A REVIEW ON SYSTEM IDENTIFICATION IN POWER GENERATION SYSTEMS

Derya OZKAYA, Ilhan KOSALAY

ABSTRACT. Power generation systems with multiple input-multiple output have a wide operating range and may not be fully defined by a fixed model due to high-order nonlinear dynamics. As the parameters of the conventional excitation and speed governor controllers are determined by the system model which is linearized around one operating point, the performances of the controllers at different operating points can be reduced. Large disturbances encountered in the system can cause the controllers to operate outside the linear region. In addition, when the plant's operating structure changes with time or with changing environmental conditions, it is necessary to readjust the controller parameters. This readjustment is needed because the controller parameters that are set to provide the best performance at one operating point may not provide the same performance when the operating points change. In order to avoid the degradation in controller performance, system identification can be performed so that the controller parameters will have an adaptive structure. At the same time, it will be possible to make predictive maintenance, determine optimum operating points, diagnose faults and estimate performance by means of the power plant model built on the basis of system identification. In order to meet these requirements, system identification methods used in power generation systems have been examined throughout this review study and the performances obtained as a result of the changes made in the controllers have been compared.

1. INTRODUCTION

The time-worn nature of the equipment, the change of the operating conditions or the change of the operating point seen as a result of the disturbances can cause a degradation of the controller performance. System identification is a functional tool for avoiding the deterioration of the controller performance in the multiple input-
multiple output systems with nonlinear dynamics and obtaining an adaptive controller structure. System identification and determination of how the variable system parameters will affect the system will help both the controllers to be more successful and contribute to the availability of the system. Thanks to the system identification, a system model in which the parameters are updated according to changing operating conditions instead of a fixed model can provide additional contributions such as creating a predictive maintenance plan, determining optimum operating points, diagnosing fault and performance estimation.

2. SYSTEM IDENTIFICATION METHODS

The system identification process can be explained as model fitting to the experimental data recorded by giving appropriate values to the system parameters. There are basically two standard methods for system identification: parametric methods and nonparametric methods [1].

- Parametric methods: The method by which the recorded data is matched to the estimated parameter vector.
- Nonparametric methods: The preferred method in the preliminary steps for estimating the structure of the system when there is no need for priori information about the model structure or where there is no priori information.

There are three main steps, as shown in Figure 1, in the system identification process that can be considered as modeling from experimental data [2].

![Figure 1 Steps of system identification](image-url)
Data collecting: It is important to make a careful plan beforehand as the system to be identified may have economic risks to do experiments in itself. The time at which the signals are to be recorded shall be predetermined and the experiment shall be designed as it shall not affect the normal working conditions and the data obtained at the end of the experiment shall contain maximum information.

Model determination: The most complex part of system identification is the selection of the model to be used. Prior knowledge and engineering intuitions about the system to be identified should be considered together with the model properties that can be used. Should a first or second transfer function be used for the system, or should it be a linear model or a non-linear model? After this decision is made, a model can be created based on some unknown physical parameters.

Model evaluation/validation: The recorded data is reproduced by the model to determine whether the selected model is suitable for the system. The final step is model validation; several tests are performed to determine if the model is good enough for the intended use.

3. Research Review

In many of the studies about system identification, the issue is accounted as an optimization problem. Such as in study [3] particle swarm optimization method is used for system identification. According to evolutionary strategies of intelligent optimization algorithms, the state of each individual is constantly corrected until the error between the outputs of the identification system and the actual system is small enough. The least squares method neglects nonlinearities when small disturbance are present, but in the case of large disturbances the least squares method significantly corrupts the output causing discontinuities at the system output due to the nonlinear nature of the system. Methods such as genetic algorithms, neural networks, fuzzy logic are quite successful at incorporating non-linear features into the model, but this requires very high number of quality training data.

The parameters of the generator and excitation system are diagnosed simultaneously by online measurement and data obtained from SCADA, error loggers or phasor measurement devices in study [4]. The reference voltage value of the excitation system is used as input, terminal voltage and active power are used as output. Parameters to be identified are divided into three classes: parameters affecting terminal voltage, parameters affecting active power, and those not affecting any output. The parameters are iteratively identified by genetic algorithm optimization. Nine parameters are diagnosed in four steps. In off-line identification, parameter
changes originating from aging cannot be detected. Also, the network connection of the unit must be corrupted for a long time. In online methods, the reaction is observed by giving small disturbance to the unit that is connected to the grid. The result of the simulation in which the actual values of the parameters are used, is compared with the estimated results. Between each step the comparison of simulation and estimation results are performed. The results exhibits that the estimated value of the output signal of which parameters are updated matches with the simulation results. Whereas the values of the output related to the not updated parameters do not match. At the end of all steps, the comparison shows that the estimation is successful. When the operating point is changed, it is observed that the estimated parameters are also successful at the different working points.

The forgetting factor updated according to the changes of input and output signals improves the performance according to the exponential forgetting coefficient used in least square method which is the key of optimization problem in [5]. The hydraulic transients of unplanned unit outages or of starting up the unit are investigated. The study, which is implemented in Simulink model, is done on a power plant operating in isolated mode. All conditions have been evaluated for different operating conditions (such as starting up or stopping the unit). The reservoir and the tail water levels are considered constant. There is a constant PID controller on one side. On the other hand, the transfer function parameters estimated according to more than one simulation data and these parameters are used to calculate the Ziegler-Nichols critical gain and critical period values. In order to meet all operating conditions optimal fitting is done from the estimated parameters. The controller that contains the PI (no D) parameters calculated according to this parameter set is called adaptive.

Genetic algorithm (GA), gravitational search algorithm (GSA) and particle swarm optimization (PSO) method are widely used for optimization problems. In study [6] the parameter identification process is handled again as an optimization problem. There are two algorithms that have proven themselves in the parameter identification studies of power systems: GA and PSO both methods give effective results for nonlinear systems, but have some disadvantages. Population loss can occur in GA and PSO cannot reach the point where better results can be obtained because it is attached to local minima. The GSA is more successful than the other two. So the combination of the GSA and the PSO is used throughout the study. It is very important that the cost function to be used in the optimization problem depends on which parameters or outputs. Normally, the cost function depends on the system output, but the intermediate values are also taken into account in this study. Also, an equivalent weighting is not used for these intermediate outputs. The weights depend on the relationship between the parameters and the outputs. PID parameters, penstock, inertia and time constants are the parameters to be determined in the system identification process. When the cost function created using the output signal
is used, the diagnostic results obtained for the intermediate values (such as torque, wicket gate opening) do not match the original system whereas the identification results for the output signal in the system match with the original system results. When the cost function involved in the evaluation of intermediate values is used, the system identification and the original system match for both system output and intermediate values. The latter cost function has been used in progressive stages, since intermediate values are more successful in the cost function involved in the evaluation. GSA & PSO, GSA, PSO, GA methods are compared. GSA & PSO cost function is more successful in finding the optimum solution point. When the simulation is repeated at different study points, the GSA & PSO parameters are updated while the GA and PSO parameters remain unchanged.

Three different identification methods are used in study [7] which again considers the system identification as an optimization problem. To determine the governor parameters, the uncertainties in the system are included in the model. Three additional evolutionary algorithms are compared with the antlion optimizer ALO, PSO, GA, GSA. When comparing the best, mean and worst solutions, ALO achieved the best success in fitness function results.

The speed governor and the generator are investigated together in [8]. The study is started by establishing a mathematical model that is compatible with the turbine prototype. Finding unknown parameters in known patterns is considered as an optimization problem. The researchers use improved GSA so that the GSA alone is not enough. Improvements are made to develop the GSA's searching capability, including the PSO's communication and memory characteristics. Identification is also done separately with normal GSA and PSO. To get rid of the randomness, each algorithm is tried 30 times and then averaged. Compared to the other 2, the identification accuracy is higher in the IGSA, the errors are lower and converging to the optimal point faster.
Derya OZKAYA, IIhan KOSALAY

Figure 2 (a) Neuroidentifier training cycle, (b) Neurocontroller training cycle [9]

Beside considering the system identification as an optimization problem, artificial neural networks (ANN) can be preferred for system identification. In study [8] where two artificial neural networks are used as shown in Figure 2, one of the neural networks is used as a neuro-identifier and provides a dynamic model that is always up-to-date. The other ANN works as a neuro controller and is used as conventional governor controller and automatic voltage regulator (AVR). Simulation is done in MATLAB/Simulink, and then the study tested in the test setup created in the laboratory. Reasons for choosing Nonlinear Auto Regressive Moving Average (NARMA) among system identification models: the aim of online training and avoiding the feedback loop in the model. In addition, static back propagation is used to update NN weights. The controller performance and training is defined with the structure defined as "Desired response predictor" with the same outputs as neuro-identifier. Thanks to this structure, the controller's dynamic performance (rise time & damping) will be adjustable, the system will be stable, will provide power system stabilizer effect. Since neuro-identifier is used, there is not any requirement to know the nonlinear mathematical model of the system, and the controller becomes adaptive. In addition, since the inputs and outputs of the controller are deviation signals from steady state, the inputs and outputs of the controller become zero once the desired point of operation is reached. Compared to the conventional governor and AVR with MATLAB/Simulink results, the overshoot and oscillations are too much in former. By changing the line impedances, the system parameters are changed, the conventional controller enters a sustained oscillation but the new one
Artificial neural network approach is used to identify the faults observed in protection relays or unit circuit breakers in study [10]. The proposed approach uses alarm signals from the protection system and can use missing or corrupted data. Emergency cases are simulated and analyzed for each part of the power system. Supervisory Control and Data Acquisition (SCADA) system defects are also taken into account and all data obtained in the event of failure are used to train the neural network. Complex emergencies such as multiple device failures are investigated. After the test and evaluation, the training results are successful and the overall study makes the operator's job easier. Two failed to operate protection relays or circuit breakers can be detected in the case of two simultaneous failures in the power system even if there exist some errors in SCADA data such as the data show that a relay, which indeed work well, tripped and a relay indeed fails to operate but work well according to the SCADA data.

In study [11] in which two different artificial neural network identification systems with different activation functions are compared, the data sets to be used in neural network training are obtained by simulation. While obtaining the data sets, both reference signals and deviation signals from the reference value are used. It has been observed that neural networks successfully track the target value changes in the experiments in which the case that can be encountered in the plant such as sensor deterioration, data loss, termination of training, etc. When the Multilayer Perceptron (MLP) neural network is trained with deviation signals from the target value, training converges faster than Radial Basis Function (RBF) neural network and more successful system identification graphs are obtained. However, MLP neural network failed to identify the system when reference signals are used. In addition, some of the data for activation function of the RBF neural network needs to be computed off-line before training, which increases the calculation-based processing load.

The nonlinear model of hydroelectric power plant is identified via neural network nonlinear autoregressive with exogenous signal (NNARX) model, the block diagram of which is shown in Figure 3, in [12]. The measurements from the simulation of power plant are used for identification. It is observed that the active power deviation amount followed the controller output through the predictive
control approach established with artificial neural networks, and the controller performance showed the same success in variable load and water flow.

![Diagram](image.png)

**Figure 3** Identification and predictive control block scheme [12]

Together with the artificial neural networks, fuzzy logic is also used in order to identify the system parameters. All of the tests are carried out using different time constants penstock in [13]. The nonlinear equations of the hydro electrical power plant are solved in the digital computer to obtain the data for system identification. 0.005 variance pseudo-random signals are superimposed on the signal with 0.01 variance to make changes in the load. With the same approach, a random wicket gate opening input signal is also generated. The disadvantage of doing such an offline system identification is that it requires much more data, but it is preferred because of lesser training time. The NNARX model is compared with the adaptive neuro fuzzy inference system (ANFIS) model and ANFIS achieves more successful estimates.

In [14], computer simulation in which the non-linear equations are used to create a mathematical model, is done for a thermal power plant. With a fuzzy neural network identifier it is tested whether the system can identify the transient conditions that occur in the system after any fault such as 3-phase short-circuit faults. The identifier predicts the action signals given at the plant input and follows terminal voltage or active power deviations. The delayed states of the plant inputs are also given as inputs to the identifier, while the other identifier inputs are: speed, actual terminal voltage and turbine power. The parameters of the identifier’s membership function are updated each time.
Using Takagi-Sueno fuzzy logic-based model, simulation studies are carried out to estimate the next system output in [15]. The center points of membership functions used in the fuzzy logic rules are updated according to the cost function calculated at each step, so that any changes in the system are detected. The cost function is chosen as the difference between the actual system output and the model output which is the deviation in speed. In addition, the learning rate used when calculating the center point of membership functions has also been addressed to be inversely proportional to the cost function. First, in the simulation studies, then in the test devices the input torque is changed and correspondingly the speed change identification is performed. As a result of the studies, model outputs tracking the real system output are obtained both in dynamic response and transient response.

Another approach for system identification problem is using either black-box or white-box modelling of the system. In black-box model, there is no physical information about the system, only the data is present and there is no physical correspondence between the parameters used in the model, parameters are set so that only the observed data is met. The disadvantage of this modeling is the inability to use expert knowledge of the system. In contrast to the black-box modeling, the physical relationship between the parameters is known in advance in the white-box model, which is considered as traditional physical modeling. But in real systems this is usually not the case. Gray-box modeling, a mixture of these two models, may be preferred. In the gray-box modeling system model can be constructed by using partial physical information and data measurement of the system.

In study [16], the Black-box and gray-box methods are used for modelling. Since the parameters may change over time after the first commissioning of power plant, system identification is implemented in order to improve the performance of the excitation system. In black-box method, discrete wavelet transform is used to find the non-linear transfer function that allows the input-output data set to be overlapped. In gray box method, the block-diagrams are generated using both the mathematical equations and data from the manufacturers. Parameter estimation is performed for each subsystem, not for an entire system, by using the input-output data set of the plant and the possible input-output data set of sub-systems. If it is not possible to measure at the sub-systems, the data in the plant documents are used and calculated from the other measured values. Both models gave satisfactory results, but compared to field measurements, the system output predicted by the black-box model gave more consistent results with the field measurements compared to the gray-box model. However, because the physical parameters were directly estimated in the gray-box model, it is much more preferable for the power system studies.
Two different estimations, in frequency domain and time domain, are made for the identification of the reactivity parameters in the nuclear power plant in [17][16]. For each domain, a gray-box model is used so that the model can include the system constraints. If a black-box model (artificial neural network) was used, an estimator with good performance could be obtained only for the region covering the training set, and a successful, unbiased estimate of the operating points where system failures occur could not be made. In the time domain, Kalman filter is used and thus unbiased results are obtained. Heuristic estimation criteria are included in the frequency domain transfer function since estimates of parameters with more complex structures will be more sensitive to modeling errors. The main reason for adding heuristic knowledge and experiences is the complexity of system noise sources in nuclear power plants. Estimates for the three predominant noise sources were made under some assumptions achieved by statistical analysis.

In [18], under normal operating conditions (at different load and excitation levels), the linear model for the armature circuit and winding parameters are estimated from the collected data by giving small disturbing signals for the excitation system. Subsequently, the nonlinear mapping function is used to identify saturable inductances with the steady state data. Using the estimated parameters, the rotor body parameters are estimated by the output error method. It is assumed that the degree of model transfer function in the estimation made in the first step is known. The parameters predicted for different operating points are then mapped into the machine variables that vary according to the operating points, with nonlinear mapping function. The reason for the deviations observed according to the generation values are: magnetic saturation, internal temperature, aging, faults that have just begun to be observed. Data sets that are not used in estimation are used for validation after modeling is finished. Compared to manufacturer parameters, the results are successful. As more data sets become more successful mapping.

Dynamic behaviors of three hydroelectric power plants with different characteristics participating in frequency control are investigated using the gray-box method in study [19]. The system identification is implemented on the linear model, but also some static nonlinearities such as backlash effect and steady state turbine gain are taken into account. The sub-systems modeled with first-order transfer functions are updated after field measurements, adding time delay and backlash effects to the required locations. When the model is constructed, some time constants are calculated using the plant physical properties. The hill-chart data generated during the plant commissioning period are used to compare the estimated steady state turbine gain. As a result of the comparison of the two test results in which the backlash effect is included and not included in the model, the simulated backlash effect gives more similar results with the real system. From these different models (with backlash or not) obtained for each power plant, the best estimators are chosen.
and once again simulated using all measurement values and the rate of validation is obtained. These selected models are evaluated in frequency and time domain one more time. The standard deviation decreases when including the backlash effect to the first and third plant models. Since the backlash effect of the second plant is less it is not resulted with an important improvement.

In study [20], both white-box and gray-box modelling are used. Since the contribution of wind turbines to system inertia is low in synchronous power systems, it is of utmost importance to estimate and monitor the inertia to be provided by wind turbines. Load inertia is estimated by the white-box method and the data related to the moment of a previous outage originating from the fault of the generator is used for estimation. By the gray-box identification method, the inertia of the synchronous generators and therefore the energy stored in the power plants are estimated. The effects of sampling rate, time shifting and averaging techniques on parameter estimation are also investigated. It is aimed to increase the system security by predicting the inertia of both the load side and the generation side. According to the results, the estimation of the load inertia is very important because of the changing nature of the load with time/day, because the minimum amount of energy to be stored is determined according to the obtained estimation information.

In study [21], another method is used for identification of inertia constant. A closed-loop micro perturbation method (MPM) is used to estimate the system equivalent inertia which is sensitive to turbine controllers and the changing operating conditions. In order to estimate the inertia constant, frequency and active power measurements are made using the phasor measurement unit at the transmission line at the point where the plant is connected to the system. In order to be able to perform an identification with sufficient performance, the energy in the disturbance signal which is injected into the system during the identification process must be greater than the energy of the system noise which are the changes in load and operating conditions. For this, either an impulse can be applied as system input signal or low level monitoring test can be done. When the impulse is used, the system inertia to be used for control strategies is not preferred because some problems can occur in the estimation process and is not suitable for real-time online identification systems. Because low-level monitoring methods are easily implemented in linear model identification and have very low effects on system safety points.

Phasor measurement units are quite preferable for data collecting step of system identification. In [22], the arx and srivc models of the MATLAB system identification toolbox are used together with the phasor measurement units for recording the system inputs and outputs. The frequency, current and voltage are
measured from the high voltage side of the generator. The active power is measured as an output simultaneously with the frequency measurement. In addition to the models in MATLAB, a model included in the manual of Power Transmission System Planning Software of Siemens is also used. The system transfer function is estimated using these models. From the estimated models, the closest to the real system is the arc model. Since the governor droop parameters are sometimes changed within the scope of ancillary services, simulations are made for different parameters and it is seen that the identification system can follow these changes. The changes in the system can be parameter changes or non-linearity in turbine characteristics.

In some cases, in order to meet the desired responses, system identification is applied for updating the controller parameters as shown in Figure 4. In study [23] the controllers of speed governor and automatic voltage regulator are upgraded by system identification. If the position of the open loop poles in the s-plane is fixed or varies in a narrow band, the poles of the closed-loop system can be placed at the points to provide the desired response from the system. If the open-loop poles of the system are far away from the closed-loop poles of the desired response, the results may not show the expected performance (damping coefficient, rise time, etc.), as the controller output signal may be saturated at the limit values. Therefore, instead of the direct pole placement method, the pole shifting method is used to prevent the controller from saturation.

![Figure 4 System identification to update controller [23]](image)

In [24] online identification is made for parameters of permanent-magnet synchronous motor (stator resistance, d and q axis inductances). These parameters are used to estimate the rotor position and the sensorless control is done through this estimation. Because the motor position and speed are not used in the identification process, the identified parameters are not affected by the estimation error. Online
identification does not require prior measurement. The observer does not directly estimate the motor position and speed, the extended electromotive force is estimated and then the position is calculated using this estimation value.

4. Conclusion

When the studies examined within the scope of literature survey are compared, it is seen that when the system parameters change, conventional controller parameters must be readjusted.

The methods such as genetic algorithm, neural networks, fuzzy logic are very successful at the point of incorporating nonlinear characteristics into the model, and they have shown an adaptive structure against changing system parameters and have achieved more successful results than conventional controllers.

In addition, systems with artificial neural networks need a very high number of quality training data, but they are still preferred because of the reduced number of training sessions while identifying the system offline.

When the optimization methods are compared, parameter updates of GSA have more accurate results than GA and PSO.

It is observed that the estimations of ANFIS model are more successful than the NNARX model.

When the black-box modeling is used more consistent results are obtained by the field measurements, on the other hand the gray-box model is more preferred in power system studies because the physical parameters are estimated directly by the gray-box.

References


Current Address: Derya OZKAYA; MRC Energy Institute TUBITAK, Ankara, TURKEY
E-mail: derya.ozkaya@tubitak.gov.tr
ORCID: https://orcid.org/0000-0003-3331-4826
Current Address: Ilhan KOSALAY; Elect. and Electr. Eng. Department Ankara University Ankara TURKEY
E-mail: ilhan.kosalay@ankara.edu.tr
ORCID: https://orcid.org/0000-0001-9231-416X