A Data-Driven Method For The Development of System Frequency Response Models For Frequency Stability Analysis

Achilleas I. Sfetkos, Eleftherios O. Kontis, Theofilos A. Papadopoulos, and Grigoris K. Papagiannis

***Abstract*—Modern power systems are characterized by reduced inertia and primary frequency response as a result of the replacement of conventional synchronous generators (SG) with converter-interfaced renewable energy sources, deteriorating frequency stability. In this context, a novel data-driven methodology is proposed to derive an equivalent aggregated system frequency response (SFR) model that is capable of simulating the power system frequency response following a disturbance. The methodology utilizes active power and frequency response measurements to derive the SFR model through a nonlinear least squares optimization approach. The accuracy of the proposed method is validated by Monte Carlo simulations conducted on the IEEE 9-Bus test system, under both transient events and normal operating conditions. The validation is based on two main aspects. Initially, the model parameters estimated using the proposed data-driven approach are compared with those obtained through analytical calculations. Further, the effectiveness of the proposed approach is evaluated by determining the frequency response of the examined power system under varying types and amplitudes of disturbances. Results verify that in all scenarios the proposed approach provides results similar to those obtained via detailed non-linear dynamic simulations.**

***Keywords*—Data-driven approach, frequency stability, inertial response, non-linear least squares, parameter estimation, primary frequency response, system frequency response models.**

I. Introduction

In modern power systems, conventional synchronous generators (SGs) are being decommissioned and replaced with converter-interfaced renewable energy sources (CI-RES). CI-RES are mechanically decoupled from the power system and therefore do not inherently contribute to the inertial response [1], [2]. As a result, higher rates of change of frequency (RoCoF), larger frequency deviations, and longer recovery times are observed, jeopardizing the stability and the reliable operation of the power system [1], [2].

For this reason, system operators should be able to predict the frequency response characteristics of their systems. Towards this objective, a computationally inexpensive low-order equivalent system frequency response (SFR) model for the inertial response and the primary frequency response (PFR) was first proposed in [3]. In [4], the authors proposed an analytical way to aggregate a multi-machine power system into such equivalent SFR. This work has been extended in [5] to include the aggregation of voltage source converter-based high voltage direct current links and wind turbines. Such approaches require the *a priori* knowledge of the frequency response parameters of all conventional and CI-RES units, which becomes especially challenging when considering distributed generation.

As a result, it is necessary to derive the equivalent SFR using exclusively system measurements. In the literature, several data-driven approaches have been proposed for the derivation of equivalent models for individual SGs. For example, in [6], the authors estimate the parameters of a low order SG model through an unscented Kalman filter approach. In [7], the authors estimate the model of an individual governor through the time-domain vector fitting (TD-VF) system identification method, focusing on hydro turbine governors. Nevertheless, no data-driven methods are available in the literature for the development of aggregated SFR models, i.e., SFR models that describe the frequency response of the overall power system.

In this paper, a novel data-driven methodology is developed for the identification of an aggregated SFR model using frequency and active power measurements. At the first step, the proposed methodology derives the inertia constants of individual system components using the TD-VF [8]. These are then used to determine the power system center of inertia (COI) and the resulting frequency response at the COI. This is an important aspect/feature to account for unspecified inertia that is not provided as a nameplate property, e.g., in dynamic loads [9] and virtual synchronous generators (VSG). At the next step, the COI frequency response and the total active power deviation are used to derive the aggregated SFR parameters on the basis of non-linear least squares (NLS) optimization. The accuracy of the proposed method is statistically validated by means of Monte Carlo (MC) analysis. The validation is performed on two aspects. Initially, SFR model parameters, estimated using the proposed data-driven method, are compared with SFR parameters determined via the aggregation formulas presented in [4]. Additionally, a second validation is performed as follows: The proposed data-driven SFR is used: i) to predict the frequency response of the power system under different disturbances and ii) to quantify several key frequency response indicators such at the RoCoF,

The research work was supported by the Hellenic Foundation for Research and Innovation (HFRI) under the 5th Call for HFRI PhD Fellowships (Fellowship Number: 19296).

A.I. Sfetkos, T.A. Papadopoulos and G.K. Papagiannis are with the School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece (e-mail: sfetkos@ece.auth.gr, thpapad@ece.auth.gr, gpapagia@ece.auth.gr).

E.O. Kontis is with the Department of Electrical and Computer Engineering, University of Thessaly, 38334 Volos, Greece (e-mail: elkontis@uth.gr).

Paper submitted to the International Conference on Power Systems Transients (IPST2025) in Guadalajara, Mexico, June 8-12, 2025.

the frequency nadir, and the post-disturbance steady-state frequency deviation. The estimated frequency response and the resulting indicators are compared with those obtained via simulations, performed on the detailed power system model. By the comparison of the results better insights regarding the accuracy and the applicability range of the proposed approach are obtained.

The rest of the paper is organized as follows: Section II provides the theoretical background to set the notation for the rest of the paper. The proposed data-driven methodology is presented in Section III. In Section IV the impact of all SFR parameters on the resulting frequency response is quantified by means of sensitivity analysis. In Section V, the proposed method is evaluated by MC analysis using transient data, considering two scenarios: one where only conventional SGs are considered, and another where a VSG is considered alongside with conventional SGs. In Section VI, the method is evaluated using data from normal operating conditions, i.e., ambient data. For the purpose of the analysis, a set of 100 MC simulations is performed on the IEEE 9-Bus benchmark system. Finally, Section VII summarizes the main findings and concludes the paper.

II. Theoretical Background

A. SG Primary Frequency Response

Fig. 1 shows the PFR control model of a SG with a reheat steam turbine, represented in the Laplace domain. ΔF_i is the per unit (p.u.) frequency deviation on the SG bus, and $\Delta P_{m,i}$ is the difference (in p.u.) of the output mechanical power. Index i refers to the i-th SG. Moreover, parameter R_i represents the droop constant, $T_{G,i}$ the governor time constant, $T_{C,i}$ the steam chest time constant, $T_{R,i}$ the reheat time constant, and $F_{H,i}$ the high-pressure turbine fraction [4].

Fig. 1 also presents two non-linearities imposed by physical system dynamics, namely the governor deadband (GDB) and the turbine generator rate constraints (GRC). The values considered are $0.036\,\text{Hz}$ and 3% per minute, respectively [10].

B. Aggregated System Frequency Response Model

Fig. 2 depicts an SFR model describing the inertial behavior and the PFR of a power system. Specifically, the model predicts the frequency deviation Δf when a change in the electric load ΔP_L occurs. Both quantities are in the p.u. system. In this model, H denotes the equivalent inertia constant, D represents the equivalent damping factor, due to the frequency dependence of the loads, and the SG damper windings [11]. A linear feedback loop of an equivalent governor-turbine model is added to represent the PFR of the power system, as detailed in Section II-A. Typical ranges of the SFR parameters are listed in Table I.

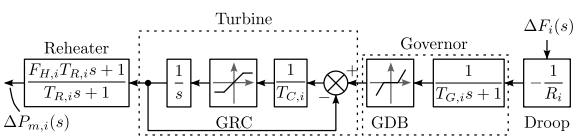

Fig. 1. Primary frequency control of a reheat steam turbine SG [4], [10].

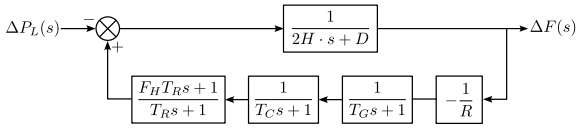

Fig. 2. Equivalent SFR model [4].

TABLE I
Typical Ranges of SFR Parameters [4].

Param.	H [s]	D [-]	R [-]	T_G [s]	T_C [s]	T_R [s]	F_H [-]
Min.	3	0	0.04	0.15	0.2	6	0.15
Max.	9	2	0.1	0.3	0.5	14	0.4

C. Frequency Response

1) Frequency Response Indicators: A graphical representation of a frequency response of a power system when excited by a step disturbance is presented in Fig. 3, where the most important frequency response indicators [12] are also illustrated. These include the pre-disturbance frequency f_{start}, the start time of the disturbance, t_{start}, RoCoF at t_{start}, the minimum (or maximum) instantaneous frequency deviation, f_{extreme}, the time at which the maximum frequency deviation occurs, t_{extreme}, and the steady-state frequency deviation, Δf_{ss}. The maximum frequency deviation and the time to reach are defined as $\Delta f_{\text{extreme}} = |f_{\text{extreme}} - f_{\text{start}}|$ and $\Delta t_{\text{extreme}} = t_{\text{extreme}} - t_{\text{nominal}}$, respectively [12].

2) RoCoF calculation: A discrete approximation of the RoCoF is given by (1) [13].

$$RoCoF = \frac{f_n - f_{n-1}}{\Delta t} \tag{1}$$

where f_{n-1} and f_n correspond to two consecutive frequency samples and Δt is the sampling rate. Estimating the RoCoF from two consecutive samples can be inaccurate due to harmonics and noise. To solve this problem, the guidelines specified in [13] are followed. First, the frequency signal is processed through a low-pass filter (LPF). The RoCoF is then determined through a sliding window of five consecutive measurements. Zero-phase filtering is used to smooth the frequency waveform using a 5-th order Butterworth LPF with a cutoff frequency of $3\,\text{Hz}$. Furthermore, a resolution of $100\,\text{ms}$ is used for the sliding window, as proposed in [13].

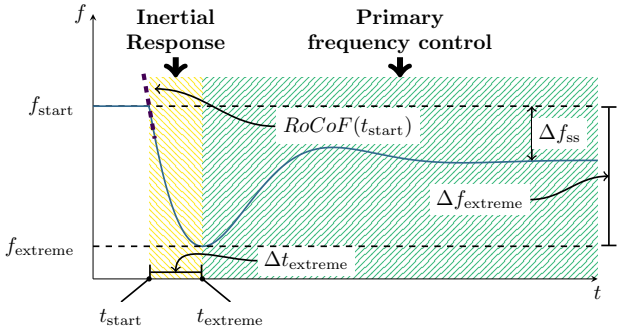

Fig. 3. Graphical representation of the frequency response with relevant indicators [12].

D. Inertia Estimation Using System Responses

The effective inertia constant at a point of measurement can be identified from active power (Δp) and frequency (Δf) deviation measurements by developing low-order input-output transfer function (TF) models. The developed TF model that relates Δp and Δf has the following general form:

$$G(s) = \frac{\Delta f(s)}{\Delta p(s)} = \frac{\beta_{n-1}s^{n-1} + \beta_{n-2}s^{n-2} + ... + \beta_0}{\alpha_n s^n + \alpha_{n-1}s^{n-1} + ... + \alpha_0}. \quad (2)$$

The unknown set of model parameters in the numerator and denominator, i.e., $\boldsymbol{\theta} = [\beta_{n-1}, \beta_{n-2}, ..., \beta_0, \alpha_n, \alpha_{n-1}, ..., \alpha_0]$, is estimated using system identification techniques. In this work, the TD-VF [8] is employed. Eventually, the unknown inertia constant is estimated using the time-domain impulse response of $G(s)$ [8], defined as $r(t) = \mathcal{L}^{-1}\{G(s)\}$.

$$H \approx -\frac{1}{2 \cdot r(0)} \quad (3)$$

where $\mathcal{L}^{-1}\{\cdot\}$ denotes the inverse Laplace transform.

III. Proposed Methodology

This Section outlines the proposed data-driven methodology for the development of SFR models, as depicted in Fig. 4. The proposed framework assumes the availability of phasor measurement units (PMUs) at generation and step-down transmission substations as part of a wide area monitoring system (WAMS). PMUs are utilized to record active power and frequency deviation measurements, denoted as Δp_i and Δf_i, for the i-th measurement point, respectively. Calculations are performed in the MATLAB 2024a [14] environment.

A. Estimation of COI Frequency Deviation

Following an active power disturbance, the frequency is not the same throughout the whole system, as the generators oscillate against each other [12]. The center of inertia (COI) is an equivalent generator that represents the average behavior of the system. Its frequency deviation, Δf_{COI}, can be defined as [9]:

$$\Delta f_{\text{COI}} = \frac{\sum_i^N H_i \cdot \Delta f_i + \sum_j^M H_j \cdot \Delta f_j}{\sum_i^N H_i + \sum_j^M H_j} \quad (4)$$

where N is the number of generators and synchronous condensers, and M is the number of load buses with significant penetration of rotating machinery. Δf and H represent the frequency deviation and inertia at the respective measurement locations.

The authors of [4] showed that aggregated SFR model of Fig. 2 outputs the COI in large power systems. Thus, in order to derive the SFR model parameters via system responses, the frequency deviation at COI (Δf_{COI}) must be initially approximated. This approximation is carried out using (5).

$$\Delta f_{\text{COI}} \approx \frac{\sum_i H_i^{TDVF} \cdot \Delta f_i}{\sum_i H_i^{TDVF}} \quad (5)$$

In this context, Δf_i represents the frequency deviation measured at substation i, and H_i^{TDVF} denotes the effective inertia values, estimated by using TD-VF. To accurately estimate the COI, it is pivotal to include measurements from both generation and step-down transmission substations with significant motor loads. The latter provide an aggregated effect of the distribution network load inertia, thus, ensuring it is properly accounted for.

This work assumes the availability of measurement units as part of a WAMS. If no such infrastructure is available, methods such as [15] have been proposed in the literature to estimate the COI using local measurements.

Prior to Δf_{COI} estimation, both Δf_i and Δp_i signals are filtered through a LPF to mitigate the effect of noise and high-order harmonics, which can influence the TF fitting.

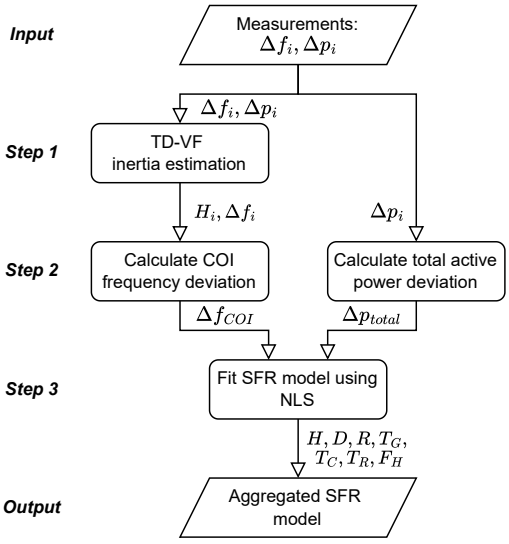

Fig. 4. Proposed method to estimate the aggregated SFR model.

B. Total Active Power Deviation

The total active power deviation, also required to derive the SFR model, is estimated from the recorded substation active power deviations Δp_i as:

$$\Delta p_{\text{total}} = \sum_i \Delta p_i. \quad (6)$$

Δp_{total} is produced by the generation units of the system to cover variations in load demand, and the total losses of the power system (sum of generator and network losses) [11].

C. Parameter Estimation

This Section focuses on fitting an SFR model to the measurement-obtained COI response, Δf_{COI}. The input of the model is the total recorded active power deviation, Δp_{total}, and its output is the estimation of the COI response, $\Delta \widehat{f}[n; \boldsymbol{\theta}]$. Here, n represents each measurement sample, and $\boldsymbol{\theta}$ is the parameter vector defined as:

$$\boldsymbol{\theta} = [H, D, R, T_G, T_C, T_R, F_H], \quad (7)$$

where the unknown $\boldsymbol{\theta}$ (see Fig. 2) are estimated via NLS approach. In more specific, $\boldsymbol{\theta}$ is estimated to minimize the

TABLE II
NLS TERMINATION CRITERIA [16].

Stopping Criterion	Value
Function tolerance	10^{-7}
Step tolerance	10^{-7}
First-order optimality tolerance	10^{-7}
Maximum number of iterations	1000

square of the residuals, i.e., $J(\boldsymbol{\theta})$, using the Trust Region Reflective algorithm [16],

$$J(\boldsymbol{\theta}) = \sum_{n} (\Delta f_{\text{COI}}[n] - \Delta \widehat{f}[n;\boldsymbol{\theta}])^2 \tag{8}$$

The parameters of $\boldsymbol{\theta}$, are bound between a lower and upper limit $\boldsymbol{\theta}_{\text{min}}$ and $\boldsymbol{\theta}_{\text{max}}$ respectively, detailed in Table I:

$$\boldsymbol{\theta}_{\text{min}} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}_{\text{max}}. \tag{9}$$

The NLS problem is solved iteratively until one or more of the predetermined criteria of Table II are met.

Due to the non-convexity of (8), the applied initial conditions might lead to local minima of J instead of the optimal solution. To this end, the MultiStart (MS) approach [17] is employed; the NLS solver is run repetitively assuming different initial conditions randomly distributed within the feasibility region defined by $\boldsymbol{\theta}_{\text{min}}$ and $\boldsymbol{\theta}_{\text{max}}$. Each solution $\boldsymbol{\theta}_i$ is stored and the one leading to the overall minimum value of J is selected. MS in this work employs 20 starting points. Once the procedure is completed, the SFR system of Fig. 2 is created using the identified model parameters.

IV. Parameter Sensitivity Analysis

It is crucial that the proposed data-driven methodology can identify accurately the most important model parameters presenting the highest sensitivity in the system dynamics, as even small deviations can affect considerably the accuracy of the developed SFR model. It is also expected that less sensitive parameters may exhibit higher estimation errors due to the inability of the NLS method to identify their influence in minimizing the cost function of (8), especially in the presence of noise and non-linearities. Therefore, a sensitivity analysis is performed to identify and characterize the obtained SFR parameters.

To perform the analysis, an initial SFR system is assumed, with parameters considered at the midpoint of the bounds listed in Table I, i.e., $[H, D, R, T_G, T_C, T_R, F_H]$ = [6, 1, 0.07, 0.225, 0.35, 10, 0.275], to calculate the original system step response. Afterwards, each system parameter is individually varied by $\pm 25\,\%$, and $\pm 50\,\%$, and the step response is recalculated. The impact of each parameter on the system step response is quantified by means of the coefficient of determination R^2 considering the original and the varied system step responses. In addition, the percentage differences in $RoCoF(t_{\text{start}})$, $\Delta t_{\text{extreme}}$, $\Delta f_{\text{extreme}}$, and Δf_{ss} between the two step responses are calculated.

From the so obtained results in Fig. 5, it is shown that H is the only parameter that affects $RoCoF(t_{\text{start}})$.

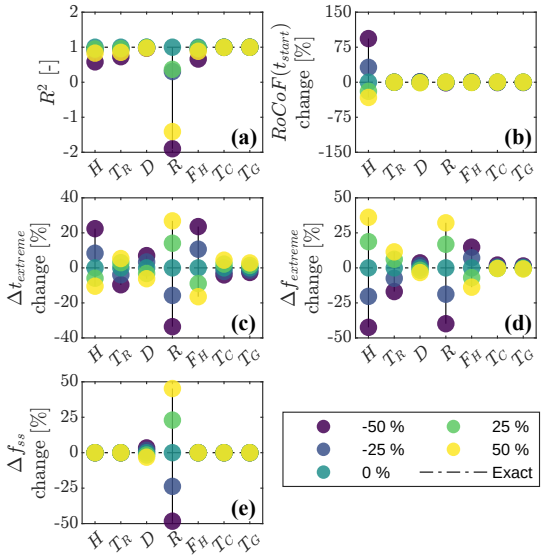

Fig. 5. Parameter sensitivity analysis: (a) R^2, (b) $RoCoF(t_{\text{start}})$, (c) t_{extreme}, (d) t_{extreme}, and (e) Δf_{ss}.

More specifically, $RoCoF(t_{\text{start}})$ is substantially increased with decreasing inertia constant. Reduced inertia levels also increase the maximum frequency deviation and shorten the time it is observed, but have no impact on the steady-state deviation. An increase in T_R causes an increase in the maximum frequency deviation while causing delay in the point observed. The opposite effect is observed for F_H. The droop constant R is a critical parameter; Higher values lead to higher maximum frequency deviations and to the times required to observe it. Furthermore, R influences the steady-state deviation, where a higher R causes a larger Δf_{ss}. This explains its strong influence on R^2. Opposite effects are observed for D, albeit at a significantly lower sensitivity level. Finally, the other time constants, namely T_C and T_G, exhibit a trivial impact on all metrics.

V. Application on Transient Responses

In this Section, the proposed methodology is assessed on the IEEE 9-Bus test system, as depicted in Fig. 6, using post-disturbance transient data obtained from 100 MC simulations. For each simulation, two distinct scenarios are considered regarding generation units: (1) the system consists of three conventional SGs, and (2), the system consists of two conventional SGs (G1 and G3) and one VSG, replacing G2. The conventional SGs provide inertial response and PFR, while the VSG provides only inertial response. The generator nominal apparent power (S), as well as H, D, R, T_G, T_C, T_R, and F_H values for both scenarios are listed in Table III. In the same Table the equivalent SFR parameters are calculated for both scenarios by the methodology of [4], assuming a p.u. base of $S_{\text{base}} = 519.5\,\text{MVA}$. Equivalent parameters R, T_G, T_C, T_R, and F_H for the second scenario display reduced values, since the VSG does not contribute to the PFR. Each conventional SG is equipped with an automatic voltage regulator (AVR). In the system, three constant power loads are considered, namely Load A ($P_A = 125\,\text{MW}, Q_A =$

TABLE III
FREQUENCY RESPONSE CHARACTERISTICS OF THE EXAMPLE SIMULATION.

Generator	S [MVA]	H [s]	D [-]	R [-]	T_G [s]	T_C [s]	T_R [s]	F_H [-]
G1	247.5	3.61	1.74	0.054	0.16	0.28	8.62	0.30
G2 - Scenario 1	192.0	7.60	0.47	0.053	0.20	0.32	11.35	0.31
G2 - Scenario 2	192.0	7.60	0.47	-	-	-	-	-
G3	128.0	4.38	0.11	0.092	0.15	0.25	11.91	0.25
Equivalent - Scenario 1	567.5	5.61	1.03	0.06	0.17	0.29	10.13	0.30
Equivalent - Scenario 2	567.5	5.61	1.03	0.09	0.16	0.27	9.39	0.29

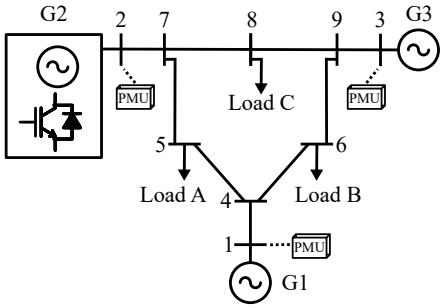

Fig. 6. Topology of the IEEE 9-Bus test system [18]; PMU locations are also shown.

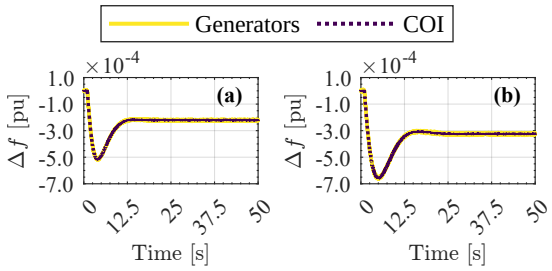

Fig. 7. Individual generator and COI frequency deviation: (a) scenario 1 and (b) scenario 2.

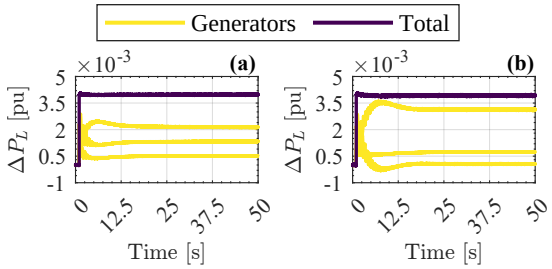

Fig. 8. Individual generator and total active power deviation: (a) scenario 1 and (b) scenario 2.

50 MVAr), Load B ($P_B = 90\,\text{MW}, Q_B = 30\,\text{MVAr}$) and Load C ($P_C = 100\,\text{MW}, Q_C = 35\,\text{MVAr}$). More details of the examined power system and the VSG implementation are provided in [18] and [19], respectively.

Simulations are performed using the RMS module of the DIgSILENT PowerFactory 2023 SP3A software [20], considering simulation duration of 50 s, i.e., $T_{\text{max}} = 50\,\text{s}$ and time step $T_s = 0.01\,\text{s}$. The frequency and active power deviation waveforms $\Delta f_i(t)$, and $\Delta p_i(t)$ are obtained through PMUs located at the three generator buses, as shown in Fig. 6. It should be noted that, due to the absence of motor loads, no additional measurement locations are considered. To simulate measurement noise, $\Delta f_i(t)$, and $\Delta p_i(t)$ responses are intentionally distorted with white Gaussian noise, assuming a signal-to-noise ratio (SNR) of 30 dB. The first 0.8 s of the signal is used for the TD-VF procedure, while the entire recorded signal is utilized for the NLS procedure. A third-order Butterworth LPF with a cutoff frequency at $f_c = 0.5\,\text{Hz}$ is used to preprocess the signal prior to inertia estimation. This cutoff frequency is selected because the inertial response frequency spectrum generally lies below this value as reported in [21].

A. Indicative Example

An indicative simulation, where a step load increase of 2 MW is applied to Load B at $t_{\text{start}} = 1\,\text{s}$ is used to demonstrate the application of the proposed method. The resulting frequency deviation waveforms of the three SG buses are shown in Fig. 7a and Fig. 7b for scenarios 1 and 2, respectively. The influence of the VSG not participating in the PFR is evident in Fig. 7b, where it can be inferred that it leads to more profound frequency nadirs and largest steady-state deviations. The identical initial RoCoF verifies that the equivalent system inertia is identical in both scenarios.

The TD-VF is used to estimate inertia constants of the individual SGs, required to calculate the COI frequency deviation. The SG inertia estimates $[\hat{H}_1, \hat{H}_2, \hat{H}_3]$ are $[3.39\,\text{s}, 7.61\,\text{s}, 4.28\,\text{s}]$ and $[3.38\,\text{s}, 6.88\,\text{s}, 4.34\,\text{s}]$ for scenarios 1 and 2, respectively. These correspond to absolute estimation errors of 6.19%, 0.15%, and 2.19% for scenario 1, and 6.27%, 9.52%, and 0.88% for scenario 2, relative to the original values reported in Table III. Using (5), the COI response is approximated and juxtaposed with the individual recorded SG responses in Fig. 7.

Fig. 8 shows the active power deviation of each SG, as obtained at their respective buses. Their sum yields the total power system active power deviation according to (6), which is also plotted in Fig. 8. As constant power loads have been considered, the total power deviation is a step change, despite that the individual active power SG responses present some oscillatory post-disturbance behavior.

The NLS-estimated parameters are summarized in Table IV. To evaluate their accuracy, the absolute percentage error (APE) is calculated:

$$APE = \left| \frac{\hat{x} - x}{x} \right| \cdot 100\%. \tag{10}$$

TABLE IV
APE IN ESTIMATING THE EQUIVALENT SFR PARAMETERS THROUGH NLS FOR THE TWO TEST SCENARIOS OF THE EXAMPLE SIMULATION.

	Scenario 1			Scenario 2		
	Equivalent	NLS-derived	APE [%]	Equivalent	NLS-derived	APE [%]
H [s]	5.61	5.46	2.66	5.61	5.60	0.15
D [-]	1.03	2.03	97.43	1.03	1.12	9.15
T_R [s]	10.13	9.89	2.31	9.39	9.24	1.58
R [-]	0.06	0.06	6.33	0.10	0.09	4.13
F_H [-]	0.30	0.26	11.75	0.29	0.29	0.13
T_C [s]	0.29	0.28	3.38	0.27	0.27	0.53
T_G [s]	0.17	0.28	59.00	0.16	0.19	20.39

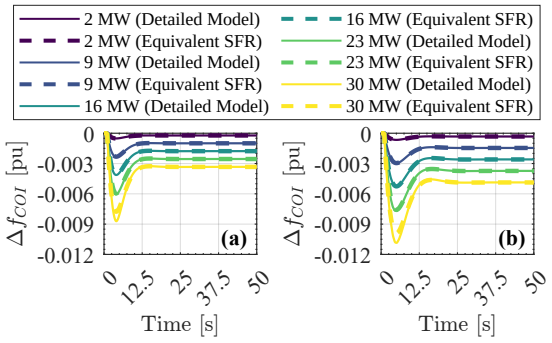

Fig. 9. Prediction of the system's response under step disturbances: (a) scenario 1, and (b) scenario 2.

In (10), x denotes the value of the parameter calculated via the method of [4] and $\hat{x}$ represents the estimates provided by the proposed data-driven method. Recall that the approach of [4] can only be applied if full knowledge of all individual components is available, i.e., information regarding time constants and individual characteristics of all SGs. As shown in Table IV, the damping constant D and the governor time constant T_G present a significant mismatch. However, results in the sensitivity analysis of Section IV show that these parameters have a minor impact on system dynamics, compared to other parameters, e.g., droop constant R which are accurately identified. The MS NLS procedure, performed on a system with an Intel(R) Core(TM) i7-13700 processor (2.10 GHz) and 32 GB of RAM, required 1.37 s to complete.

After deriving the equivalent SFR model, its accuracy is assessed on simulating the frequency response of the examined power system under different disturbances, i.e., disturbances different to the original ones used for the development of the model. This validation test practically verifies the generalization capability of the derived SFR model. For this purpose, the IEEE 9-bus test system is excited under five distinct step disturbances in Load B of i) 2 MW ii) 9 MW iii) 16 MW iv) 23 MW and v) 30 MW and five ramp disturbances with slopes i) 0.5 MW/s ii) 1 MW/s iii) 1.5 MW/s iv) 2 MW/s and v) 2.5 MW/s. All disturbances are simulated in the DigSILENT PowerFactory software using the detailed nonlinear model of the examined power system.

The continuous lines in Fig. 9a and Fig. 9b demonstrate the actual frequency step response of the multi-machine system derived through DigSILENT PowerFactory, while the dashed lines show the step response simulated by the

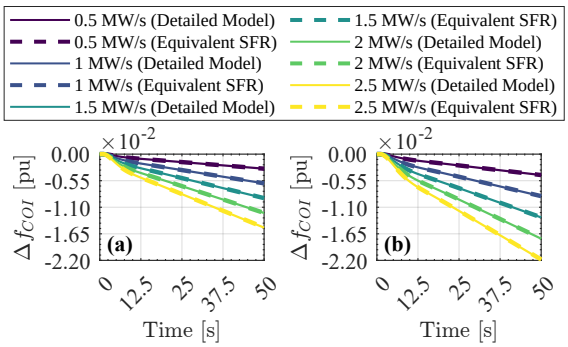

Fig. 10. Prediction of the system's response under ramp disturbances: (a) scenario 1, and (b) scenario 2.

derived SFR model for scenarios 1 and 2. Based on the results it can be concluded that the developed SFRs models successfully capture the frequency response of the first four step disturbances, including RoCoF at t_{start}, $\Delta f_{\text{extreme}}$, and Δf_{ss}. Conversely, it underestimates the frequency nadir for the largest disturbance by 10.80 % and 8.77 % in the two scenarios. This can be attributed to the activation of the GRC, which was triggered by the large disturbance amplitude. Nevertheless, it should be indicated that 30 MW corresponds to 9.5% of the total load, making this an extreme disturbance case. Even in that case, Δf_{ss} is correctly estimated.

Lastly, Fig. 10 illustrates the response of the NLS-derived system under ramp disturbances, along with the response simulated by the proposed data-driven SFR method for both scenarios. Results demonstrate that the proposed method correctly identifies the frequency response of the system during also ramp disturbances.

B. Monte-Carlo Simulations

To further evaluate the performance of the proposed method in a wide range of operating conditions, 100 MC simulations are applied on the 9-bus system for the two scenarios. In each MC simulation, all frequency-relevant parameters (H, D, R, T_G, T_C, T_R, F_H) are randomly varied between the respective lower and upper limits of Table I. To excite the system, a 2 MW step disturbance is applied randomly to one of the loads at $t_{\text{start}} = 1$ s.

Fig. 11 shows the distribution of APE for each inertia constant in all MC simulations using boxplots. The estimated inertia constants are used to derive the COI frequency as per (5). All generators exhibit similar errors, with median values around 3%, and upper quartile less than 11%. Finally, a limited number of outliers is presented.

The total active power and COI measurements are further used to derive the SFR model through NLS optimization. Fig. 12 shows the APE on the SFR parameters derived via the proposed data-driven methodology, with respect to the values obtained on the basis of the aggregation method of [4]. Specifically, Fig. 12a shows the APE on the most sensitive, i.e., H, T_R, R, and F_H and Fig. 12b shows the APE on the least sensitive parameters, i.e, D, T_C, and T_G. From Fig. 12b it can be inferred that the latter parameters exhibit significant differences with respect to the aggregation calculations of [4].

As these parameters have only a trivial influence on the overall system dynamics, the level of inaccuracy is inconsequential. In scenario 1, the most impactful parameters are estimated with low APE. Specifically, median and maximum values (non-outlier) are 1.70%, 1.41%, 3.40%, and 6.58%, and 4.45%, 5.21%, 9.76%, and 18.06% for parameters H, T_R, R, and F_H, respectively. The same can be concluded for the most impactful parameters in scenario 2, where median and maximum values (non-outlier) are 1.25%, 0.77%, 2.85%, and 7.97%, and 3.78%, 3.26%, 10.07%, and 27.40% for parameters H, T_R, R, and F_H, respectively.

Next, both the DigSILENT PowerFactory model of each MC and the derived SFR model are subject to different step disturbances, namely those detailed in Section V-A, in order to statistically quantify the generalization capability of the latter.

Fig. 13 and Fig. 14 show the APE of the main frequency indicators, comparing the multi-machine model to the data-driven SFR across all MC simulations and step disturbances for scenarios 1 and 2. In both figures, a small number of outliers were removed to better highlight the underlying trends. As verified by the corresponding APE, the RoCoF at the moment of the disturbance is accurately estimated by the proposed SFR, since the mean APE is less than 3.5%, regardless of the level of the disturbance. On the other hand, indicators influenced by the GRC, i.e, $\Delta t_{\text{extreme}}$ and $\Delta f_{\text{extreme}}$, show progressively increasing maximum errors as the disturbance level increases, reaching almost 3.5% and 17%, respectively, for a 30 MW disturbance in scenario 2. However, as previously outlined, a 30 MW disturbance represents an extreme case relative to the examined system. Even in such case, the median values of these errors remain within acceptable limits. For smaller disturbances, the SFR model accurately captures the characteristics of the frequency extrema. The Δf_{ss} prediction error is practically constant and always less than 1%, since the influence of non-linearities is diminished following the dynamic phenomenon, thus allowing for accurate prediction by the SFR model.

VI. Application on Normal Operating Condition Data

In this Section the proposed methodology is evaluated using normal operating condition (ambient) data, through 100 MC simulations. The system of Fig. 6 is utilized, assuming only conventional SGs. To simulate normal operating conditions, the total load of the system is randomly varied within $\pm 0.1\,\%$ of its nominal value every 5 s. The simulation time is $T_{\text{max}} = 600\,\text{s}$ and the time step is $T_s = 0.01\,\text{s}$.

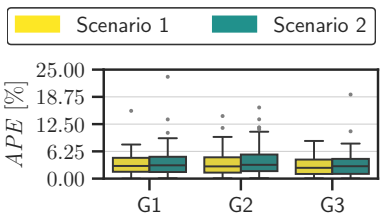

Fig. 11. APE distribution for each TD-VF-estimated inertia constant.

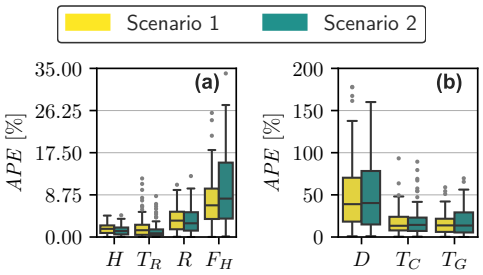

Fig. 12. APE distribution in estimating the equivalent SFR parameters through NLS. (a) Most sensitive parameters, and (b) least sensitive parameters.

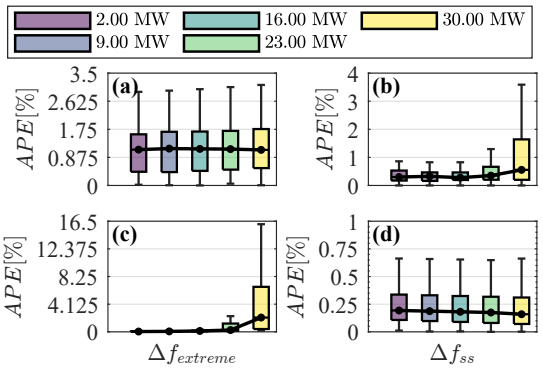

Fig. 13. Distribution of APE in frequency response indicators for different step disturbances for scenario 1 MC simulations: (a) RoCoF at the start of the disturbance, (b) time to maximum frequency deviation, (c) maximum frequency deviation and (d) steady-state frequency deviation.

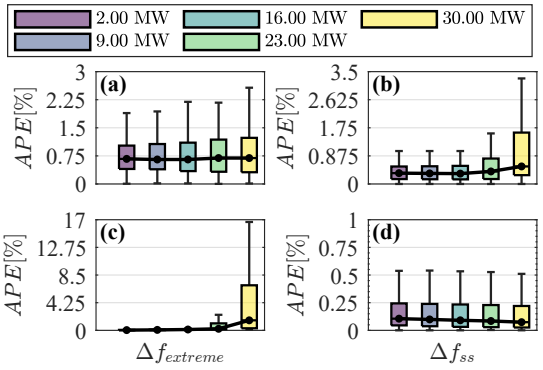

Fig. 14. Distribution of APE in frequency response indicators for different step disturbances for scenario 2 MC simulations: (a) RoCoF at the start of the disturbance, (b) time to maximum frequency deviation, (c) maximum frequency deviation and (d) steady-state frequency deviation.

The TD-VF method is utilized to estimate the COI by using (5). To increase the estimation accuracy, a sliding window of 120 s is applied on measurements, with a 30 s offset between consecutive windows. To reduce computational burden, the signals are downsampled to 50 samples per second. Additionally, to reduce the influence of higher-order harmonics on inertia estimation, the responses are filtered by a 15-th order LPF with cut-off frequency at $f_c = 5$ Hz [8]. The final estimate is obtained as the mean value of all window-based estimations that result in inertia values within the limits of Table I and are derived through stable transfer functions $G(s)$. Fig. 16a, depicts the APE for each SG inertia constant. The median APE

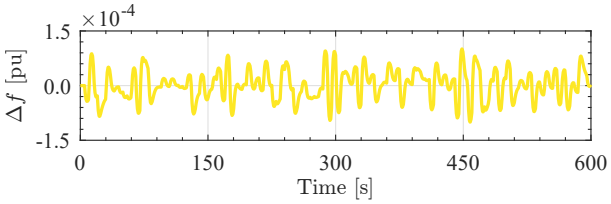

Fig. 15. Normal operating condition COI frequency deviation.

is 5.45%, 4.63%, and 4.85% for G1, G2, and G3 respectively, while the maximum (non-outlier) values are 23.24%, 14.62%, and 19.54%. Compared to the application on transient data, the estimation presents higher APE, more outliers and requires higher computational time. This can be attributed to the influence of both PFR and inertial response during normal operation, unlike transient conditions where inertial response is dominant following the onset of a disturbance [8].

The aggregated SFR parameters are then identified through the NLS procedure for all MC simulations. For this procedure, the full-length 600 s signals are used. Fig. 16b shows the APE on the most sensitive parameters, i.e., H, T_R, R, and F_H and Fig. 16c depicts the APE on the least sensitive parameters, i.e, D, T_C, and T_G. Similarly to the transient response case, the latter exhibit significant differences with respect to the aggregation calculations. On the contrary, the most impactful parameters are estimated with low APE, similar to the transient response case. The median and maximum (non-outlier) values are 1.39%, 3.17%, 3.26%, and 5.93%, and 3.85%, 9.30%, 11.16%, and 16.54% for H, T_R, R, and F_H, respectively.

In summary, the proposed method is effective with both normal operating condition data and post-disturbance transient data. The first type is more computational demanding, requiring longer time windows for both TD-VF and NLS, alongside high-order LPF and signal preprocessing.

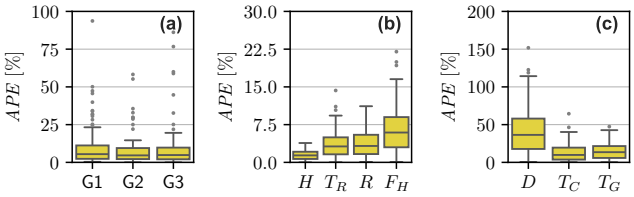

Fig. 16. Normal operating condition case study estimation APE: (a) inertia constants, (b) most sensitive parameters, and (c) least sensitive paramaters.

VII. Conclusions

In this paper, a robust data-driven methodology is proposed to estimate an aggregated SFR model. The methodology employs frequency and active power measurements in order to estimate the constants of the SFR model through a NLS parameter estimation procedure. The methodology is validated on the IEEE 9-bus test system via 100 MC simulations using transient and normal operating condition data. An additional scenario where one SG is replaced with a VSG was also examined. Results show that the derived SFR effectively captures the dynamic behavior and predicts frequency response indicators with high accuracy under various disturbances, even under the impact of governor non-linearities and the simulated measurement noise. Moreover, the most important SFR parameters according to the performed sensitivity analysis (that is, the inertia constant, the droop constant, the reheat time constant, and the high-pressure turbine fraction), are correctly identified by the procedure with respect to the analytically calculated aggregated parameters. Future research will focus on the extension to different generation mixes, i.e., different types of generation devices and technologies and different load types including rotating machinery.

References

[1] B. Tan, J. Zhao, M. Netto, V. Krishnan, V. Terzija, and Y. Zhang, "Power System Inertia Estimation: Review of Methods and the Impacts of Converter-Interfaced Generations," *Int. J. Electr. Power Energy Syst.*, vol. 134, 2022.

[2] "CIGRE WG C2C4. Impact of high penetration of inverter-based generation on system inertia of networks," *CIGRE TB 851*, 2021.

[3] P. Anderson and M. Mirheydar, "A low-order system frequency response model," *IEEE Trans. Power Syst.*, vol. 5, no. 3, pp. 720–729, 1990.

[4] Q. Shi, F. Li, and H. Cui, "Analytical Method to Aggregate Multi-Machine SFR Model With Applications in Power System Dynamic Studies," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6355–6367, 2018.

[5] K. Yan, G. Li, R. Zhang, F. Li, T. Jiang, X. Li, and H. Chen, "Aggregated SFR model for VSCHVDC interconnected power systems with high penetration of wind power," *Electr. Power Syst. Res.*, vol. 216, p. 109018, 2023.

[6] H. G. Aghamolki, Z. Miao, L. Fan, W. Jiang, and D. Manjure, "Identification of synchronous generator model with frequency control using unscented kalman filter," *Electr. Power Syst. Res.*, vol. 126, pp. 45–55, 2015.

[7] S. H. Jakobsen and K. Uhlen, "Vector fitting for estimation of turbine governing system parameters," in *2017 IEEE Manchester PowerTech*, 2017, pp. 1–6.

[8] N. E. Skopetou, A. I. Sfetkos, E. O. Kontis, T. A. Papadopoulos, and A. I. Chrysochos, "Identification of inertia constants using time-domain vector fitting," *Electr. Power Syst. Res.*, vol. 236, p. 110924, 2024.

[9] L. L. Fernandes, M. R. A. Paternina, D. Dotta, and J. H. Chow, "Data-driven assessment of center of inertia and regional inertia content considering load contribution," *Int. J. Electr. Power Energy Syst.*, vol. 156, p. 109733, 2024.

[10] H. Bevrani, *Robust Power System Frequency Control*. New York, NY, USA: Springer, 2009.

[11] F. Casado-Machado, J. L. Martinez-Ramos, and A. Marano-Marcolini, "Simplified Models for Frequency Studies in Electrical Power Systems," in *2019 IEEE Milan PowerTech*, 2019, pp. 1–6.

[12] ENTSO-E, "Future system inertia 2," Technical Report, 2018.

[13] ——, "Frequency measurement requirements and usage," Technical Report, 2018.

[14] The MathWorks, Inc., "MATLAB Version: 24.1.0 (R2024a)," Natick, Massachusetts, United States, 2024. [Online]. Available: https://www.mathworks.com

[15] M. Sun, G. Liu, M. Popov, V. Terzija, and S. Azizi, "Underfrequency load shedding using locally estimated rocof of the center of inertia," *IEEE Trans. Power Syst.*, vol. 36, no. 5, pp. 4212–4222, 2021.

[16] The MathWorks, Inc., *Optimization Toolbox User's Guide*, version R2024a. [Online]. Available: https://www.mathworks.com/help/pdf_doc/optim/index.html

[17] ——, *Global Optimization Toolbox*, version R2024a. [Online]. Available: https://www.mathworks.com/help/pdf_doc/gads/index.html

[18] DIgSILENT GmbH, "Nine-bus System," DIgSILENT PowerFactory 2023 SP3A, Gomaringen, Germany, 2023.

[19] ——, "Grid-forming Converter Templates," DIgSILENT PowerFactory 2023 SP3A, Gomaringen, Germany, 2023.

[20] ——, "DIgSILENT PowerFactory 2023 SP3A," Gomaringen, Germany, 2023. [Online]. Available: https://www.digsilent.de

[21] P. M. Ashton, C. S. Saunders, G. A. Taylor, A. M. Carter, and M. E. Bradley, "Inertia Estimation of the GB Power System Using Synchrophasor Measurements," *IEEE Trans. Power Syst.*, vol. 30, no. 2, pp. 701–709, 2015.