Model Reduction of Swing Equations with Physics Informed PDE

Laurent Pagnier, Michael Chertkov
Program in Applied Mathematics,
University of Arizona,
Tucson, USA
{laurentpagnier,chertkov}
@math.arizona.edu

Julian Fritzsch, Philippe Jacquod
School of Engineering,
University of Applied Sciences
of Western Switzerland HES-SO,
CH-1951 Sion, Switzerland
{julian.fritzsch,philippe.jacquod}@hevs.ch

***Abstract*—This manuscript is the first step towards building a robust and efficient model reduction methodology to capture transient dynamics in a transmission level electric power system. Such dynamics is normally modeled on seconds-to-tens-of-seconds time scales by the so-called swing equations, which are ordinary differential equations defined on a spatially discrete model of the power grid. We suggest, following Seymlyen (1974) and Thorpe, Seyler and Phadke (1999), to map the swing equations onto a linear, inhomogeneous Partial Differential Equation (PDE) of parabolic type in two space and one time dimensions with time-independent coefficients and properly defined boundary conditions. The continuous two-dimensional spatial domain is defined by a geographical map of the area served by the power grid, and associated with the PDE coefficients derived from smoothed graph-Laplacian of susceptances, machine inertia and damping. Inhomogeneous source terms represent spatially distributed injection/consumption of power. We illustrate our method on PanTaGruEl (Pan-European Transmission Grid and ELectricity generation model) [1], [2]. We show that, when properly coarse-grained, i.e. with the PDE coefficients and source terms extracted from a spatial convolution procedure of the respective discrete coefficients in the swing equations, the resulting PDE reproduces faithfully and efficiently the original swing dynamics. We finally discuss future extensions of this work, where the presented PDE-based reduced modeling will initialize a physics-informed machine learning approach for real-time modeling, $n-1$ feasibility assessment and transient stability analysis of power systems.**

***Index Terms*—Power system dynamics, disturbance propagation, electromechanical waves, inter-area oscillations, continuum, physics-informed machine learning**

I. Introduction

This manuscript is focused on building a reduced – thus computationally efficient – and sufficiently accurate model describing the transient response of a transmission level electric power system to a significant perturbation – for example the disconnection and/or re-connection of a large generator. We consider transmission level power system dynamics on a continental scale and focus on sub-minute transients on time scales ranging from one second to few tens of seconds. We follow an approach that is standard in power system studies [3], [4], [5] and assume that the so-called swing equations, giving the time-evolution of the voltage angles at all nodes on the power grid, provide a sufficiently accurate representation of the power system dynamics within the considered spatio-temporal scales. Stated differently in the language of modern machine learning, the spatio-temporal integration of the swing equations provide a high-fidelity representation of the ground truth. There are two competing aspects of the swing equations. On the one hand, they are based on physically meaningful quantities and parameters such as line capacities, machine inertia and damping. Accordingly they are expected to correctly capture the physics of the system. On the other hand, integrating these equations on a large, continental scale grid can be computationally very expensive, even for a single run corresponding to a specific initial condition. Obviously, it becomes even more expensive if the task is to screen many possible initial conditions, and often prohibitively expensive when the screening need to be repeated numerous times, testing many possible control actions. Model reduction in this type of real-time, so-called "on-line" applications [6] comes as a way to strike a balance between accuracy and computational complexity. Central to this optimization is to keep in mind that the transient dynamics of interest occurs over time scales up to few tens of seconds while the goal is numerically resolve all the multiple scenarios (of initial conditions and controls) faster than real time.

How does model reduction work? In the current era of deep learning, many model reduction techniques rely on neural networks and other tools of modern data science and machine learning, see e.g. [7], [8], [9], [10]. The idea is to use the ground truth model – the swing equations in our case – to produce dynamical data, and then to train a pre-selected reduced model on these datasets to fine-tune the parameters of the model. If the reduced model is of an application agnostic type, as is customary in mainstream machine learning, the scheme relies on very large datasets. However, recall that running our ground truth model is computationally expensive. Then, if producing the needed training datasets is not an option, can we still hope to build a reliable reduced model? Our only hope in this case is to inject the relevant, application-specific information – in our case information about the power system physics – into the model reduction framework. Physics Informed Machine Learning (PIML) is the modern approach to resolve the model reduction bottleneck – that is to compensate

for the lack of data (typical of the on-line applications) by building models that are aware of the underlying physics [11], which may be expressed in terms of differential equations [12], [13], [14]. (See also [15], [16] for discussion of the application of PIML to power systems.)

Why is Partial Differential Equation (PDE) modeling a sound option for power system model reduction? In this manuscript, we propose a first step towards developing PIML for general on-line applications and specifically advancing model reduction of PIML applications to power systems, as e.g. developed earlier by members of our team [16], [17]. Similarly to [15], we take advantage of the PIML approach and construct an on-line framework for simulating power system dynamics faster than real time. We are however aiming to capture the transient dynamics in a very large, continental-scale power system, a goal that has not been addressed by any earlier related approach we are aware of. Accordingly, we choose to build our reduced model on the continuous PDE approach to modeling power system dynamics pioneered by Semlyen [18] and later extended by Thorpe, Seyler and Phadke [19] (see also [20]). These works were however restricted to spatially-continuous one dimensional system, i.e. with 1+1 dimensional PDE. Our PDE approach to power system, to be presented below, inherits all the relevant physics of the original swing Ordinary Differential Equations (ODEs), accordingly it is 2+1 dimensional. Thus it resolves power grid dynamics over a spatially-continuous two-dimensional domain associated with the power system's geographical area of service. Approximating the swing ODEs by a PDE may seem strange at first sight, as naively, this transition dramatically increases the number of degrees of freedom. However, this naive thinking is not quite right for a number of reasons. First, numerical solutions of linear 2+1 dimensional PDE assumes spatial regularization via two-dimensional grid, where the grid size can be chosen according to the desired spatial resolution. Therefore, the number of grid points may eventually be comparable of even smaller than the number of nodes in the original grid. Second, basic operations, such as matrix inversion, can be numerically implemented much more efficiently on a regular grid than on a complex meshed graph. Third, the number of physical parameters in the original power grid model (line capacities, machine inertia and damping coefficients) may be reduced significantly within the PDE approach. Indeed, within the reduced model, we want to faithfully capture only the large-wavelength component of the swing dynamics. This justifies using only a coarse-grained/filtered expression for all the coefficients in the linear PDE, therefore representing the coefficients via only a few long-wavelength harmonics.

Our Contribution: In this manuscript we make the first steps towards a novel on-line methodology for multi-scenario testing and control based on modeling the dynamics of a large, continental scale power system within a novel 2+1 PDE modeling framework. We show how a properly coarse-grained PDE model faithfully captures the power grid transient responses to disturbances of a high fidelity model. Our model reduction methodology is illustrated on the PanTaGruEl model of the synchronous grid of continental Europe introduced in [1], [2]. Specifically, PanTaGruEl simulates power flow and swing equations with high fidelity to produce the ground truth data. The latter in their turn are used to infer a spatially continuous 2+1 dimensional PDE model. The quality of the reconstruction is judged, first, by its ability to mimic power system dynamics and, second, by a faithful reconstruction of spatially coarse-grained and physically meaningful static – spatial distribution of line impedances – and dynamic – spatial distribution of damping and inertia – parameters. We conclude the manuscript with a suggestion for a path towards using the PDE based reduced modeling framework for efficient online screening of multiple failure scenarios on large transmission grids, faster than real time.

II. Problem Formulation

A. Power Flow and Swing Equations (system of ODEs)

AC Power Flow (PF) equations describe steady distributions of electric power flows over an AC power grid. The equations connect complex power injections $\{s_i \equiv p_i + \boldsymbol{i} q_i\}$ to complex voltages $\{V_i \equiv v_i \exp(\boldsymbol{i}\theta_i)\}$, where p_i, q_i, v_i and θ_i denote the active and reactive power injections, and the voltage magnitude and angle at node $i \in \mathcal{V}$ respectively:

$$p_i = \sum_j v_i v_j \Big[g_{ij} \cos\big(\theta_i - \theta_j\big) + b_{ij} \sin\big(\theta_i - \theta_j\big) \Big], \quad \text{(1a)}$$

$$q_i = \sum_j v_i v_j \Big[g_{ij} \sin\big(\theta_i - \theta_j\big) - b_{ij} \cos\big(\theta_i - \theta_j\big) \Big]. \quad \text{(1b)}$$

Here, b_{ij} and g_{ij} are elements of the susceptance and conductance matrices, see e.g. [3], [4], [5] for more details.

Suppose that a steady solution of the PF Eqs. (1) is perturbed, for example by a fast disconnection and reconnection of a large generator or a load. In this particular case, stabilization of voltage amplitudes to their pre-perturbation values occurs within a few AC cycles, on a time scale of tens of milliseconds. However, voltage angle relaxation takes longer, and the transient dynamics of these angles on time scale ranging up to few tens of seconds is governed by the swing equations (see e.g. [4]). In their linearized form they read

$$m_i \ddot{\theta}_i + d_i \dot{\theta}_i = p_i - \sum_j v_i v_j b_{ij} (\theta_i - \theta_j)\,, \quad (2)$$

where the voltage amplitudes v_i and v_j are considered constant, already stabilized to the steady-state solution of Eqs. (1) and m_i and d_i denote the inertia and the damping (i.e. primary control) of the generators. Eq. (2) describes the relaxation dynamics of voltage angles towards a steady-state solution, $\ddot{\theta}_i = \dot{\theta}_i = 0$, corresponding to the lossless, $g_{ij} = 0$, linearized version of Eqs. (1a). Two comments are in order here. First, the linearized approach used here is in practice quite accurate to reproduce the transient dynamics following not too strong perturbations. Nevertheless, the approach to be presented below can be extended to the nonlinear case, with $(\theta_i - \theta_j) \to \sin(\theta_i - \theta_j)$ in Eq. (2). Second, the swing equation approach is not restricted to the just discussed case of a fast

disconnection-reconnection fault, but also captures the voltage angle dynamics following a fault which is not immediately cleared, such as the removal of a generator or a load without reconnection. In such cases, the final, relaxed state is not balanced, i.e. $\sum_i p_i \neq 0$, and the power mismatch is compensated by the second, damping term in Eq. (2), leading to an AC frequency shift $\dot{\theta}_i = \omega_{pf} \forall i$, with $\sum_i p_i = \omega_{pf} \sum_i d_i$.

In the next paragraph, we construct a reduced model by mapping the discrete system of ODEs (2) into a continuous PDE. Before we do that, we re-emphasize why a model reduction is needed at all. The motivation was lucidly expressed in Ref. [21] as follows: "The focus is on the construction of low-order models which closely approximate the global behavior of the hybrid nonlinear system. There is a growing recognition of the strong need for rapid and reliable computation of the system dynamics." Comprehensive discussions of model reduction in a general context as well as for specific applications to slow coherency and inter-area oscillations can further be found in Ref. [22].

B. Mapping the power network to a two-dimensional continuum: the Swing PDE Model

Consider a two-dimensional domain $\boldsymbol{\Omega} \subset \mathbb{R}^2$, with coordinates $\boldsymbol{r} = (x, y)$, inside which the discrete, planar or quasi-planar network is embedded. The boundary of the domain is denoted by $\boldsymbol{\partial\Omega}$ and $\forall \boldsymbol{r} \in \boldsymbol{\partial\Omega}$, $\boldsymbol{n} \equiv (n_x, n_y)$ denotes the normal vector to the boundary at $\boldsymbol{r}$. Imagine that the swing Eqs. (2) are derived by discretizing a PDE describing the dynamics of a scalar field $\theta(t; \boldsymbol{r})$ over an irregular mesh which corresponds to the original network. Then, following [19], one naturally asks: what is the PDE corresponding to the swing Eqs. (2)? We answer this question by writing the following most general form of the swing PDE on $\boldsymbol{\Omega}$:

$$m(\boldsymbol{r})\frac{\partial^2}{\partial t^2}\theta(t;\boldsymbol{r}) + d(\boldsymbol{r})\frac{\partial}{\partial t}\theta(t;\boldsymbol{r}) = p(t;\boldsymbol{r}) + \sum_{\alpha,\beta=1,2} \partial_{\boldsymbol{r}_\alpha} b_{\alpha\beta}(\boldsymbol{r}) \partial_{\boldsymbol{r}_\beta}\theta(t;\boldsymbol{r}), \quad (3)$$

where $\boldsymbol{r}_1 = x,\ \boldsymbol{r}_2 = y$. One of our main task is to map the physical parameters of (2) into the continuum as follows

$$\forall i: \quad \theta_i(t) \to \theta(t;\boldsymbol{r}),\ m_i \to m(\boldsymbol{r}),\ d_i \to d(\boldsymbol{r}) \\ p_i(t) \to p(t;\boldsymbol{r}),\ b_{ij} \to b_{\alpha\beta}(\boldsymbol{r}),\ \forall \alpha,\beta = 1,2. \quad (4)$$

We discuss a procedure for reconstruction of the parameters in Section III.

Next, the swing PDE (3) need to be equipped with physically appropriate boundary conditions, which in our case are Neumann boundary conditions

$$\forall t,\ \forall r \in \partial\Omega: \ n_\alpha b^{\alpha\beta} \partial_{\boldsymbol{r}_\beta}\theta(t;\boldsymbol{r}) = 0, \quad (5)$$

corresponding to a vanishing normal derivative of the angle field on the domain boundary $\partial\Omega$. These boundary conditions directly follow from the condition that post-perturbation frequencies in the continuous model correspond to those in the original swing equations, i.e. $\omega(t;\boldsymbol{r}) \equiv \frac{\partial}{\partial t}\theta(t;\boldsymbol{r}) = \omega_{\rm pf}$, $\forall \boldsymbol{r} \in \Omega$. This condition translates into

$$D\,\omega_{\rm pf} = \int_{\boldsymbol{\Omega}} p(t;\boldsymbol{r}) \mathrm{d}\boldsymbol{r} + \int_{\boldsymbol{\partial\Omega}} b_{\alpha\beta}(\boldsymbol{r}) \partial_{\boldsymbol{r}_\alpha}\theta(t;\boldsymbol{r}) n_\beta(\boldsymbol{r}) dr, \quad (6)$$

which directly corresponds to the frequency shift $D\omega_{\rm pf} = \sum_i p_i$ in the swing Eqs. (2) if the second term in the right-hand side of Eq. (6) vanishes identically. This is guaranteed by the Neumann boundary conditions (5).

In the following we simplify our PDE model, assuming that the b-tensor is diagonal $b_{12} = b_{21} = 0$, accordingly, we use a shorter notation, $b_{11} \to b_x,\ b_{22} \to b_y$. Additional details on the derivation of the swing PDE Eqs. (3) can be found in the Appendix.

C. Parameter Reconstruction and Validation Experiments

Once the general structure of the swing PDE (3) is established we need to reconstruct the PDE parameters $b_{x,y}(\boldsymbol{r}), m(\boldsymbol{r})$ and $d(\boldsymbol{r})$. This is achieved in the next Section, and proceeds along a smoothing procedure of the original parameters defined on the discrete grid of Eq. (2). Obviously, there are many different choices for this coarse-graining/filtering procedure. Therefore, it is crucial to develop a validation criteria. Our validation criterion proceeds via post-factum dynamic tests, described in the following section, where we compare the dynamics following a perturbation in the original, discrete swing equations with that in the spatially continuous model.

III. Reconstruction of the PDE Parameters via Artificial Diffusion

The PDE model (3) is a continuous version of the swing equations (2) defined on a discrete power network. Accordingly, parameter reconstruction consists in developing a smoothing procedure of the spatially discrete physical parameters defined.

Such a simple smoothing procedure was proposed in Ref. [18], [19] focusing on a 1+1 dimensional PDE representation of a linear power network, where all parameters in the 1+1 dimensional PDE were chosen to be spatially constant. This homogeneous smoothing was improved in [20], where non-uniform parameters of the 1+1 PDE system were derived by means of a convolution with a fixed Gaussian kernel.

Our Artificial Diffusion (AD) approach generalizes the Gaussian kernel approach of [20] by (a) extending from the 1+1 PDE case to the 2+1 PDE case, and (b) making it flexible in terms of adjusting spatial scaling (width) of the Gaussian kernel. Implementation of the AD approach starts with initializing the spatial distribution of one of the physical coefficients [$x(\boldsymbol{r}) = m(\boldsymbol{r}), d(\boldsymbol{r})$ or $b_{\alpha\beta}(\boldsymbol{r})$], assigning discrete values from the swing model to the nearest nodes of the regular square lattice discretizing the PDE (3). (See the Appendix for a discussion of the discretization procedure.) Next we introduce

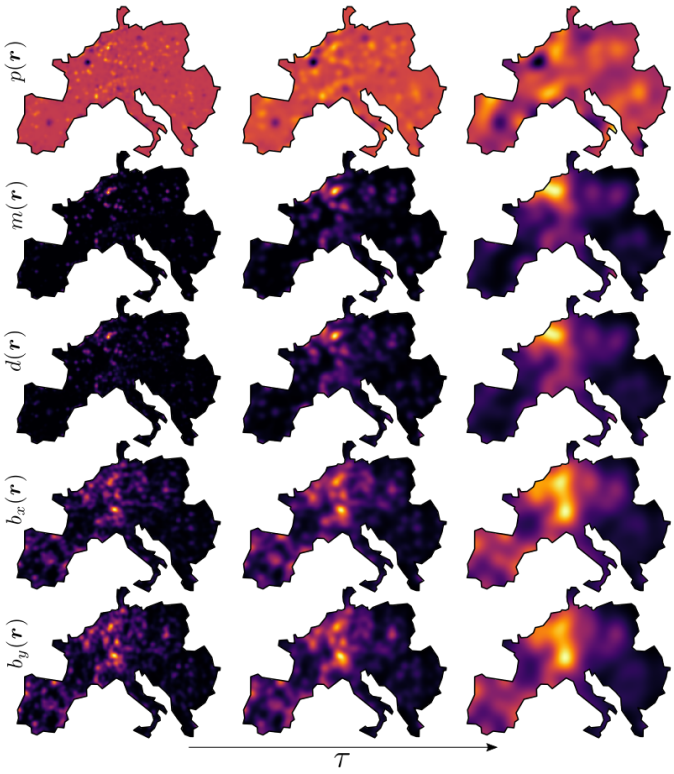

Fig. 1. Evolution in the artificial time τ of the system parameters for the continuous PDE model from the PanTaGruEl grid. (See text for details.)

an AD process which propagates these initial values (in a fictitious algorithmic time τ) over the lattice according to

$$\frac{\partial}{\partial \tau} x = \kappa \boldsymbol{\nabla}^2 x \,. \tag{7}$$

Here, x is one of the physical system parameters, while κ and τ are the artificial (purely algorithmic, non-physical) time and diffusion coefficient characterizing the smoothing procedure. They are not independent of one another and can be absorbed into the rescaling of the other. The diffusion process governed by Eq. (7) is stopped after some time, τ_*, which controls the width of the convolution kernel, and which is determined by an *a posteriori* validation criterion.

We illustrate the artificial diffusion procedure on PanTaGruEl grid. Fig. 1 shows the artificial dynamics of the diagonal components of the susceptance tensor, $b_x(\boldsymbol{r})$ and $b_y(\boldsymbol{r})$, the damping and inertia coefficients and the power injections. These numerical results suggest in particular that the resulting spatial distribution of the diagonal part of the susceptance tensor is isotropic, i.e. $b_x(\boldsymbol{r}) \approx b_y(\boldsymbol{r})$. This, together with the neglect of off-diagonal elements for the susceptance tensor, means that the entire susceptance tensor is well approximated by a scalar function, $b_{\alpha\beta}(\boldsymbol{r}) \approx \delta_{\alpha\beta} b(\boldsymbol{r})$, with the Kronecker symbol $\delta_{\alpha\beta}$.

We validate the AD procedure on the PanTaGruEl grid as follows. We find the grid points closest to the location of the buses within PanTaGruEl and terminate the AD process when the voltage angles of the two steady-state solutions are as close

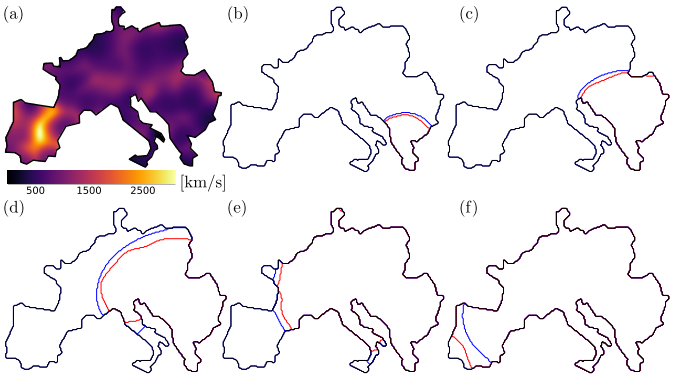

Fig. 2. (a) assessment of the local propagation speed by $c(\boldsymbol{r}) = \sqrt{b(\boldsymbol{r})/m(\boldsymbol{r})}$. (b)-(f) Fronts of the perturbation at incremental times, with $\Delta t = 0.6$s, after the fault for inhomogeneous parameters (red) and for average parameters (blue).

to one another as possible, $\theta_i^{\rm cont} \simeq \theta_i^{\rm disc}$. (See Section IV) for details.) Notice that, not surprisingly, the best fit is achieved at setting sufficiently small τ_*, corresponding to the Gaussian kernel size corresponding roughly to a few lattice spacing in the grid discretization of the PDE (3).

We conclude this Section, clarifying, on one example, how the calibration of the parameters in the PDE (3) can be used to gain intuitive insight into the system behavior. Once the filtered values of the susceptance tensor, $\boldsymbol{b}(\mathbf{r})$ and of the inertia parameter, $\boldsymbol{m}(\boldsymbol{r})$ are discovered (via the aforementioned validation procedure), we can immediately use it to build a spatial map for the speed of the electro-mechanical waves, $c(\boldsymbol{r}) = \sqrt{b(\boldsymbol{r})/m(\boldsymbol{r})}$, shown in Fig. (2). The Figure also illustrates how knowledge of $c(\boldsymbol{r})$ allows to reconstruct dynamics of a localized perturbation spread. (The illustration is based on the example of the perturbation front spread within the PanTaGruEl grid initiated in Greece. See Fig. 4, and related discussions in the following Section, for details.)

IV. Experiments

In this Section, we test our PDE model and AD procedure (direct reconstruction of the PDE model parameters) against the original swing model (the ground truth).

A. Steady State Experiments

We start with comparative analysis of the PDE model vs original swing model performances in reproducing the steady (time independent) solution. The two are juxtaposed in Fig. 3. We report a good agreement, even though observing some discrepancies, in particular in the Italian and Greek part of the PanTaGruEl. We can attribute some of the discrepancies to transformers, which are present in the original model (PanTaGruel) but were not represented proper in the continuous model. We plan to correct for this caveat in the future. (See discussion in the following Section.)

B. Dynamic Experiments

Next we describe the ultimate (as dynamic) comparison of the reduced (PDE) model and the basic (swing) model

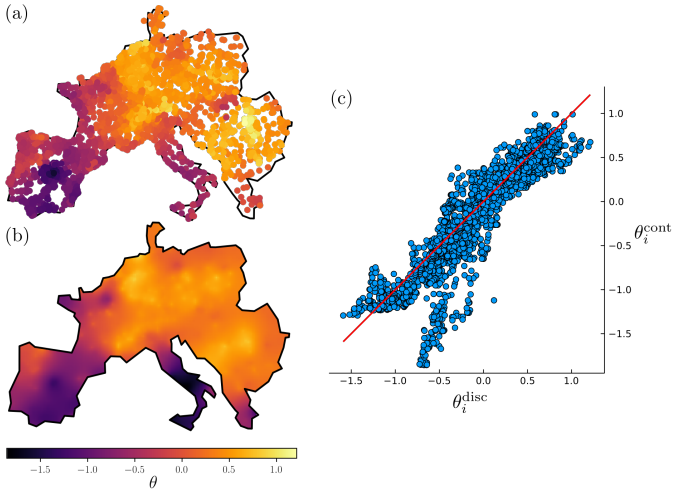

Fig. 3. Comparison of discrete and continuous steady state solutions. (a) Discrete solution θ^{disc}. (b) Continuous solution $\theta(\boldsymbol{r})$. The same color map is used in panels (a) and (b). (c) One-to-one comparison: for each bus in the discrete model the nearest node in the continuous mesh is selected. Red line marks a perfect match.

in describing exemplary response of the PanTaGruEl system to a loss of 900 MW of production in Greece. Fig 4 shows frequency response at four generators across Europe. We register a good agreement. Notice that the largest deviations are observed close to location of the fault.

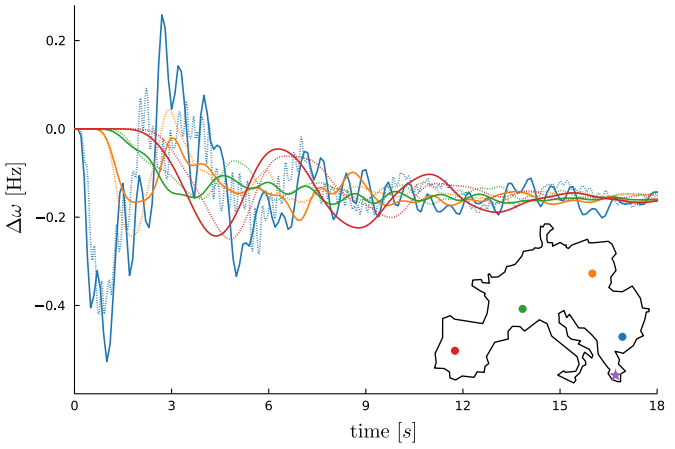

Fig. 4. Frequency response of the system to a 900 MW loss of power in Greece (marked by a purple star) at the locations shown in the inset. Data from the continuous (PDE) and discrete (system of original ODEs) models are displayed as dotted and solid lines respectively.

C. Experiments towards Future Machine Learning Explorations

This Subsection is devoted to experiments aimed at provoking future work where we plan to improve the AD procedure of the PDE model parameter calibration with machine learning (following our general logic described in the introductory Section I, see also the next concluding Section V) with the (Physics Informed) Machine Learning.

Moving forward we are interested to see if the number of parameters describing the system's physical quantities can be drastically reduced, i.e. filtered out, without introducing a significant loss of accuracy into description of the system spatio-temporal dynamics at sufficiently large scale (say with the resolution of above 50km). Specifically, we will be aiming to improve the AD procedure finding optimal parameters in the PDE (3) parameterized with Neural Networks. (The AD procedure, per se, is still expected to be useful to initialize the future PIML schemes.) Planning for this work we should take into account a number of considerations, such as related to keeping number of training epoch and overall the training time under control – consistently with the goal of making the future Machine Learning schemes capable of achieving the goal of evaluating in parallel multiple perturbation scenarios in the time which is comparable or faster than dynamic simulations of the ground truth (swing) model. This NN-based approaches will be physics informed in multiple ways, in particular via (a) diminishes the risk of over-fitting (typical in deep learning); and (b) using the set of eigen-modes of the continuous model (e.g. with the parameters found via the AD procedure), possibly filtered additionally with a low-pass Fourier filter, as a set of features for the NN representation.

To elaborate on the last point, related to additional Fourier filtering of the results of the AD procedure, we conducted the following preliminary experiments illustrated in Fig. 5. We apply a low pass filter with a cut-off frequency of 20% of the largest frequency. We compare the stable solution that is found before and after Fourier filtering. We find that this additional Fourier filtering returns results which are, generally, in a good qualitative agreement with the ground truth. (Deviations between the two are mainly found in the peripheral areas such as the Balkans.)

V. Critical Evaluations, Discussions and Future Work

We start this concluding Section reflecting on what we managed to achieve, transitioning to discussion of some preliminary, but not yet fully confirmed observations, and finally turning the next steps planned to advance the ambitious program set up in the introductory Section I.

Major accomplishment of the manuscript is in construction and validation of the PDE version of the swing equations. The construction included accurate resolution of the boundary conditions and development of an efficient and flexible filtering procedure – Gaussian convolution achieved via "artificial diffusion". We also saw how the PDE model can be used to analyze perturbations, also showing results in a transparent and intuitive manner.

We have also made a number of interesting observations which are clearly preliminary. Comparing the performance of the continuous model on a regular grid on the size comparable to the size of the original graph (the size is measured in the number of nodes) we saw that the PDE model performance an order of magnitude faster. This is consistent with what is expected (see respective discussion in the introductory Section

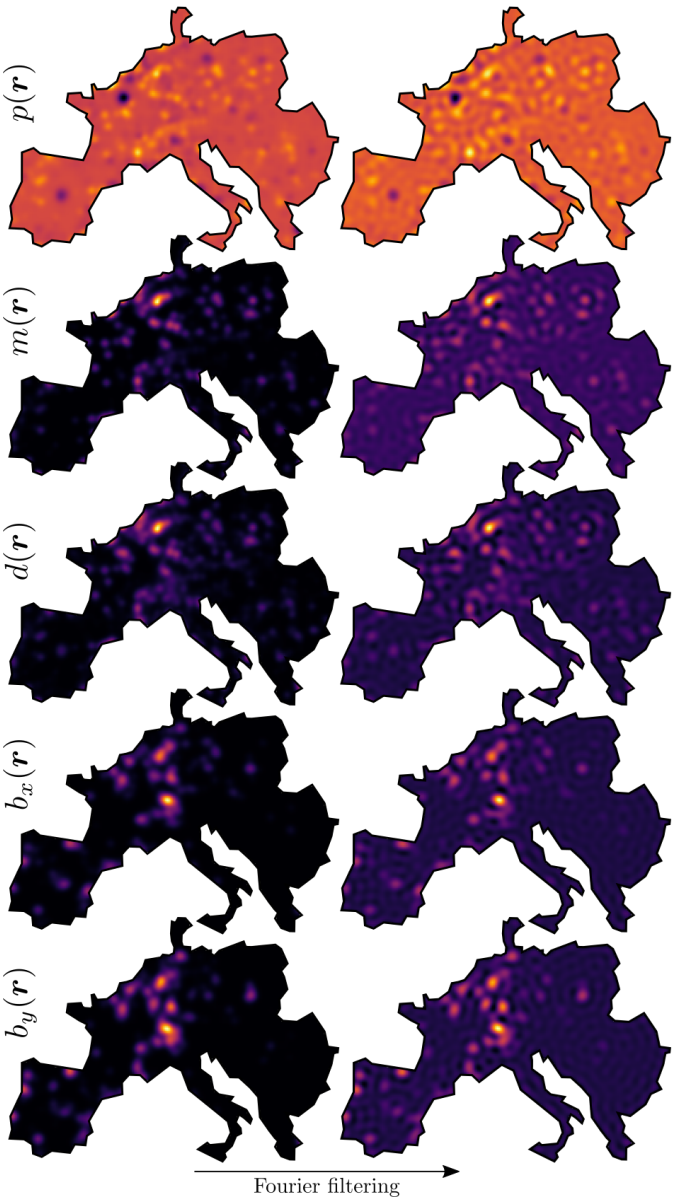

Fig. 5. Distribution of the grid parameters before and after a Fourier low pass filter with a cut-off frequency of 20% (of the maximum frequency) applied.

I). We also observed that in many cases a significant filtering, reviled via choice of a rather large value of the coarsening scale (large value of τ_* within the "artificial diffusion" process) does not hurt accurate resolution of the underlying phenomena at sufficiently large scales (hundreds of kilometers) and sufficiently long times (seconds).

Finally, the most important message this manuscript sends is related to the path forward towards using the methodology developed here to empower further work on the PDE-based model reduction with the PIML approach. Specifically, our next step consists in using the approach developed here as a warm start for learning physical parameters of the PDE model (3). (See also discussions of the preceding Section.) Indeed, we envision modeling the functional maps for, $m(\boldsymbol{r}), d(\boldsymbol{r})$ and

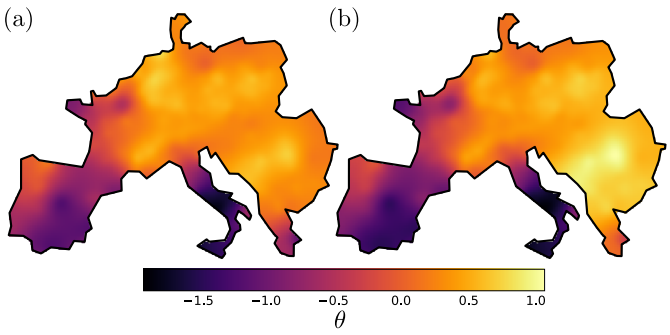

Fig. 6. Comparison of the steady state solution before (a) and after (b) application of the Fourier low-pass filter applied. The solutions are in good agreement. (Deviations are mainly found in a grid peripheral regions, e.g. in the Balkans.)

$b^{\alpha\beta}(\boldsymbol{r})$ as Neural Networks. Our preliminary tests show that an arbitrary initial choice for the maps complicates training, presumably due to a rather degenerate form of the respective loss function landscape. We expect however that running the same procedure with the initial condition computed according to the "artificial diffusion" approach discussed above will guarantee fast convergence of the NN training.

ACKNOWLEDGMENT

The authors thank Robert Ferrando and Christopher Koh for participating in the discussions which led to the manuscript.

APPENDIX

A. Details of the PDE discretization

We use the same spatial increment Δ for x and y axes, subsequently $\boldsymbol{r} = (i\Delta, j\Delta)$.

$$
\begin{aligned}
b_x(\boldsymbol{r}) &= \frac{b^x_{i,j-1} + b^x_{i,j}}{2}\,,\\
\partial_x b_x &= \frac{b^x_{i,j} - b^x_{i,j-1}}{\Delta} + O(\Delta^4)\,,\\
\partial_x \theta &= \frac{\theta_{i+1,j} - \theta_{i-1,j}}{2\Delta} + O(\Delta^4)\,,\\
\partial^2_x \theta &= \frac{\theta_{i-1,j} - 2\theta_{i,j} + \theta_{i+1,j}}{\Delta^2} + O(\Delta^4)\,.
\end{aligned}
$$

Similar expressions are obtained for the y-axis. Then discretization of the last term in Eq.(3) becomes

$$
\begin{aligned}
&\partial_x b_x \partial_x \theta + b_x \partial^2_x \theta + \partial_y b_y \partial_y \theta + b_y \partial^2_y \theta \approx \big(b^x_{i,j-1}\theta_{i,j-1}\\
&+ b^x_{i,j}\theta_{i,j+1} + b^y_{i-1,j}\theta_{i-1,j} + b^y_{i,j}\theta_{i+1,j} - \beta\theta_{i,j}\big)\Big/\Delta^2, \quad (8)
\end{aligned}
$$

where $\beta = b^x_{i,j-1} + b^x_{i,j} + b^x_{i-1,j} + b^y_{i,j}$. In order to make our numerical scheme more efficient we vectorize (re-index) the grid, and the field, $\theta(t;\boldsymbol{r})$ defined over the grid, according to $\theta_{i,j} \to \tilde{\theta}_k$, where $k = N_y(i-1) + j$. It results in the following re-indexing of the grid-neighbors: $i-1, j \to k-1,\ i+1, j \to k+1,\ i, j-1 \to k-N_y,\ i, j+1 \to k+N_y$. This results in reformulation of the principal part of Eq. (8) in terms of a matrix $\boldsymbol{\Xi}$ acting on the vector $\tilde{\boldsymbol{\theta}}$. Furthermore, with the convention that inner nodes, i.e. nodes that aren't on the

boundary layer, have a zero normal vector, $n_x = 0$ and $n_y = 0$, and introducing $\eta_{\pm}(x) = \{1 \text{ if } \pm x \geq 0\,;\ 0 \text{ otherwise}\}$, we rewrite $\boldsymbol{\Xi}$, therefore accounting for the Neumann boundary conditions (5),

$$\begin{aligned}\Xi_{kl} = &- \tilde{\beta}_k \delta_{k,l} + \eta_+(n_x)\tilde{b}^x_{k-N_y}\delta_{k-N_y,l} + \eta_-(n_x)\tilde{b}^x_k\delta_{k+N_y,l} \\ &+ \eta_+(n_y)\tilde{b}^y_{k-1}\delta_{k-1,l} + \eta_+(n_y)\tilde{b}^y_k\delta_{k+1,l}\,, \end{aligned} \tag{9}$$

where $\tilde{\beta}_k = \eta_-(n_x)\tilde{b}^x_k + \eta_+(n_x)\tilde{b}^x_{k-N_y} + \eta_-(n_y)\tilde{b}^y_k + \eta_+(n_y)\tilde{b}^y_{k-1}$ and $\delta_{.,.}$ is the Kronecker product. It is important that the method used for the numerical integration of the PDE is a finite volume method. This class of methods is conservative. This means that there is zero flux leakage at the boundary by construction which, in particular, guaranties that the post-fault system frequency is indeed at the value it is expected to be.

B. Details of the PDE numerical scheme

We use the Crank–Nicolson method [23] to integrate PDE (3). At each time step we solve the following system of linear equations

$$\boldsymbol{A}\begin{bmatrix}\tilde{\boldsymbol{\theta}}(t+\Delta t)\\ \tilde{\boldsymbol{\omega}}(t+\Delta t)\end{bmatrix} = \boldsymbol{B}\begin{bmatrix}\tilde{\boldsymbol{\theta}}(t)\\ \tilde{\boldsymbol{\omega}}(t)\end{bmatrix} + \boldsymbol{C}\,,$$

where

$$\begin{aligned}\boldsymbol{A} &= \begin{bmatrix}\mathbb{1} & -\frac{\Delta t}{2}\mathbb{1}\\ -\frac{\Delta t}{2}\boldsymbol{M}^{-1}\boldsymbol{\Xi} & \mathbb{1} + \frac{\Delta t}{2}\boldsymbol{\Gamma}\end{bmatrix},\\ \boldsymbol{B} &= \begin{bmatrix}\mathbb{1} & \frac{\Delta t}{2}\mathbb{1}\\ \frac{\Delta t}{2}\boldsymbol{M}^{-1}\boldsymbol{\Xi} & \mathbb{1} - \frac{\Delta t}{2}\boldsymbol{\Gamma}\end{bmatrix},\\ \boldsymbol{C} &= \begin{bmatrix}\mathbb{0}\\ \frac{\Delta t}{2}\boldsymbol{M}^{-1}\Big(\tilde{\boldsymbol{p}}(t+\Delta t) + \tilde{\boldsymbol{p}}(t)\Big)\end{bmatrix},\end{aligned}$$

with $\boldsymbol{M} = \mathrm{diag}(\tilde{\boldsymbol{m}})$ and $\boldsymbol{\Gamma} = \mathrm{diag}(\tilde{\boldsymbol{m}}^{-1}\tilde{\boldsymbol{d}})$.

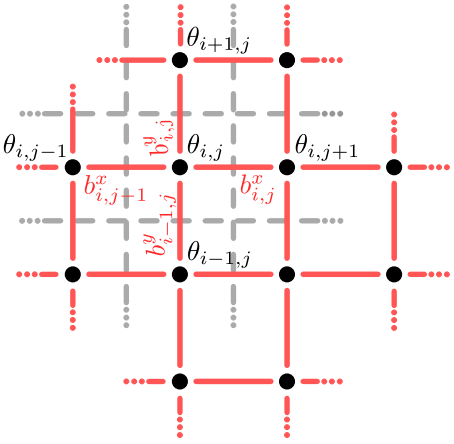

Fig. 7. For this work, we adopt a regular lattice to represent of the continuum. The dashed square represent a unit cell.

REFERENCES

[1] L. Pagnier and P. Jacquod, "Inertia location and slow network modes determine disturbance propagation in large-scale power grids," *PloS one*, vol. 14, no. 3, p. e0213550, 2019.

[2] M. Tyloo, L. Pagnier, and P. Jacquod, "The key player problem in complex oscillator networks and electric power grids: Resistance centralities identify local vulnerabilities," *Science advances*, vol. 5, no. 11, p. eaaw8359, 2019.

[3] P. Kundur, *Power System Stability and Control*. MacGraw-Hill, Inc., 1993.

[4] J. Machowski, J. Bialek, J. R. Bumby, and J. Bumby, *Power system dynamics and stability*. John Wiley & Sons, 1997.

[5] A. R. Bergen and V. Vittal, *Power systems analysis*. Prentice Hall, 2000.

[6] E. Hazan, "Introduction to online convex optimization," *Foundations and Trends in Optimization*, vol. 2, no. 3-4, pp. 157–325, 2016. [Online]. Available: http://dx.doi.org/10.1561/2400000013

[7] C. Ma, J. Wang, and W. E, "Model reduction with memory and the machine learning of dynamical systems," *arXiv:1808.04258*, 2018.

[8] R. Swischuk, L. Mainini, B. Peherstorfer, and K. Willcox, "Projection-based model reduction: Formulations for physics-based machine learning," *Computers & Fluids*, vol. 179, pp. 704–717, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045793018304250

[9] K. Bhattacharya, B. Hosseini, N. B. Kovachki, and A. M. Stuart, "Model reduction and neural networks for parametric pdes," 2021.

[10] W. Chen, Q. Wang, J. S. Hesthaven, and C. Zhang, "Physics-informed machine learning for reduced-order modeling of nonlinear problems," *Journal of Computational Physics*, vol. 446, p. 110666, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0021999121005611

[11] R. King, O. Hennigh, A. Mohan, and M. Chertkov, "From Deep to Physics-Informed Learning of Turbulence: Diagnostics," *arXiv:1810.07785*, 2018.

[12] I. Lagaris, A. Likas, and D. Fotiadis, "Artificial neural networks for solving ordinary and partial differential equations," *IEEE Transactions on Neural Networks*, vol. 9, no. 5, p. 987–1000, 1998. [Online]. Available: http://dx.doi.org/10.1109/72.712178

[13] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics informed deep learning (part ii): Data-driven discovery of nonlinear partial differential equations," *arXiv:1711.10566*, 2017.

[14] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, pp. 422–440, 2021.

[15] G. S. Misyris, A. Venzke, and S. Chatzivasileiadis, "Physics-informed neural networks for power systems," *arXiv:1911.03737*, 2020.

[16] L. Pagnier and M. Chertkov, "Physics-informed graphical neural network for parameter & state estimations in power systems," *arXiv:2102.06349*, 2021.

[17] ——, "Embedding power flow into machine learning for parameter and state estimation," *arXiv:2103.14251*, 2021.

[18] A. Semlyen, "Analysis of disturbance propagation in power systems based on a homogeneous dynamic model," *IEEE Transactions on Power Apparatus and Systems*, no. 2, pp. 676–684, 1974.

[19] J. S. Thorp, C. E. Seyler, and A. G. Phadke, "Electromechanical wave propagation in large electric power systems," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 45, no. 6, pp. 614–622, 1998.

[20] M. Parashar, J. S. Thorp, and C. E. Seyler, "Continuum modeling of electromechanical dynamics in large-scale power systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 51, no. 9, pp. 1848–1858, 2004.

[21] P. Parrilo, S. Lall, F. Paganini, G. Verghese, B. Lesieutre, and J. Marsden, "Model reduction for analysis of cascading failures in power systems," in *Proceedings of the 1999 American Control Conference (Cat. No. 99CH36251)*, vol. 6, 1999, pp. 4208–4212 vol.6.

[22] J. H. Chow, *Power system coherency and model reduction*. Springer, 2013, vol. 84.

[23] J. Crank and P. Nicolson, "A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 43, no. 1, p. 50–67, 1947.