solving the following equation

$$\Theta = \mathrm{B}^{-1}\mathrm{P} \tag{4}$$

where, $\mathrm{P} = [P_1, \dots, P_{\mathrm{N}}]^\top$ contains the net (active) power generation at each bus, and $\Theta = [\theta_1, \dots, \theta_{\mathrm{N}}]^\top$ contains the bus voltage angles. The B is a two-dimension matrix defined based on the susceptance of the transmission lines, where $b_{ij} = 1/x_{ij}$ (the x_{ij} is the reactance of transmission line k which connects bus i and j), such that $\mathrm{B}_{ij} = \mathrm{B}_{ji} = -b_{ij}$; $\forall i \neq j$ and $\mathrm{B}_{ii} = \sum_{j\neq i}^{\mathrm{N}} b_{ij}$.

The P_i (in (4)) is also the sum of net export of the transmission lines connected to bus i. Clearly, as d is well-defined (i.e., by the ex-post power flow data) while the M is unknown, the problem is an inverse modeling problem (as in (2)). The problem is solved by adopting a G^{-1} for finding the value of X (the M) by performing the following iterative steps: (i) find the approximate X (henceforth denoted as X^*), which comprises the estimated value of x_k of the transmission lines, and compute the B matrix, (ii) calculate the Θ and the estimated F_k (henceforth denoted as F_k^*) using (3), and (iii) compare the F_k^* with the known F_k. A solution (the optimum X^*) is found when either all of the found F_k^* is equal to the F_k, or deviate from it by a small error margin, ϵ. The objective function, ϕ of the inverse problem is as follows

$$\phi = \sum_{k=1}^{\mathrm{K}} |F_k^* - F_k| \approx 0 \tag{5}$$

The authors adopted the Genetic Algorithm (GA) [10] as the solver of the mentioned iterative steps because the problem is nonlinear [5]. Both X^* and Θ are unknown, and has to be solved in an iterative manner. The GA is thus the G^{-1}. Once the X^* is derived, we can calculate the PTDF matrix of the transmission network, Γ as follows

$$\Gamma = \mathrm{B}_f \mathrm{B}_r^{-1} \tag{6}$$

where, B_r^{-1} is the inverse of the reduced B matrix (as in (4)), within which the B_{ij} associated with an arbitrarily selected reference bus (bus i_0) is removed. The B_f is a two-dimension matrix where each of its row is associated with transmission line k; in each row, the value is b_{ij} at column i, and $-b_{ij}$ at column j. The column associated with the reference bus is as well removed. The derived Γ is a two-dimension matrix where the row index corresponse to transmission line k and each column corresponds to bus i. As the Γ is computed using the X^*, it is as well an approximated parameter. It is henceforth denoted as Γ^*.

The derived Γ^* is a two-dimension matrix where the row index corresponds to transmission line k and each column corresponds to bus i. The elements of Γ^* are henceforth denoted as $\beta_{i\longrightarrow i_0,k}$. They represent the change on transmission line k, when a unit of power is dispatched from bus i to bus i_0. For an inversed zonal modeling, the computation of (6) would yield the $\Gamma^{\mathrm{A},*}$, within which are all of the $\beta^{\mathrm{A}}_{z\longrightarrow z_0,m}$ values, which are the linear sensitivity factor which reflect the change in the power flow on the interconnector m when a unit of power is dispatched from zone z to a reference zone, z_0. Subsequently, the calculation of the inter-zonal power flow in a dispatch plan is simplified as follows

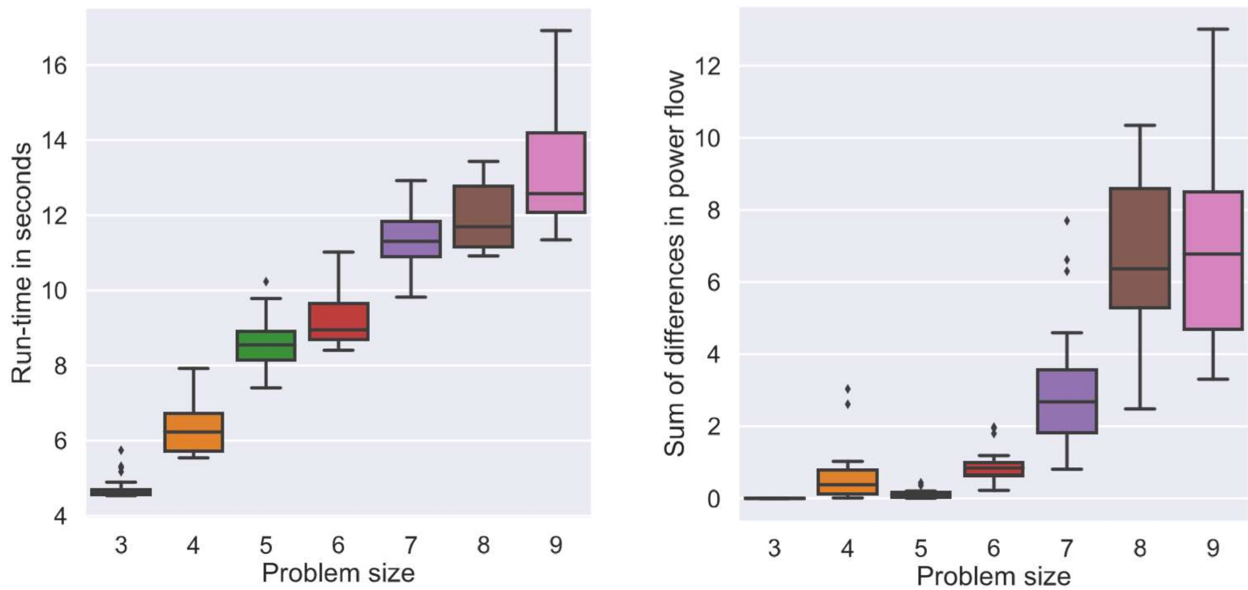
$$F_m = \sum_{z=1}^{\mathrm{Z}} P_z^{\mathrm{A}} \beta^{\mathrm{A}}_{z\longrightarrow z_0,m} \tag{7}$$

where, the F_m is the power flow on an inter-zonal interconnector m; and P_z^{A} is the net power generation in zone z.

2.1 Decomposition of the inverse problem

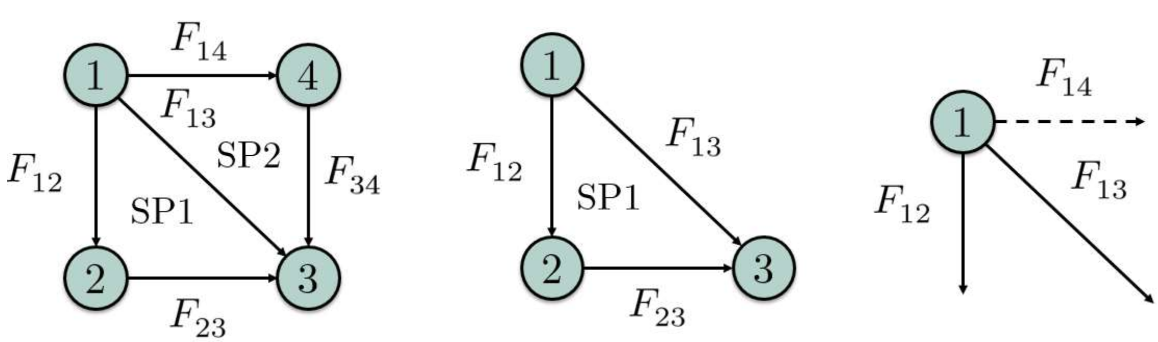
The main issue of using the GA as the G^{-1} is the stochastic nature of its implementation (and its solution), and the rapid increase in its run-time when solving larger problems. The GA is an approximate method, which is an analytical procedure for developing solutions that are close to the exact but unknown solution of the nonlinear problem [11]. Thus, the solution derived at the end of its implementation may vary in different repetition of its implementation. Nonetheless, the performance of GA is often comparable to other approach when it is solving a small problem. An example is given in Figure 2, which shows (i) the run-time (in seconds) and (ii) the sum of differences in power flow on the transmission lines (i.e., the ϕ) derived from testing the GA in solving the inverse problems of size 3 to 9. The GA is repeated 30 times when solving each of the problems.

The median of both run-time and ϕ increases with the size of the problem, N solved. The range of the data also increased with N. A significant deviation in ϕ is seen in the 8-bus and 9 bus problems. A greater spread in run-time too is seen in the 9-bus problem. Ideally, the ϕ of the solution should be kept as small as possible (or $\phi < \epsilon$) (as in (5)). The figure has also showed that the quality of the solution is poor when the problem size is larger than six. The best performance is achieved for the 3-bus problem, as its run-time is minimum and its $\phi \approx 0$. In other words, problems of smaller size are more tractable.



(a) (b)

Figure 2. The performance (run-time and quality of solutions found) of GA for solving the inverse problems of different size: (a) the run-time required by the solver for finding an optimal solution; (b) the sum of differences in power flow at the interconnectors for the solutions found



(a) (b) (c)

Figure 3. A simple example of decomposition performed on a 4-bus network model. The model is separated as two sub-problems (SP1 and SP2) of three buses: (a) the original 4-bus transmission network model, (b) one of the decomposed sub-problems, created by virtually removing bus 4 from the original network, and (c) the net power generation at bus 1 is readjusted after the decomposition by subtracting F_{14} from its original value

Fortunately, the inverse problem can always be decomposed as several small sub-problems. As shown in Figure 3, a 4-bus network model can be separated as two sub-problems (SP1 and SP2).

The ex-post power flow on all of the transmission lines is same after the separation. The net power generation on each of the common buses is adjusted by subtracting the power flow of the "disconnected" transmission lines from the original value. For example, the net power generation at bus 1 (as in Figure 3 (a)), $F_{12}+F_{13}+F_{14}$ becomes $F_{12}+F_{13}$ in SP1 (Figure 3(b)) as the transmission line connecting bus 1 and bus 4 is virtually removed from it after the decomposition (Figure 3 (c)).

Its validity could be explained by observing the P_i at a bus that co-exists in a sub-problem before and after the decomposition based on (3):

$$P_i = \sum_{k=1}^{\mathrm{K}} \mathrm{B}_k \Delta\theta_k \tag{8}$$

where, k is the index of transmission line connecting bus i and j, and $\Delta\theta_k$ is the difference in voltage angle of transmission line k, or $\Delta\theta_k = \theta_i - \theta_j$. Before a decomposition is conducted, P_i at the bus is as follows

$$P_i \;=\; \mathrm{B}_1\Delta\theta_i \;\dots\; + \mathrm{B}_{\mathrm{K}-1}\Delta\theta_{\mathrm{K}-1} + \mathrm{B}_{\mathrm{K}}\Delta\theta_{\mathrm{K}} \tag{9}$$

The net power generation at the bus is reformulated as below after the decomposition as transmission line K is disconnected from it. The P_i is therefore, reduced by $\mathrm{B}_{\mathrm{K}}\Delta\theta_{\mathrm{K}}$:

$$\begin{aligned} P_i - \mathrm{B}_{\mathrm{K}}\Delta\theta_{\mathrm{K}} &= \mathrm{B}_1\Delta\theta_i \;\dots\; + \mathrm{B}_{\mathrm{K}-1}\Delta\theta_{\mathrm{K}-1} + \mathrm{B}_{\mathrm{K}}\Delta\theta_{\mathrm{K}} - \mathrm{B}_{\mathrm{K}}\Delta\theta_{\mathrm{K}} \\ &= \mathrm{B}_1\Delta\theta_i \;\dots\; + \mathrm{B}_{\mathrm{K}-1}\Delta\theta_{\mathrm{K}-1} \end{aligned} \tag{10}$$

As the B_k value is constant, the $\Delta\theta_k$ must be equal in both (9) and (10). All of the remaining transmission line that are connected with bus i would retain the same value of F_k after the decomposition. Thus, each of the sub-problems can be solved independently. The fact that the power flow remain the same on the transmission line that co-exist in the sub-problems too allows the recombination of the solutions of the sub-problems to reconstruct the complete zonal model. This further justified the validity of the decomposition strategy.

2.2 The clustering of ex-post power flow data

Another issue of the inverse modeling is the existence of multiple sub-optimal X^* which fulfills the requirement of $\phi \approx 0$ in (5). Only one of them is the real solution of the problem. This is because of the existence of multiple X^* that could yield the same B matrix used in (3). The existence of multiple solutions for a power flow problem is also a well-known issue [12]. This can be solved by including more than one ex-post power flow cases in (5) as follows [5] [13]

$$\phi = \sum_{s}^{\mathrm{S}} \sum_{k=1}^{\mathrm{K}} |F_k^{*,s} - F_k^s| \approx 0 \tag{11}$$

where, s is the index of a power flow case.

The cases has to be selected carefully. This is because, several Γ^{A} actually exist in the zonal model. Each of them is implicitly related to a unique Π_z [13]. In other words, several clusters of cases could be found in the ex-post data. The selected cases must be originated from the same cluster. Otherwise, the GA would converge to a solution with poor ϕ value in (11). Theoretically, the $\beta^{\mathrm{A}}_{z\longrightarrow z_0,m}$ is calculated as follows

$$\beta^{\mathrm{A}}_{z\longrightarrow z_0,m} = \sum_{g} \left(\sum\nolimits_{k} \pi_{g,z}\beta_{i^*\longrightarrow i_0,k}\right), \forall k \in K_m, z \in \{z_0, z\}, z \neq z_0 \tag{12}$$

where, $\pi_{g,z} \in \Pi_z$ is the share of power generation committed by generator g in zone z to the generation of the next unit of power in the zone; i^* is the index of the bus which is connected with generator g, z_0 is the index of the reference zone, and K_m is the set of indices of all of the constituting transmission lines of interconnector m.

Because of the existence of varying Π_z, various $\beta^{\mathrm{A}}_{z\longrightarrow z_0,m}$ (and thus the Γ^{A}) could exist. Nevertheless, only a finite number of Π_z exists in the actual operation of the transmission network and they tend to repeat themselves over the time. Consequently, the domain of the Γ^{A} is as well finite. A simple heuristics approach has been developed in [13] for selecting cases that belong to a same cluster. It is formulated based on the characteristics of the cases of the cluster as follows

a) Each of the zones in the cases would have an approximately identical value in power generation ratio, w^{P}_z compare to their counterpart in the other cases of the same cluster. For each case, the w^{P}_z is calculated by dividing the net power generation at zone z, P^{A}_z by a weighting factor, W which is the largest absolute value of all of the P^{A}_z in the case. For a case denoted with s, the W^s is calculated as follows

$$W^s = \max(|P^{\mathrm{A},s}_z|);\ \forall z \in \{1, \dots, \mathrm{Z}\} \tag{13}$$

Hence, the $w^{\mathrm{P},s}_z$ is defined as:

$$w^{\mathrm{P},s}_z = \frac{P^{\mathrm{A},s}_z}{W^s} \tag{14}$$

The cases could potentially be part of the same cluster if they meet the following criteria:

$$w^{\mathrm{P},s}_z \approx w^{\mathrm{P},s_{\mathrm{ref}}}_z;\ \forall s \in \{1, \cdots, \mathrm{S}\};\ \forall z \in \{1, \cdots, \mathrm{Z}\} \tag{15}$$

where, s_{ref} is an arbitrarily selected case within the same cluster.

b) The cases that met the criteria of (15) could be part of the same cluster if they have an approximately identical value in their power flow ratio, $w^{\mathrm{F},s}_m$ compared with their counterpart in the other cases of the same cluster. The ratio is calculated by dividing the F^s_m by the W^s in (13):

$$w^{\mathrm{F},s}_m = \frac{F^s_m}{W^s} \tag{16}$$

The cases from the same cluster (possessing the same $\Gamma^{A,*}$) would meet the following criteria:

$$w^{\mathrm{F},s}_m \approx w^{\mathrm{F},s_{\mathrm{ref}}}_m;\ \forall s \in \{1, \dots, \mathrm{S}\};\ \forall z \in \{1, \dots, \mathrm{Z}\} \tag{17}$$

For instance, the w^{F}_m of I1, w^{F}_1 of Figure 1 (b) of two arbitrary cases are formulated as follows. For the first case (case 1):

$$\begin{aligned} w^{\mathrm{F},1}_1 &= \frac{F^1_1}{W^1} \\ &= \frac{P^{\mathrm{A},1}_2 \beta^{\mathrm{A},1}_{2\longrightarrow 1,1} + P^{\mathrm{A},1}_3 \beta^{\mathrm{A},1}_{3\longrightarrow 1,1}}{W^1} \\ &= w^{\mathrm{P},1}_2 \beta^{\mathrm{A},1}_{2\longrightarrow 1,1} + \ w^{\mathrm{P},1}_3 \beta^{\mathrm{A},1}_{3\longrightarrow 1,1} \end{aligned} \tag{18}$$

and, the second case (case 2)

$$\begin{aligned} w^{\mathrm{F},2}_1 &= \frac{F^2_1}{W^2} \\ &= \frac{P^{\mathrm{A},2}_2 \beta^{\mathrm{A},2}_{2\longrightarrow 1,1} + P^{\mathrm{A},2}_3 \beta^{\mathrm{A},2}_{3\longrightarrow 1,1}}{W^1} \\ &= w^{\mathrm{P},2}_2 \beta^{\mathrm{A},2}_{2\longrightarrow 1,1} + \ w^{\mathrm{P},2}_3 \beta^{\mathrm{A},2}_{3\longrightarrow 1,1} \end{aligned} \tag{19}$$

In (18) and (19), if $w_1^{\mathrm{F},1} \approx w_1^{\mathrm{F},2}$ and $[w_2^{\mathrm{P},1}, w_3^{\mathrm{P},1}]^\top \approx [w_2^{\mathrm{P},2}, w_3^{\mathrm{P},2}]^\top$, then $[\beta^{\mathrm{A},1}_{2\longrightarrow 1,1}, \beta^{\mathrm{A},1}_{3\longrightarrow 1,1}]^\top = [\beta^{\mathrm{A},2}_{2\longrightarrow 1,1}, \beta^{\mathrm{A},2}_{3\longrightarrow 1,1}]^\top$. Both case 1 and 2 would have the same $\Gamma^{\mathrm{A},*}$.

This is a strict clustering approach because the cases in each cluster must fulfil (14) and (16). In reality, the $\Gamma^{\mathrm{A},*}$ of several of such clusters are identical. This issue can be handled by testing a new case with $\Gamma^{\mathrm{A},*}$ of identified clusters. The cases belong to the same (identified) cluster if its $\Gamma^{\mathrm{A},*}$ could reproduce the F_m of all of the interconnectors in the new case.

III The implementation

The inverse modeling is implemented in five steps: (i) decomposition is performed by splitting the complete zonal network model into small sub-problems; (ii) the ex-post power flow data is collected for each of the sub-problems, and the $w_z^{\mathrm{P},s}$ and $w_m^{\mathrm{F},s}$ is calculated for each of them; (iii) groups of (ex-post) power flow cases possessing closest possible $w_z^{\mathrm{P},s}$ and $w_m^{\mathrm{F},s}$ values are selected from the ex-post power flow data; (iv) inverse modeling is performed on each of the groups to derive their respective X^* and $\Gamma^{\mathrm{A},*}$ and (v) the $\Gamma^{\mathrm{A},*}$ derived is tested on all of the other (untreated) cases; and the cases which also achieved $\phi \approx 0$ in (5) when computed using the $\Gamma^{\mathrm{A},*}$ is automatically included as part of the group, thus expanding the size of the cluster.

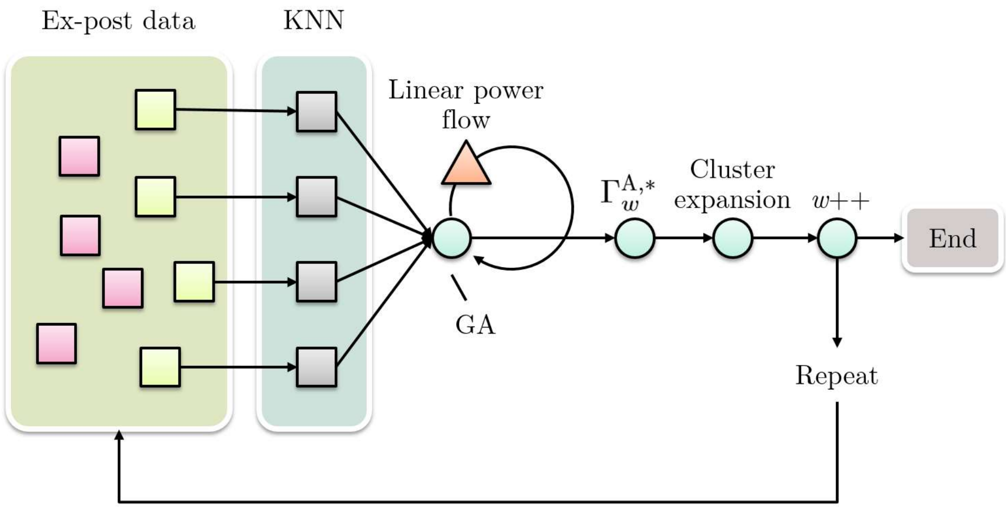


Figure 2. The workflow of the inverse modeling for computing $\Gamma^{\mathrm{A},*}$

The authors applied the k-Nearest Neighborhood (KNN) classification algorithm [14] in step (iii) for selecting the groups of power flow cases. The KNN is an efficient supervised machine learning algorithm for identifying cases of closest proximity in their $w_z^{\mathrm{P},s}$ and $w_m^{\mathrm{F},s}$. The steps are repeated until all of the $\Gamma^{\mathrm{A},*}$ existed in the zonal model are fully identified. The complete work-flow for the inverse modeling is shown in Figure 2.

IV Test case

The authors applied the method to a test case of three market zones in the south of Norway. They are market zone NO1, NO3, and NO5. The three zone network is a sub-problem decomposed from the Nordic zonal model published by Statnett [15]. The hourly data of the first week of 2020 is used as the input. In the week, the areas have a total of 37.3% of all of the hydro-reservoir of the Norwegian system [16], as shown in Table I.

Table I. Hydro-reservoir in the Norwegian markets within the first week of 2020

Market zone	Hydro-reservoir (in GWh)
NO1	4007
NO3	5594
NO5	11182
All market zones	55699
Share (in %)	37.3%

An example of a group of three power flow cases identified by the KNN is given in Table II and III. The tables respectively shows the net power generation and the inter-zonal power flow of the cases. They are the ex-post power flow data on January 2, in hour 11, 13 and 14) as well as their respective $w_z^{\mathrm{P},s}$ and $w_m^{\mathrm{F},s}$. Evidently, the KNN algorithm could find cases with very close $w_z^{\mathrm{P},s}$ and $w_m^{\mathrm{F},s}$ values.

Table II. Net zonal power generation of the KNN identified cases

Zone	$P_z^{\mathrm{A},1}$	$P_z^{\mathrm{A},2}$	$P_z^{\mathrm{A},3}$	$w_z^{\mathrm{P},1}$	$w_z^{\mathrm{P},2}$	$w_z^{\mathrm{P},3}$
NO1	-1304.51	-1316.38	-1374.84	-1.00	-1.00	-1.00
NO3	601.54	609.30	648.04	0.46	0.46	0.47
NO5	702.97	707.08	726.80	0.54	0.54	0.53

Table III. Inter-zonal power flow of the cases in Table II

Interconnector	Start	End	F_m^1	F_m^2	F_m^3	$w_m^{\mathrm{F},1}$	$w_m^{\mathrm{F},2}$	$w_m^{\mathrm{F},3}$
I1	NO1	NO3	-312.36	-314.24	-328.71	-0.24	-0.24	-0.24
I2	NO1	NO5	-992.15	-1002.14	-1046.13	-0.76	-0.76	-0.76
I3	NO3	NO5	289.18	295.06	319.33	0.22	0.22	0.23

The X^* found by GA is shown in Table IV. The $\Gamma^{\mathrm{A},*}$ calculated with X^* is shown in Table V. The F_m, F_m^*, and ϕ of three of the cases calculated using the X^* is given in Table VI. The solution is tested on the other cases and found 14 additional cases that belong to the same cluster. All cases in the same cluster and their respective maximum error in approximation, $\gamma = \max(\frac{|F_m^* - F_m|}{F_m} \times 100))$ are listed in Table VII. The entire domain of $\Gamma^{\mathrm{A},*}$ could be fully identified by repeating the steps.

As shown, the inverse modeling had identified a cluster in the ex-post power flow data of the first week of 2020. The cluster comprises 17 cases with highest γ of 8.03%. All of them share the same $\Gamma^{\mathrm{A},*}$ that is given in Table V. The inverse modeling is repeated until all of the clusters are fully identified. It is evident that the $\Gamma^{\mathrm{A},*}$ has repeated itself within the zonal model as it appears in several consecutive hours of two occasions (from hour 9 to 21 in January 2, and from hour 9 to 11 in January 3). Because of the validity of the decomposition, the X^* found in this sub-problem remain valid as they could be recombined with the X^* of the other sub-problem to form the complete zonal model of the Nordic transmission system.

Table IV. The X^* found by GA for the KNN identified cases

Interconnector	X^*
I1	1.0000
I2	0.3099
I3	0.0139

Table V. The $\Gamma^{\mathrm{A},*}$ for the three bus model computed using X^* (in Table IV)

Zone	Interconnector	$\beta^{\mathrm{A}}_{z \longrightarrow z_0,m}$
NO1	I1	-
NO1	I2	-
NO1	I3	-
NO3	I1	-0.2446
NO3	I2	-0.7553
NO3	I3	0.7553
NO5	I1	-0.2341
NO5	I2	-0.7658
NO5	I3	-0.2341

Table VI. The F_m, F^*_m, and ϕ of the cases identified by KNN

Case	Interconnector	F^*_m (MW)	F_m (MW)	$\lvert F^*_m - F_m \rvert$ (MW)
1	I1	-311.76	-312.36	0.60
1	I2	-992.76	-992.15	0.61
1	I3	289.79	289.18	0.61
2	I1	-314.24	-314.62	0.38
2	I2	-1002.14	-1001.77	0.37
2	I3	295.06	294.68	0.38
3	I1	-328.71	-328.71	0
3	I2	-1046.13	-1046.13	0
3	I3	319.33	319.33	0
			ϕ	2.95

Table VII. All other cases in the same cluster of the ones shown in Table II and III

Case	Date	Hour	γ (%)
1	2020-01-02	9	5.16

2	2020-01-02	10	2.15
3	2020-01-02	11	0
4	2020-01-02	12	2.12
5	2020-01-02	13	0.21
6	2020-01-02	14	0.13
7	2020-01-02	15	6.82
8	2020-01-02	16	2.64
9	2020-01-02	17	8.03
10	2020-01-02	18	0.04
11	2020-01-02	19	0.06
12	2020-01-02	20	0.57
13	2020-01-02	21	4.16
14	2020-01-03	9	1.88
15	2020-01-03	10	2.59
16	2020-01-03	11	4.57
17	2020-01-04	20	6.03

V Discussions and Conclusions

The use of the zonal model with correct representation of the inter-zonal power flow is beneficial to the long-term generation dispatch planning. A correct zonal model would reflect how the local planning of bulk power dispatch would affect the inter-zonal bottleneck in any part of the transmission network. This is very important for minimizing the corrections (and the associated opportunity costs) needed at the time of dispatch. The zonal model is presently the most important component in the FBMC design in Europe. Additionally, the policy makers in Europe prefer to adopt the zonal model for institutional barriers and socio-political reasons [3] [17].

The authors have pointed out that there are two approaches to conduct zonal modeling: (i) the model-based (i.e., bottom-up), and (ii) the data-driven (i.e., top-down) approach. The current state-of-the-arts are mostly model-based approach. Meanwhile, the data-driven approach is relatively unexplored. The usability of the model-based approach is limited for those who do not have the knowledge on the transmission network (especially the impedance of the transmission lines). Also, the direct dependency of the Γ^{A} with the Π_z could render it impractical.

Then, the authors propose a data-driven approach as a better alternative for zonal modeling. The method takes the ex-post power flow data as input, and reconstruct all Γ^{A} (the most important parameter of the zonal model) by means of inverse modeling. It is a highly tractable and feasible method as it inherits two significant characteristics of the linear power flow theory: (i) the ex-post power flow data can be decomposed into small sub-problems; where each of them can be solved independently. The sub-problems could be recombined with the others to reconstruct the complete zonal model; and, (ii) the ex-post power flow data can be grouped into several clusters, where each of the clusters comprises a number of cases that have an approximately identical $\Gamma^{\mathrm{A},*}$. The clustering can be done because the $\Gamma^{\mathrm{A},*}$ tend to repeat themselves. Also, the $\Gamma^{\mathrm{A},*}$ is not sensitive to minor changes in Π_z (see [13] for more information). The use of the linear power flow theory for representing the inter-zonal power flow is reasonable as it deviates from a complete AC power flow analysis by about 5% on average for high-voltage transmissions [3].

The authors have contributed to the field of study by introducing an approach to effectively integrate the power flow characteristics within the zonal model for use in the long-term power dispatch planning. Many long-term simulation tools (e.g., the EMPS model) computes the inter-zonal power exchanges based on the network flow programming (i.e., the available transfer capacity) approach [18]. Although their solutions could be insightful, they are imprecise in reality because the modeling neglected many electrical characteristics [4]. Presently, there is also a need for a formal methodology to determine the future $\Gamma^{A,*}$ prediction of zonal prices beyond the day-ahead time-frame [4]. The inverse modeling can effectively identify all $\Gamma^{A,*}$ which had physically existed in the past. The trend and probability distribution of the identified $\Gamma^{A,*}$ at a certain time in the past (e.g., the seasonal factor) could serve as a valid reference for long-term future price determination.

VI Acknowledgements

This work would not be possible without the Nordic inter-zonal power flow data published by Statnett on its website

References

[1] M. G. Pollitt, The European single market in electricity: an economic assessment, Review of Industrial Organization, vol. 55, pp. 63-87, 2019

[2] B. Felton, T. Felling, and C. Weber, Inefficiencies in zonal market coupling due to uncertainties in generation shift keys, 40th IAEE International Conference, Singapore, June 19, 2017

[3] K. Van den Bergh, and E. Delarue, An improved method to calculate injection shift keys, Electric Power Systems Research, vol. 134, pp. 197-204, 2016

[4] T. Kristiansen, The flow based market coupling arrangement in Europe: implications for traders, Energy Strategy Reviews, vol. 27, pp. 1-7, 2020

[5] P. C. See, O. B. Fosso, K. Y. Wong, and M. Molinas, An approach for establishing a common grid model for flow-based market mechanism simulations, CSEE Journal of Power and Energy Systems, vol. 5, no. 3, pp. 374-381, 2019

[6] T. M. Hansen, A. G. Journel, A. Tarantola, and K. Mosegaard, Linear inverse Gaussian theory and geostatics, Geophysics, vol. 71, no. 6, pp. 101-111, 2006

[7] S. Delsanto, M. Griffa, and L. Morra, Inverse problems and genetic algorithms, In: P. P. Delsanto, Universality of Nonclassical Nonlinearity, Springer, United States of America, 2006

[8] J. A. Scales, and R. Sneider, The anatomy of inverse problems, Geophysics, vol. 65, no. 6, pp. 1708-1710, 2000

[9] B. Stott, J. Jardim, and O. Alsac, DC power flow revisited, IEEE Transactions on Power Systems, vol. 24, no. 3, pp. 1290-1300, 2009

[10] D. E. Goldberg, Genetic algorithms in search, optimization and machine learning, Addison-Wesley, Massachusetts., United States of America, 1989

[11] W. F. Ames, Nonlinear ordinary differential equations in transport processes, Academic Press, New York, United States of America, 1968

[12] Y. Levron, and Y. S. Barel, Close multiple power flow solutions in power networks, IEEE International Conference on Microwave, Communications, Antennas and Electronic Systems, Tel Aviv, Israel, November 2-4, 2015

[13] P. C. See, O. B. Fosso, and M. Molinas, Inverse modeling of zonal transmission network, submitted to Frontiers in Energy Research, 2020

[14] T. Cover, and P. Hart, Nearest neighbor pattern classification, IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21-27, 1967

[15] https://www.statnett.no/en/for-stakeholders-in-the-power-industry/data-from-the-power-system

[16] https://www.nordpoolgroup.com/Market-data1/Power-system-data/hydro-reservoir1/ALL/Hourly/?view=table

[17] I. Aravena, and A. Papavasiliou, Renewable energy integration in zonal markets, IEEE Transactions on Power Systems, vol. 32, no. 2, pp. 1334-1349, 2017

[18] H. K. Ringkjøb, P. M. Haugan, and I. M. Solbrekke, A review of modelling tools for energy and electricity systems with large share of variable renewables, Renewable and Sustainable Energy Review, vol. 96, pp. 440-459, 2018