Ph.D. Dissertation

Program: Sistemas de Energía Eléctrica

Application of Kalman filter – based estimation techniques to electric power systems

Author: Miguel Ángel González Cagigal

Advisors: José Antonio Rosendo Macías

Antonio Gómez Expósito

Department of Electrical Engineering

Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Seville, 2020

UNIVERSIDAD DE MÁLAGA

UPC

UPV EHU

Ph. D. Dissertation
Program: Sistemas de Energía Eléctrica

Application of Kalman filter – based estimation techniques to electric power systems

Author:

Miguel Ángel González Cagigal

Advisors:

José Antonio Rosendo Macías

Antonio Gómez Expósito

Department of Electrical Engineering

Escuela Técnica Superior de Ingeniería

Universidad de Sevilla

Seville, 2020

Ph. D. dissertation: Application of Kalman filter – based estimation techniques to electric power systems

Author: Miguel Ángel González Cagigal

Advisors: José Antonio Rosendo Macías y Antonio Gómez Expósito

The thesis committee, composed by the following members:

Agree to the calification:

Seville, 2020

The commitee Secretary.

A mi familia

A mis maestros

Acknowledgement

A mis padres, por facilitarme estudiar lo que me gusta; a mis hermanos, amigos y compañeros por apoyarme y a mis directores por transmitirme y ayudarme a adquirir los conocimientos que he podido aplicar en este trabajo.

Dado que este es el primer paso en mi camino, espero que me sigáis apoyando en los que quedan.

El autor de este trabajo quiere agradecer al Ministerio de Educación y Formación Profesional la financiación recibida a través de la beca FPU17/06380.

Abstract

This thesis presents several applications of dynamic state estimators based on Kalman filtering to different fields of the electric power systems.

First, a parameter estimation technique is proposed, applied to a generation set composed by the synchronous machine along with the frequency regulation (speed governor) and the voltage controllers (automatic voltage regulator and power system stabilizer). The proposed method is based on a formulation of the unscented Kalman filter, being this study the first attempt, to the authors' knowledge, to include the full generation set in the estimator model, with the corresponding state variables and parameters, using just external measurements taken at the generator terminal bus.

A similar estimation technique, using the cubature Kalman filter, is implemented subsequently for a joint estimation of the dynamic state and the model parameters of a variable-speed wind turbine with permanent magnet synchronous generator and back-to-back voltage source converter. In this case, the major contribution consists of the inclusion of the control parameters in the state vector to be estimated.

Finally, three Kalman filter formulations (unscented Kalman filter, cubature Kalman filter and ensemble Kalman filter) are implemented to address the problem of identifying the electrical phase of single-phase consumers in distribution grids, using for this purpose hourly energy measurements exclusively. The accuracy and robustness of these estimators are compared in different case studies with variations in the number of loads and errors in the measurements and the considered model.

Contents

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

This introductory chapter is aimed to establish the context embracing this doctoral thesis, which is split into three parts subsequently described. Along with the technical framework, a thorough revision of the state of the art is presented regarding the considered dynamic state estimator applications, allowing to assess the contributions of this study.

1.1. Parameter estimation in electric power plants. Context and motivation

Electric power systems (EPS) are currently experiencing a remarkable period of change, given the inclusion of new technologies based on power electronics, and the consolidation of the renewable sources of energy, which have modified the existing paradigm in the conventional energy production system.

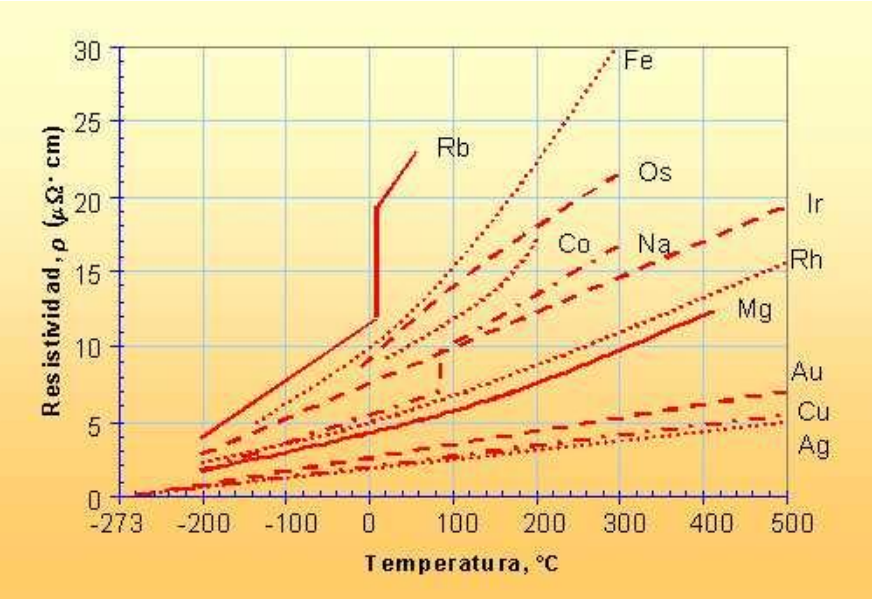
The latter aspects make more difficult the operation and control in electrical networks, increasing the vulnerability of the system towards the possible disturbances that might appear. Moreover, the mentioned tasks of operation and control require a deep knowledge of the dynamics associated to each of the elements comprising the grid, so that the response of this components can be predicted, and precautionary measures can be taken.

The transient behavior is determined by a set of differential-algebraic equations, establishing the dynamic models for the different units. These equations involve a collection of characteristic parameters, for which an adequate knowledge is essential in order to correctly identify the state of the system and take the

corresponding actions.

The values of these parameters are sometimes provided by the manufacturer of the equipment, while in several cases this information is not available. Regardless the situation, a usual practice is assuming the parameters to be invariable with respect to the operating conditions. A particular example might be the widely studied dependence of some parameters, as the electrical resistivity, with the temperature (Figure 1-1, [1]).

Figure 1-1. Variation of the resistivity with the temperature for different materials



Besides the temperature, other factors might affect significantly the values of the parameters involved in the dynamic models of the components of an EPS, such as:

- The operating point of the elements, determined by the voltage and the exchanged power in a certain moment.

- The degradation caused by unusual working conditions or aging.

The variation of the parameters due to the latter factors is not studied as deeply as that caused by the temperature, given its potential complexity, and resulting in the previously mentioned assumption of neglecting those changes. Finally, the absence of information provided by the manufacturer may also lead to a lack of knowledge on their equipment parameters.

In light of the above mentioned aspects, it is clear the usefulness of a method that allows to establish a reliable estimation of the parameters involved in the dynamics of the components of an EPS, given the external factors previously highlighted. This is the main goal of parts 1 and 2 of this thesis, which will approach the described issues with a parameter estimation technique for two particular elements of the system: the synchronous generator and the variable speed-wind turbine.

1.2. Electrical phase assignment in distribution grids

Focusing on distribution networks, operation and control tasks are essential in order to guarantee the good quality of the service provided to the consumers. In this context, grid operators require an unambiguous knowledge of the load distribution in each of the three phases of the system, so that this information allows to state a correct load equilibrium in those phases. The importance of this problem is even more remarkable when considering renewable sources of energy connected to the distribution grids, since the phase identification helps to establish a better generation-consumption balance in each phase.

In this regard, despite the efforts undertaken by distribution companies, they frequently lack enough information about the phase connection of their single-phase customers, owing, for instance, to network reconfiguration after faults, phase switching derived from improper maintenance, or inaccurate recording of the true load-to-phase connectivity. In these circumstances, a method must be developed to estimate as accurately as possible the actual phase to which a customer is connected in LV feeders, which is known as the customer-phase identification (CPI) problem.

The CPI problem has been approached in several ways by previous works. In [2], a signal processing perspective is applied to voltage observations, which are also

used in [3], for a correlation-based methodology. Two techniques are highlighted in this section: a method proposed in [4], using Least Absolute Shrinkage and Selection Operator (LASSO) and a novel approach in [5] for phase idenfication using graph theory and principal component analysis (PCA). The latter methodologies will be included in a comparative study with the estimation technique proposed in this thesis for the CPI problem.

1.3. Dynamic state estimators. Kalman filter

A dynamic state estimator (DSE) is aimed to identify the state of the system which is unobservable through direct measurement. This state, defining the system under study, is considered to be time-variant, unlike the traditional static estimators.

An example of DSE might be the Luenberger observer, [6], based on the definition of an equivalent system for which it is possible the direct observation of the state variables whose knowledge is desired.

However, this work is focused on other DSE, the so-called Kalman filter (KF). Chapter 2 includes a brief historical context of this estimator and the filter equations, while the particular applications considered will be described in chapters 3-5.

The KF has been widely used in several works related to EPS. A major división can be made regarding those studies related to large-scale estimation (the reader may refer to [7–9]), and those works focused on the characterization of single elements of the system, being the latter field of investigation closer to the objectives of this document.

The original formulation of the KF, [10], assumes a linear dynamic of the system, being this condition quite uncommon in the equations of the elements in EPS. To approach this issue, different formulations have arised to deal with non-linear dynamics. Among them, the extended KF (EKF), is based on the calculation of the jacobian in order to linearize the state equations. This scheme has been widely used in several works related to EPS, as that in [11], where this estimator is used for state estimation with unkown inputs. Regarding synchronous machines, [12] makes use of the EKF to estimate the initial rotor position of permanent magnets synchronous machines (PMSMs), while the flux losses are included in the estimation process of [13], using vector control, and in [14], where demagnetization problems are

considered.

A study in [15] concludes that the EKF and another non-linear formulation, the unscented KF (UKF), are the most computationally efficient when the synchronous generator dynamic is considered. The latter formulation has been used in [16] for state estimation in single machines and in [17–19] for multimachine systems.

In the particular case of synchronous generators, the joint estimation of state and parameters has been approached in several works using the UKF (see [20] and [21]). However, the application of those estimation techniques in real systems is hindered by the use of internal measurements.

To avoid these difficulties, different strategies have been developed, as that in [22], using a set of measurements obtained from phasor measurement units (PMUs). The method proposed in [23] broadens the problem by including in the model the parameters of the speed governor (SG) and the automatic voltage regulator (AVR), studying the effect of a variable error (up to ±30%) in the initialization of those parameters.

The power system stabilizer (PSS) is introduced in the KF model in [24], although the parameters of this regulator are taken as known inputs of the system. This assumption might not be accurate enough given the variability of these parameters and the possible lack of information.

With the above-mentioned considerations, and to the author's knowledge, the publication included in annex II.1 is the first attempt to consider a fully regulated generation set, including the synchronous machine itself along with the whole set of regulators (SG, AVR and PSS). Both state and parameters from these elements are jointly estimated using UKF and measurements obtained at the genartor terminal bus, which could be provided by a PMU in real applications.

Furthermore, a recent formulation of the KF, the so-called cubature KF (CKF) has proven a correct performance when it is used for the dynamic state estimation in synchronous generators, [25]. To see the theoretical aspects of this scheme, the reader may refer to [26], where some limitations of other KF formulations, such as UKF, which are not suffered by the CKF, are highlighted. Joint state and parameter estimation using CKF is studied in [27], applied to a vehicle model, and in [28], to permanent magnet synchronous motors.

In the second part of this thesis, the CKF is considered for the state and parameter

estimation of a variable-speed wind turbine, coupled to a PMSM, and connected to the external network through a back-to-back voltage source converter (VSC).

In a considerable number of the works previously mentioned, the KF-based DSE is used for the estimation of the electrical parameters of the system. The inclusion of these quasi-invariant elements in the state vector has led to the consideration of this methodology in a static problem as the CPI, being this particular approach described in the third part of the thesis.

1.4. Kalman filter applied to the monitorization of a virus spread

Related to the undesirable situation derived from the appearance of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), cause of the coronavirus disease 2019 (Covid-19), which has severely affected the health of millions of people and the global economy in 2020, the Ph.D. student and his thesis advisors developed a KF-based estimation technique for the monitoring and tracking of the evolution of this virus.

Although this study is not related to electric power systems and therefore it is not incorporated into the body of this document, it has been considered appropriate to include a description of the proposed methodology in the annex I, jointly with the most relevant results obtained. Additionally, annex II.4 incorporates the latest version of the paper submitted to "IEEE Journal of Biomedical and Health Informatics", and whose current status is minor revision.

1.5. Document structure

The structure of this document responds to the regulation considered by the University of Seville for Ph.D. thesis presentations in the form of a compendium of publications, where only the main contributions, results and conclusions are remarked. The reader may refer to the publications included in annex II to see the details of each of the three parts of the doctoral thesis, namely:

- Parameter estimation in fully regulated synchronous generators using UKF.
- Parameter estimation in variable-speed wind turbines using CKF.
- Application of KF-based estimation techniques to the identification of the phase connection in distribution grids.

For each of these fields, chapters 3-5 of this document describe:

- The particular implementation of the KF-based estimator.
- The estimation results for the different scenarios considered.

The conclusions derived from the presented results are included in chapter 6, jointly with the possible future lines of investigation in each study field. Finally, chapter 6 also highlights the most relevant publications of the Ph.D. student with his thesis advisors. The three published papers related to the different parts of the thesis can also be found in the annexes II.1-II.3 to this document.

1.6. Financing

This Ph.D. thesis has been developed with the economic support of:

- Ministry of Education and Professional Training of Spain (grant FPU17/06380).
- Project: "Pastora: Análisis preventivo de redes inteligentes en tiempo real e integración de recursos renovables" (PI-1897/12/2019).
- Competitiveness grants ENE2015-69597 and PCIN-2015-043.

2. KALMAN FILTER CONTEXT

The objective of this chapter is to briefly introduce the different formulations of the KF considered in this thesis, namely the UKF, the CKF and the EnKF.

First, the context of the original KF estimator is included, while the particular equations of each formulation are subsequently analyzed.

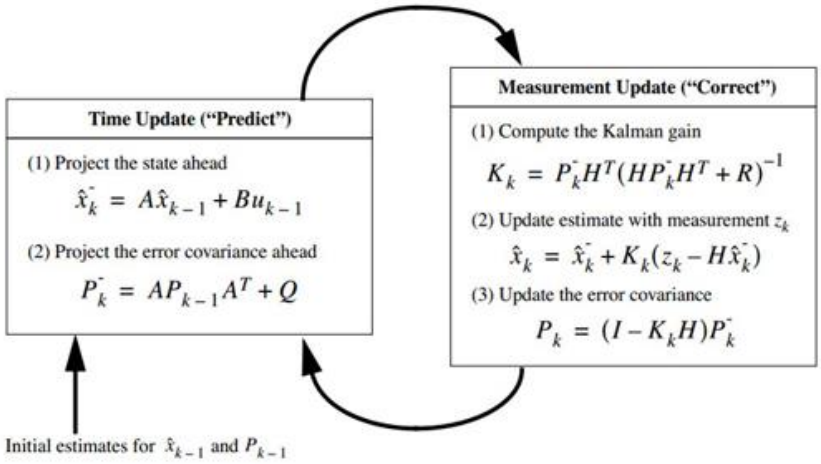
2.1. Context

From the information in [29], the Kalman filter was developed in 1960 by Rudolf E. Kalman as a method to identify the non-observable state of a system whose dynamic response had to be determined by a set of linear equations, although this idea would be extended to a wider range of situations.

The major advantage of the KF with respect to previous techniques, such as Luenberger observer, is that the system under study can be affected by an additive Gaussian noise, being this characteristic of great importance in the applications presented in this thesis, since the measurements obtained from real systems present random errors, whose probability distribution functions are usually modelled as Gaussian.

Another distinctive aspect of this estimator is its defining recursive process, which is composed by prediction and correction stages (Figure 2-1). At each iteration, a *Kalman gain* is calculated to optimally correct the *a priori* estimation of the system state with the available measurements.

Figure 2-1. Diagram of the recursive estimation process of the KF



At the beginning, the interest in the KF was based on aeronautic applications, being really useful in craft positioning and control in the Apollo program. From this point, its implementation in vehicle control, guide and navigation is remarkable.

Nevertheless, the use of KF-based techniques has been diversified to fields as heterogeneous as signal processing and econometric studies (even to modeling the spread of the Covid-19 disease).

2.2. Original equations

In the original implementation of the KF, the system dynamic is modelled by the following discrete state and measurement linear equations:

$$x_k = A_{k-1} \cdot x_{k-1} + w_k \tag{1}$$

$$z_k = H_k \cdot x_k + v_k \tag{2}$$

where x_k is the state vector and z_k the available measurements at instant k, being A_k and H_k the model transition and measurement matrices. Regarding w_k and v_k, they are the process and measurement Gaussian noises, with zero mean and covariance matrices Q_k and R_k respectively.

At instant k, the KF uses the information of the previous iteration $(k-1)$, determined by the estimated state vector, $\hat{x}_{k-1}$, and the covariance matrix of the estimation error, P_{k-1}. The *a priori* estimation of the state vector is calculated as:

$$\hat{x}_k^- = A_{k-1} \cdot \hat{x}_{k-1} \tag{3}$$

with the covariance matrix of the estimation error,

$$P_k^- = A_{k-1} P_{k-1} A_{k-1}^t + Q_{k-1} \tag{4}$$

For the correction stage, the Kalman gain, K_k is calculated as follows:

$$K_k = P_k^- H_k^t [H_k P_k^- H_k^t + R_k]^{-1} \tag{5}$$

Finally, the corrected *a posteriori* prediction is obtained as:

$$\hat{x}_k = \hat{x}_k^- + K_k \cdot (z_k - H_k \cdot \hat{x}_k^-) \tag{6}$$

The continuous-time counterpart of the original KF formulation, the so-called Kalman-Bucy filter [30], is used for continuous systems represented by the following state and measurement equations:

$$\dot{x}(t) = A \cdot x(t) + B \cdot u(t) + w(t) \tag{7}$$

$$z(t) = C \cdot x(t) + v(t) \tag{8}$$

being $u(t)$ the system input. However, since the measurements are collected at discrete time instants in real applications of the estimator, a continuous-time, discrete-measurement framework is defined with the following equations:

$$\dot{x}(t) = A \cdot x(t) + B \cdot u(t) + w(t) \tag{9}$$

$$z_k = C_k \cdot x_k + v_k \tag{10}$$

Where a discretization of the state equation (9) is used to calculate the *a priori* estimation of the state vector at instant k.

Finally, in most applications related to EPS, and specifically those presented in this thesis, the KF is applied to discrete-time, discrete-measurement nonlinear systems. In the discrete-time framework, the associated equations may be expressed as,

$$x_k = f(x_{k-1}, u_{k-1}) + w_k \tag{11}$$

$$z_k = g(x_k, u_k) + v_k \tag{12}$$

where $f(\cdot)$ and $g(\cdot)$ are nonlinear functions.

In order to approach the nonlinearities in the system equations, several formulations of the KF have been developed, from which three of them are presented in the next sections, with their corresponding equations.

2.3. Unscented Kalman filter

At instant k, a cloud of $2L + 1$ vectors, the so-called σ-points, is obtained from the previously estimated state vector, $\hat{x}_{k-1}$ (dimension L), and the covariance matrix of the state estimation error, P_{k-1}, as follows [31]:

$$X_{k-1}^{0} = \hat{x}_{k-1} \tag{13}$$

$$X_{k-1}^{i} = \hat{x}_{k-1} + [\sqrt{(L+\lambda)P_{k-1}}]_i \qquad i = 1, \dots, L$$

$$X_{k-1}^{i} = \hat{x}_{k-1} + [\sqrt{(L+\lambda)P_{k-1}}]_i \qquad i = 1, \dots, L$$

$[\sqrt{(L+\lambda)P_{k-1}}]_i$ being the i^{th} column of the corresponding matrix, and λ a scaling factor calculated as $\lambda = \alpha^2(L + \kappa) - L$, where α and κ are two filter parameters to be tuned.

These σ-points are evaluated through equation (11), yielding $2L + 1$ vectors, X_k^{i-}, from which the a priori estimations $\hat{x}_k^-$ and P_k^- are obtained:

$$\hat{x}_k^- = \sum_{i=0}^{2L} W_{mi} X_k^{i-} \tag{14}$$

$$P_k^- = \sum_{i=0}^{2L} W_{ci}(X_k^{i-} - \hat{X}_k^-)(X_k^{i-} - \hat{X}_k^-)^T + Q_k \tag{15}$$

where the weighting vectors W_m and W_c are calculated from

$$W_{m0} = \frac{\lambda}{L+\lambda} \tag{16}$$

$$W_{c0} = \frac{\lambda}{L+\lambda} + 1 - \alpha^2 + \beta$$

$$W_{mi} = \frac{1}{2(L+\lambda)} \qquad i = 1, \dots, 2L$$

β being another tunable parameter.

On the basis of the a priori information, the correction stage starts with the calculation of a new cloud of vectors, X_k^i, which are evaluated with the measurement function $g(\cdot)$ in equation (12), and weighted with the vectors W_m, yielding

$$\gamma_k^{i-} = g\left(X_k^i, u_k\right) \qquad \text{i} = 0, \dots, 2\text{L} \tag{17}$$

$$\hat{z}_k^- = \sum_{i=0}^{2L} W_{mi}\gamma_k^{i-} \tag{18}$$

Then, the covariance matrix of the measurement estimation error, P_{zk}^-, and the cross-covariance matrix of state and measurements, P_{xzk}^-, are obtained using the vector W_c as follows:

$$P_{zk}^- = \sum_{i=0}^{2L} W_{ci}(\gamma_k^{i-} - \hat{z}_k^-)(\gamma_k^{i-} - \hat{z}_k^-)^T + R_k \tag{19}$$

$$P_{xzk}^- = \sum_{i=0}^{2L} W_{ci}(X_k^{i-} - \hat{x}_k^-)(\gamma_k^{i-} - \hat{z}_k^-)^T \tag{20}$$

The correction stage concludes with the *a posteriori* predictions,

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - \hat{z}_k^-)^T \tag{21}$$

$$P_k = P_k^- - K_k P_{zk}^- K_k^T \tag{22}$$

which are based on the $a\ priori$ predictions at instant k and the so-called Kalman gain, K_k, calculated from

$$K_k = P_{xzk}^- (P_{zk}^-)^{-1} \tag{23}$$

2.4. Cubature Kalman filter

The CKF formulation, more recent than the UKF, was developed in 2010 as a technique to estimate the positioning in the navigation field. This estimator uses $2L$ cubature points, which are calculated from the previous information, $\hat{x}_{k-1}$ and P_{k-1}, through the following equations, [32]:

$$S_k^- S_k^{-T} = P_k^- \tag{24}$$

$$x_k^{i-} = S_k^- \xi_i \sqrt{L} + \hat{x}_k^- \qquad i = 1, \ldots, 2L \tag{25}$$

where S is a positive-definite square root matrix of P, and ξ_i is the i^{th} cubature node, obtained as the intersection of the unit sphere and the $\mathbb{R}^{\mathrm{L}}$ axis.

The state function $f(\cdot)$ in equation (11) is evaluated for the set of cubature points, yielding $2L$ vectors X_k^{i-}, from which the *a priori* estimation is computed,

$$\hat{x}_k^- = \frac{1}{2L}\sum_{i=1}^{2L} X_k^{i-} \tag{26}$$

$$P_k^- = \frac{1}{2L}\sum_{i=1}^{2L} x_k^{i-} x_k^{i-T} - \hat{x}_k^- \hat{x}_k^{-T} + Q_k \tag{27}$$

For the correction stage, the covariance matrix is factorized in order to calculate both the matrix S_k^-

$$S_k^- S_k^{-T} = P_k^- \tag{28}$$

and a new set of $2L$ cubature points, X_k^{i-}, at which function $g(\cdot)$ is evaluated to obtain γ_k^{i-}.

Then, the measurement estimation, $\hat{z}_k^-$, its covariance matrix, P_{zk}^-, and the cross-covariance matrix of state and measurement, P_{xzk}^-, are calculated as follows:

$$\hat{z}_k^- = \frac{1}{2L}\sum_{i=1}^{2L} \gamma_k^{i-} \tag{29}$$

$$P_{zk}^- = \frac{1}{2L}\sum_{i=1}^{2L} \gamma_k^{i-} \gamma_k^{i-T} - \hat{z}_k^- \hat{z}_k^{-T} + R_k \tag{30}$$

$$P_{xzk}^- = \frac{1}{2L}\sum_{i=1}^{2L} x_k^{i-} \gamma_k^{i-T} - \hat{x}_k^- \hat{z}_k^{-T} \tag{31}$$

The *a posteriori* predictions of the state vector, $\hat{x}_k$, and the covariance P_k are calculated with the Kalman gain using the same equations (21) and (22) as in the UKF.

2.5. Ensemble Kalman filter

Finally, the EnKF formulation, [33], is a Monte Carlo approximation of the original KF, which has proven accurate enough in high-dimensional state-space problems. The ensemble is represented by a LxN matrix, N being the number of samples considered.

This ensemble is first propagated through the state and measurement functions [34],

$$x_k^{i-} = f\left(x_{k-1}^i, u_{k-1}\right) \tag{32}$$

$$z_k^{i-} = g\left(x_k^i, u_k\right)$$

$$i = 1, \dots, N$$

and then the mean values are calculated:

$$\overline{x_k^-} = \frac{1}{N}\sum_{i=1}^{N} x_k^{i-} \tag{33}$$

$$\overline{z_k^-} = \frac{1}{N}\sum_{i=1}^{N} z_k^{i-} \tag{34}$$

The EnKF correction stage is based on the calculation of the intermediate matrices

$$\overline{P_k}H_k^T = \frac{1}{N}\sum_{i=1}^{N}(x_k^{i-} - \overline{x_k^-})\left(z_k^{i-} - \overline{z_k^-}\right)^T \tag{35}$$

$$H_k\overline{P_k}H_k^T = \frac{1}{N}\sum_{i=1}^{N}\left(z_k^{i-} - \overline{z_k^-}\right)\left(z_k^{i-} - \overline{z_k^-}\right)^T \tag{36}$$

which allow the Kalman gain and the updated values of each sample in the ensemble to be obtained:

$$K_k = \overline{P_k}H_k^T(H_k\overline{P_k}H_k^T + R)^{-1} \tag{37}$$

$$x_k^i = x_k^{i-} + K_k\left(z_k - z_k^{i-}\right) \qquad i = 1, \dots, N \tag{38}$$

Finally, the corrected covariance matrix, P_k, is calculated as follows:

$$\overline{x_k} = \frac{1}{N}\sum_{i=1}^{N}x_k^i \tag{39}$$

$$P_k = \frac{1}{N}\sum_{i=1}^{N}(x_k^{i-} - \overline{x_k})(x_k^{i-} - \overline{x_k})^T \tag{40}$$

3. APPLICATION 1: FULLY REGULATED SYNCHRONOUS GENERATOR

This section presents the application of a KF-based technique to the estimation of the parameters involved in the dynamic equations of a synchronous generator and its regulators (SG, AVR and PSS).

First, the proposed implementation of the UKF formulation is described, as well as the tuning of the estimator. The considered case studies are presented subsequently, showing the corresponding estimation results.

3.1. Kalman filter implementation

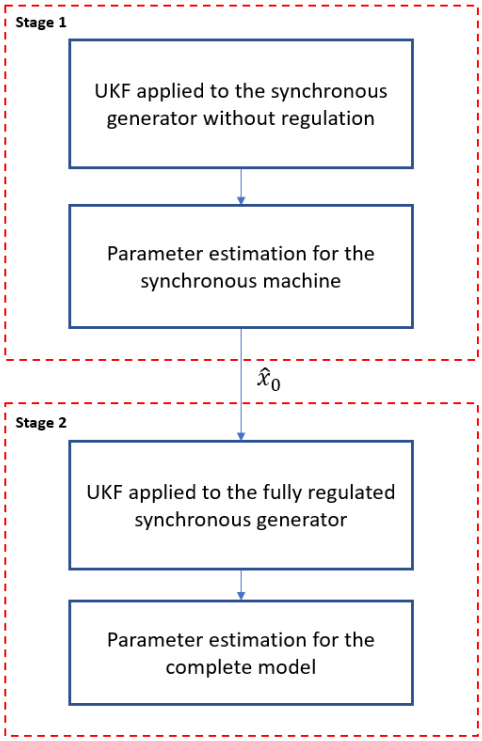
The UKF scheme is used in this part of the thesis for the joint estimation of the state and parameters of a fully regulated synchronous generator. The algorithm of this formulation has been coded using MATLAB.

First attempts with a vanilla implementation of the joint estimator, directly using the whole set of parameters in the system equations, led to unacceptable convergence behavior. Therefore, as proposed in [35] for a simpler case with single-axis generator and only frequency regulation, a new modified parameter vector has been considered using different combinations with their original values.

Finally, all the convergence problems were solved using a novel two-stage formulation, as illustrated in Figure 3-1, so that the modified parameters from the synchronous machine are identified in the first stage. These results are incorporated

to the full model in the second stage for the estimation of the parameters from the machine regulators.

Figure 3-1. Diagram of the two-stage implementation for the UKF estimator



The UKF tuning is determined by the following points:

- Values of the UKF tuning parameters.
- Definition of the system inputs and measurements.
- Initial estimation.

- Process and measurement noise covariance matrices.

First, the UKF algorithm involves a set of tuning parameters, α, β and κ. As proposed in [36], where the influence of these values in the performance of the estimator is analyzed and accordingly the values $\alpha = 10^{-4}$ y $\beta = 2$ have been taken in this work. Regarding the optimal tuning for κ, it is obtained using the following expression:

$$\kappa + L = 3$$

being L the size of the state vector.

With respect to the measurements taken from the system, five magnitudes can be easily measured at the generator terminal bus, namely:

- Voltage magnitude, V
- Voltage angle, θ_v
- Angular speed, ω
- Active power exchanged, p_e
- Current magnitude, I

Those magnitudes are divided, as proposed in [37], into inputs, $u(t)$, and measurements, $z(t)$, as it is summarized in Table 3-1.

Table 3-1. Distribution of inputs and measurements for the synchronous generator

Input vector, $u(t)$	Measurement vector, $z(t)$
V	I
θ_V	p_e
	ω

The initial estimation of the UKF implementation is determined by the initial value of the state variables and the model parameters (i.e. the initial estimation augmented state vector, $\hat{x}_0$), jointly with covariance matrix of the initial estimation error, P_0. A steady-state calculation has been considered to obtain the initial values of the state variables, while the parameters have been randomly initialized in the range $\pm 20\%$ to $\pm 40\%$ of their real values in the simulation. In the second stage of the proposed methodology, the generator parameters are initialized with their estimated values from the first stage. Regarding the covariance matrix P_0, it has been defined as a diagonal matrix,

$$P_0 = diag\left(\left[P_{x0}^T, P_{\psi 0}^T\right]\right)$$

where the vector P_{x0} corresponds to the state and $P_{\psi 0}$ to the modified model parameters. The reader may refer to the paper included in the annex II.1 to consult the particular values taken in both stages of the proposed estimation technique.

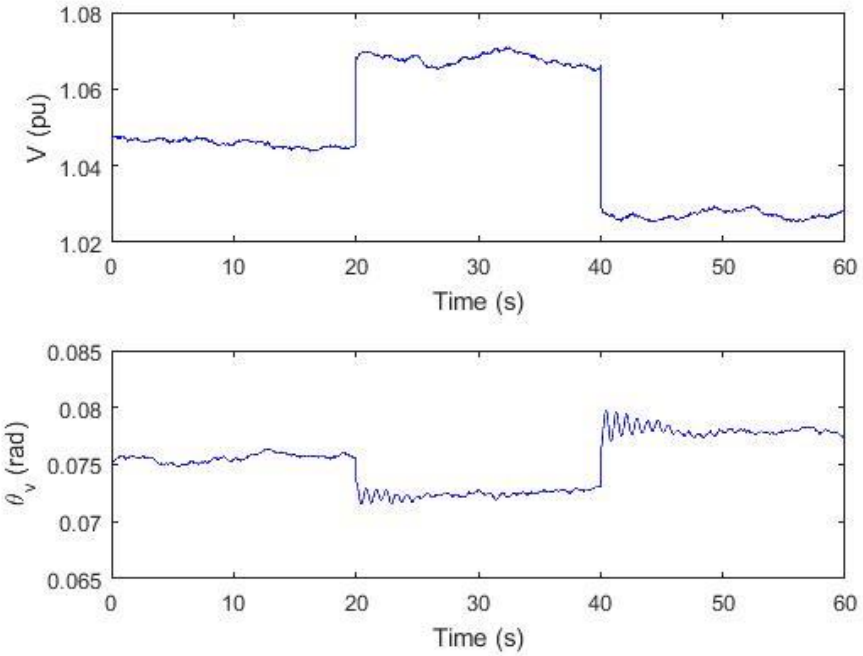
Finally, the covariance matrices of the process and measurement noises, respectively Q and R, are both considered in this work as diagonal matrices, with $Q_{ii} = 10^{-8}$ *y* $R_{ii} = 10^{-4}$.

3.2. Results

First, the estimation results are shown for the first stage of the proposed technique, including the state variables and the parameters of two-axis model considered for the synchronous generator, [38].

In the base case, the generator is connected to an infinite busbar through a tie-line with known impedance. The voltage at the infinite busbar is modelled with magnitude and angle evolving smoothly, as Gaussian random walks, with standard deviations $R_w = 10^{-5}$. Two abrupt voltaje magnitude steps (3% upwards and 5% downwards) are applied, as represented in Figure 3-2.

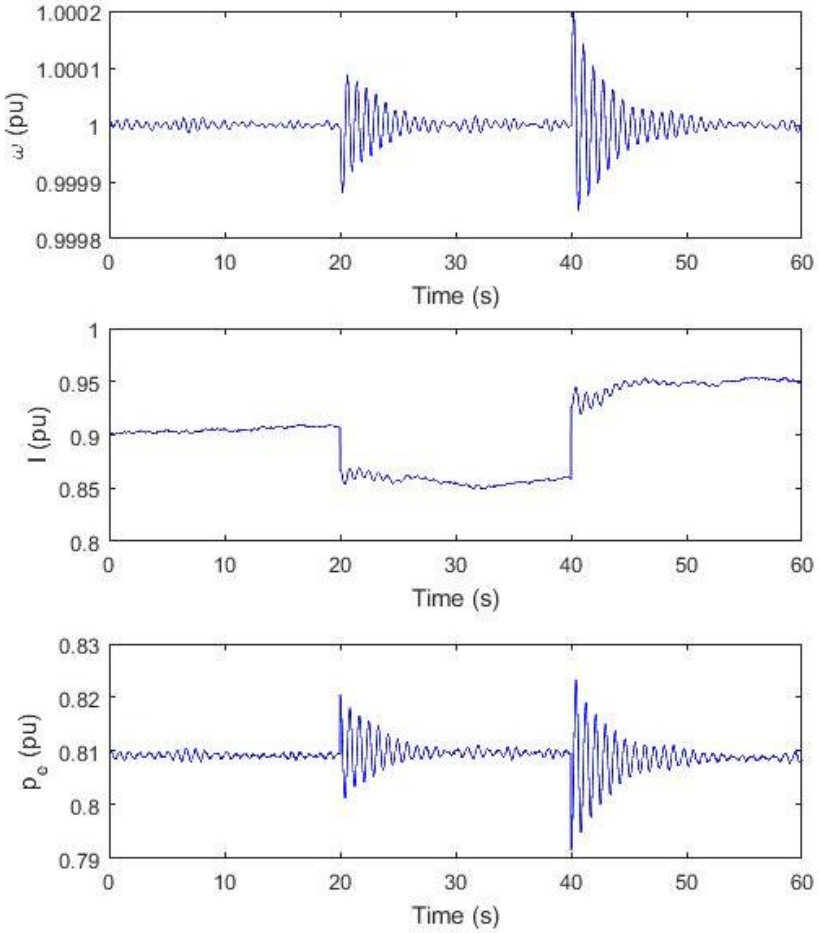
Figure 3-2. Voltage magnitude and angle at the generator bus



Regarding the measurements used in the correction stage of the UKF algorithm, Figure 3-3 shows the evolution of those signals with the above-mentioned abrupt changes in the voltage magnitude.

Please note that, as remarked in the introduction of this document, only external measurements are considered, so that they can be obtained using measurement equipment already installed in the system, as might be the PMUs, given that the sampling frequency has been taken as 100 Hz, a typical value for those elements.

Figure 3-3. Considered measurements for the estimation process



As it can be noticed in the representations, the total simulation time is set to 60 s, including the voltage steps at 20 and 40 s.

3.2.1. First estimation stage

Using the previously mentioned signals as inputs and measurements, together with the described UKF tuning, the joint state-parameter estimation was carried out for the synchronous generator.

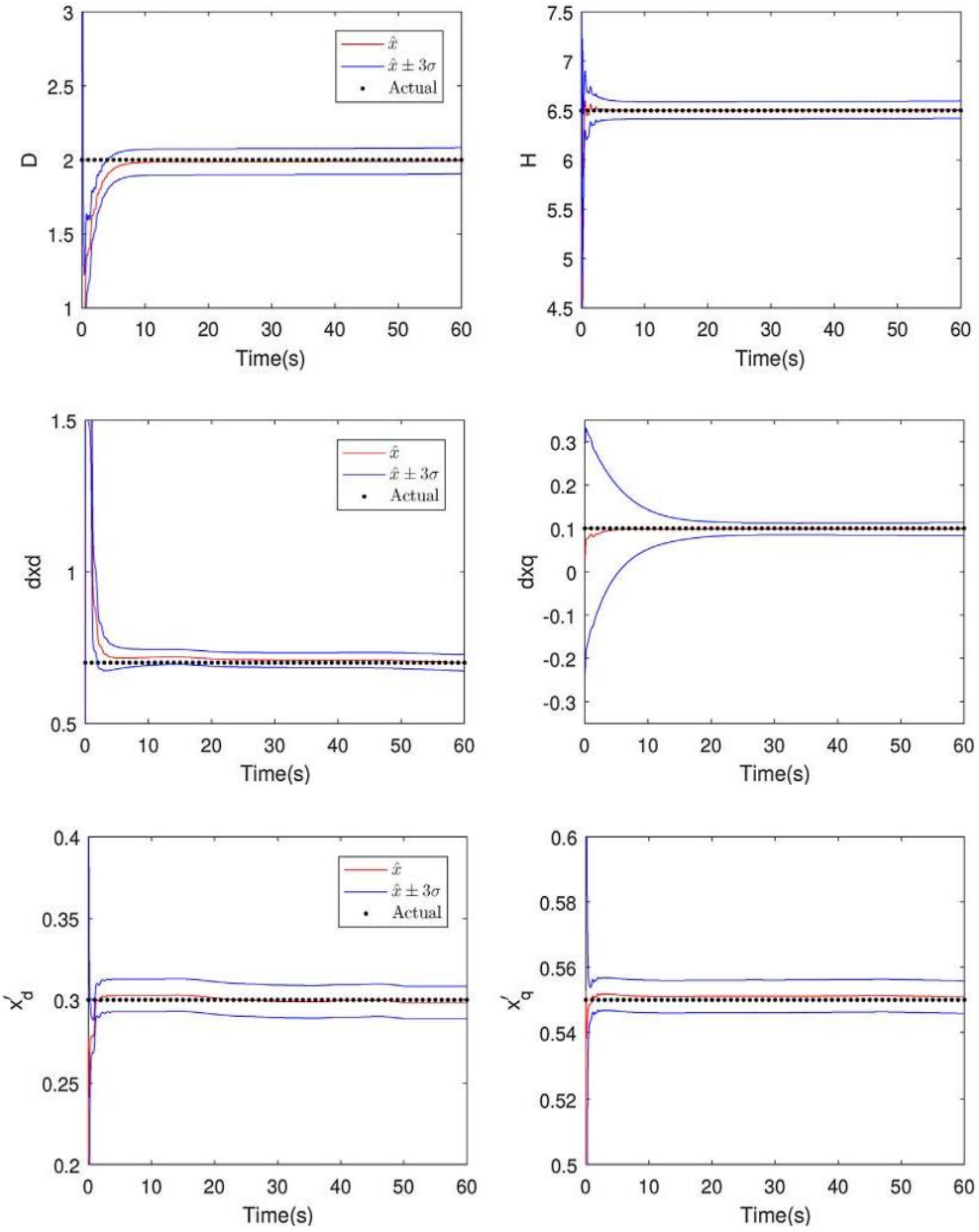
The good performance of the estimator in the first stage is proven with the subsequently presented results for the modified parameters of the synchronous machine. However, the convergence of the KF-based technique is not only defined by the stable value achieved by a certain parameter, but also by the covariance of the estimation error, given by the corresponding component P_{ii} for the considered variable. An accurate estimated value with high covariance results in a low reliability in the estimation, whereas a value of P_{ii} close to the process noise covariance, Q_{ii}, would imply a correct performance of the proposed estimator.

For this reason, the evolution of the estimated values of the parameters is represented in Figure 3-4 jointly with a deviation $\hat{x}_\mathrm{i} \pm 3\sigma_i$ being $\sigma_i = \sqrt{P_{ii}}$. In such graphics, it can be observed that the estimated values are close to those used in the simulation and the corresponding deviations tend to reduce remarkably. These aspects give evidence of the accuracy and reliability of the first stage of the estimation process.

With the obtained values for the modified parameters it is possible to easily calculate the estimation of the original parameters as they are used in the model equations. These results, with the corresponding relative errors, are summarized in Table 3-2.

As it can be noted, the results in the first stage, with a maximum error lower than 2.5%, are accurate enough to proceed with the second stage of the estimation technique proposed in this thesis.

Figure 3-4. Estimation of the modified parameters in the first stage



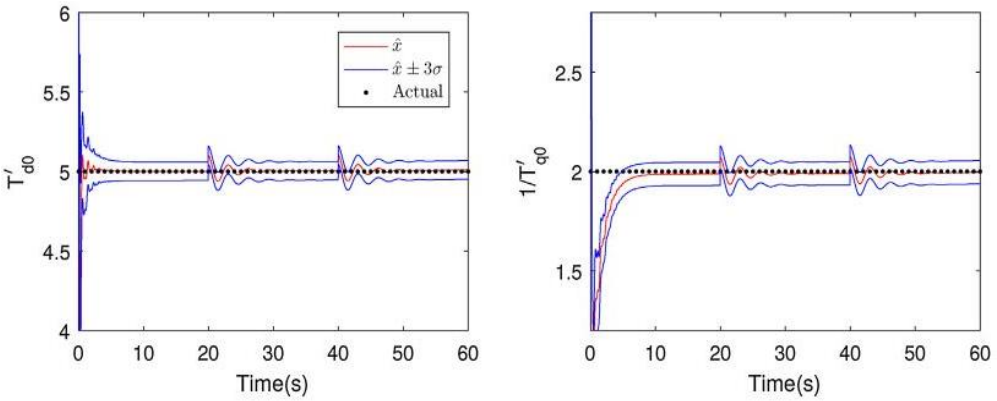


Table 3-2. Relative errors for the parameters of the synchronous generator

Parameter	Estimated value	Relative error (%)
x_d	0.987	1.30
$x_d^{'}$	0.298	0.67
x_q	0.658	1.28
$x_q^{'}$	0.559	1.43
H	6.341	2.45
D	1.999	0.05
T'_{d0}	0.502	0.40
$T_{q0}^{'}$	5.117	2.34

3.2.2. Second estimation stage

In the second stage of the estimation process, the modified parameters of the three machine regulators considered (SG, AVR and PSS) are included in the augmented state vector of the KF, to be estimated. In this case, the information obtained from the first stage is used, given that the initial values of the parameters of the synchronous machine are taken as their previous estimation (Table 3-2), with a high confidence in the components of $\hat{x}_0$, meaning low values for the corresponding elements of the matrix P_0.

The system inputs and the measurements considered for the UKF implementation remain the same as those used in the first stage, including the considered variations in the terminal voltage magnitude.

For the representation of the modified parameters, the same considerations as in the first stage have been taken, showing the estimated value jointly with a $\pm 3\sigma_i$ deviation. Figure 3-5 illustrates the evolution of the modified parameters of the considered regulators. The graphics corresponding to the estimation of the machine parameters are not included, since they suffer no significant variation with respect to the initial estimation and, therefore, no additional information is provided.

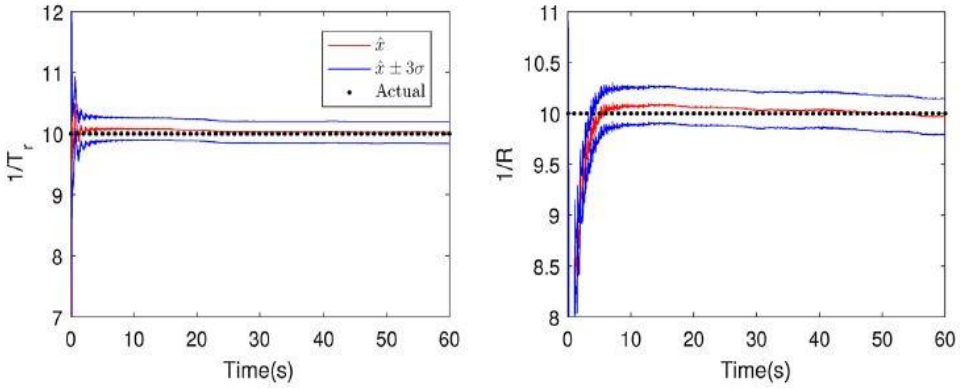
In this figure it can be observed, as with the generator modified parameters, that those of the three regulators reach stable values close to the simulated ones, with a reduced covariance, giving evidence of the good performance of the proposed estimation methodology.

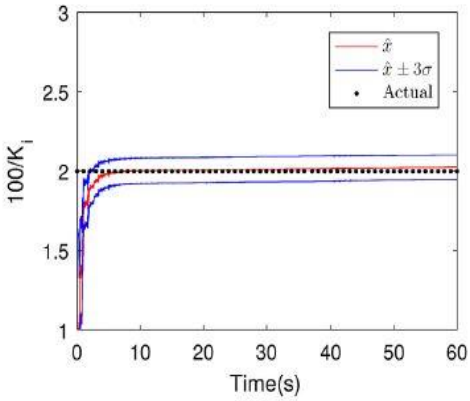
Finally, Table 3-3 summarizes the estimated values of the original parameters of the regulators, with their corresponding relative errors. The fact that the maximum error is only slightly higher than 3% also shows that the accuracy of the estimation technique is more than acceptable.

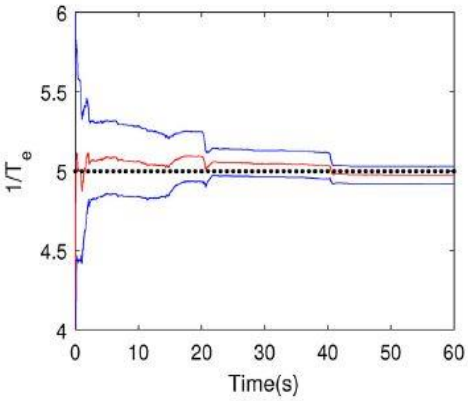
Table 3-3. Relative error in the estimation of the parameters of the machine regulators

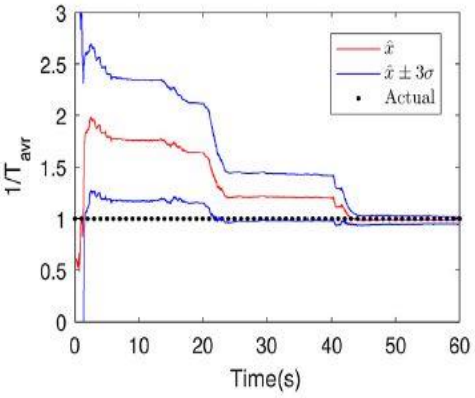
Parámetro	Valor estimado	Error relativo (%)
R	0.1032	3.20
T_r	0.1006	0.64
K_i	49.75	0.498
T_{avr}	0.988	1.186
T_e	0.2002	0.101
K_0	2.479	0.84
K_w	30.78	2.596
T_w	10.135	1.35

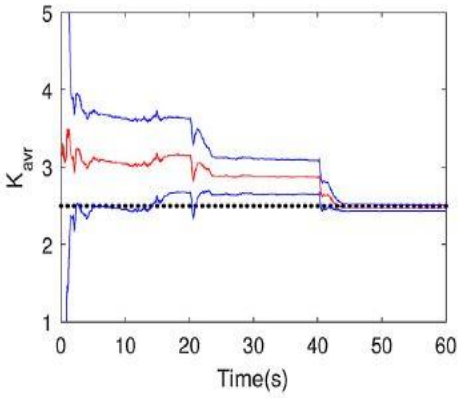
Figure 3-5. Estimation of the modified parameters in the second stage

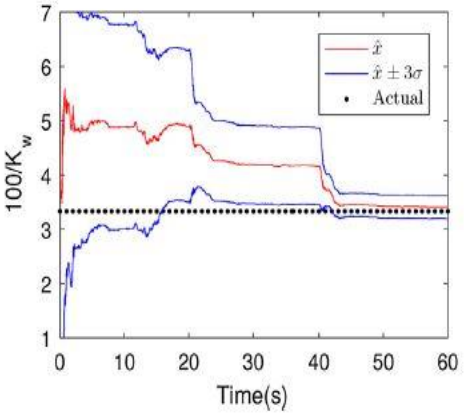


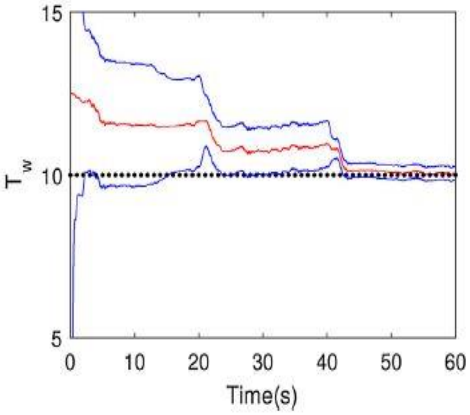

$\hat{x}$
$\hat{x} \pm 3\sigma$
Actual
100/K_i
Time(s)


1/T_e
Time(s)


$\hat{x}$
$\hat{x} \pm 3\sigma$
Actual
1/T_avr
Time(s)


K_avr
Time(s)


$\hat{x}$
$\hat{x} \pm 3\sigma$
Actual
100/K_w
Time(s)


T_w
Time(s)

3.2.3. Variations of the base case

To asses the robustness of the UKF-based estimator proposed in this thesis, three additional scenarios have been considered with different values in the standard deviation of the random walk, R_w, defining the magnitude and angle of the external voltage.

The comparison is based on the estimated values of the parameters and their corresponding relative errors. Tables 3-4 to 3-6 include these estimated values for $R_w = 10^{-6}$, 10^{-4} and 10^{-3} respectively.

To make the analysis easier, Table 3-7 summarizes the mean and maximum relative errors in each scenario, including the base case.

Table 3-4. Estimated values of the parameters and relative errors for $R_w = 10^{-6}$

Parameter	Estimated value	Relative error (%)
x_d	0.985	1.50
$x_d^{'}$	0.298	0.67
x_q	0.656	0.92
$x_q^{'}$	0.562	2.18
H	6.345	2.38
D	1.994	0.03
T'_{d0}	0.506	1.20
$T^{'}_{q0}$	5.116	2.32
R	0.1031	3.05
T_r	0.1006	0.60
K_i	49.75	0.50
T_{avr}	0.987	1.31
T_e	0.198	1.02
K_0	2.479	0.84
K_w	30.65	2.17
T_w	10.115	1.15

Table 3-5. Estimated values of the parameters and relative errors for $R_w = 10^{-4}$

Parameter	Estimated value	Relative error (%)
x_d	0.982	1.78
$x_d^{'}$	0.296	1.12
x_q	0.661	1.64
$x_q^{'}$	0.563	2.35
H	6.340	2.51
D	1.983	0.84
T'_{d0}	0.510	2.03
$T^{'}_{q0}$	5.137	2.74
R	0.1041	4.15
T_r	0.1011	1.11
K_i	49.519	0.97
T_{avr}	0.984	1.59
T_e	0.203	1.33
K_0	2.473	1.09
K_w	31.203	4.01
T_w	10.309	3.10

Table 3-6. Estimated values of the parameters and relative errors for $R_w = 10^{-3}$

Parameter	Estimated value	Relative error (%)
x_d	0.980	2.04
$x_d^{'}$	0.296	1.33
x_q	0.661	1.74
$x_q^{'}$	0.567	2.89
H	6.308	3.04
D	2.023	1.14
T'_{d0}	0.511	2.20
$T^{'}_{q0}$	5.136	2.72
R	0.1048	4.84
T_r	0.1013	1.27
K_i	49.363	1.29
T_{avr}	1.024	2.35
T_e	0.204	2.06
K_0	2.453	1.88
K_w	31.503	5.01
T_w	10.407	4.07

Table 3-7. Mean and máximum relative errors for different values of R_w

R_w	10^{-6}	10^{-5}	10^{-4}	10^{-3}
$\lvert E_r^{max} \rvert$ (%)	3.05	3.2	4.15	5.01
$\lvert E_r^{med} \rvert$ (%)	1.36	1.27	2.02	2.49

As it is observed, the results are similar in the first two cases, with $R_w = 10^{-6}$ and 10^{-5}, becoming the errors higher as the value of R_w increases. Nevertheless, in all the considered scenarios, the mean and maximum errors remain within an acceptable range, showing the accuracy and robustness of the two-stage estimation technique presented in this thesis using UKF.

4. APPLICATION 2: VARIABLE-SPEED WIND TURBINE

This section presents the application of a KF-based technique to the estimation of the parameters involved in the dynamic equations of a variable-speed PMSM wind turbine connected to the grid through a back-to-back pair of VSCs.

First, the proposed implementation of the CKF formulation is described, as well as the tuning of the estimator. The considered case studies are presented subsequently, showing the corresponding estimation results.

4.1. Kalman filter implementation

In this part of the thesis, the CKF scheme is used for a joint estimation of the state and parameters involved in the dynamic equations of the wind turbine, the PMSM, the VSC and the pitch angle controller, [37]. The algorithm of the CKF, presented in chapter 2, has also been coded using MATLAB.

As described in chapter 3 for the fully regulated synchronous generator, the proposed methodology is based on the modification of the original parameters in the equations. Regarding the CKF tuning, it consists of the same points presented for the UKF, except for the use of tuning parameters, which are not included in the CKF algorithm.

The following signals have been taken from a simulated model, representing the measurements obtained from a real system:

- Rotor angular speed, ω.
- Wind speed, v_w.
- Pitch angle, θ_p.
- Magnitude and angle of the voltage external to the generation set, V and θ_V.
- Magnitude and angle of the current external to the generation set, I and θ_I.

These magnitudes have been split into inputs and measurements, as it is represented in Table 4-1, defining the vectors $u(t)$ and $z(t)$.

Table 4-1. Distribution of inputs and measurements for the wind turbine

Input vector, $u(t)$	Measurement vector, $z(t)$
v_w	ω
V	I
θ_V	θ_I
	θ_p

Regarding the initial estimation, a different treatment is considered for the state variables and the model parameters:

- For the state variables in the model, the simulated initial values have been taken, given that the initialization process is independent of the dynamic parameters to be estimated, being these initial values a function of the operating steady-state point, which is supposed to be known.
- For the model parameters, a total lack of knowledge on their values is supposed and they are initialized to 1. This assumption shows the

robustness and accuracy of the proposed estimation method, given that no previous information of the parameters is considered.

Finally, the tuning of the matrices P_0, Q and R is similar to that presented for the UKF in application 1. The reader may refer to the paper in annex II.2 for the particular values taken.

4.2. Results

The estimation process considered is tested in a first scenario, denominated as base case, where the model parameters are estimated using exclusively the dynamic due to the wind variability. Then, using the obtained results as initial estimation, the robustness of the KF-based algorithm is proven when three disturbances are simulated, all of them typical in the operating conditions of a wind turbine.

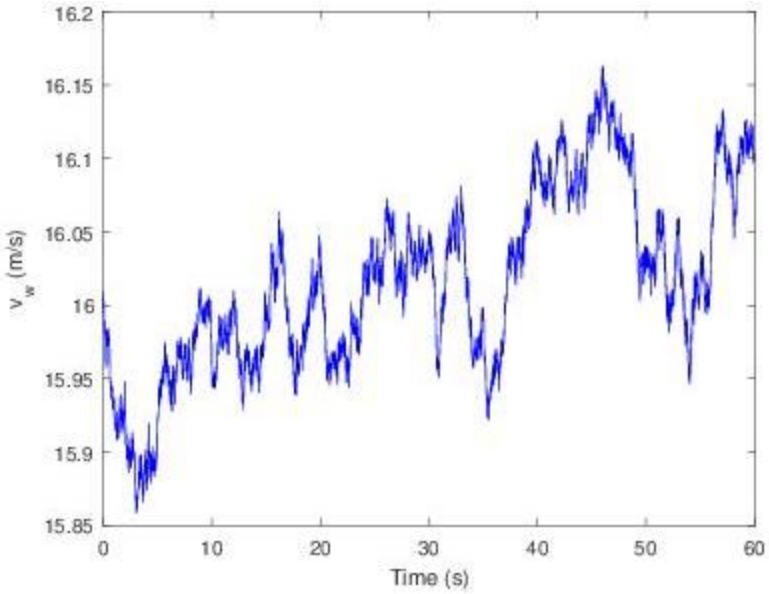
Finally, the CKF formulation is compared to other KF schemes when different noise levels are considered for the measurements included in the estimation techniques.

4.2.1. Base case

As mentioned above, in this first scenario, the variations considered in the operating point are exclusively caused by two reasons:

- The random walk assumed for the voltage in the external network (magnitude and angle), with a standard deviation $R_w = 10^{-4}$.
- A smooth wind variability, taken with a mean value of 16 m/s and a standard deviation of 1%. As an illustrative example, Figure 4-1 shows the evolution described for the wind speed.

Figure 4-1. Wind speed variation in the base case



With the CKF tuning described and the considered inputs and measurements, a total simulation time of 60 s has been taken, with a sample frequency of 100 Hz. The proposed technique has provided accurate results in the estimation of the modified model parameters. As described for the synchronous generator, the evolution of the estimated values is represented jointly with a 99,73% confidence interval ($\hat{x}_i \pm 3\sigma_i$). Figure 4-2 shows the evolution of the modified shaft inertia, while the parameters from the pitch angle controller and the VSC are represented in Figure 4-3.

It is observed that all the modified parameters eventually converge to the neighborhood of the simulated values, with a remarkable decrease in the confidence interval. Table 4-2 collects the estimation values of the original parameters, with their corresponding relative error.

With a máximum relative error of 2.1% the accurate performance of the estimation technique is concluded, so that the robustness of the estimator can be tested in the following scenarios.

Figure 4-2. Modified inertia estimation. Base case

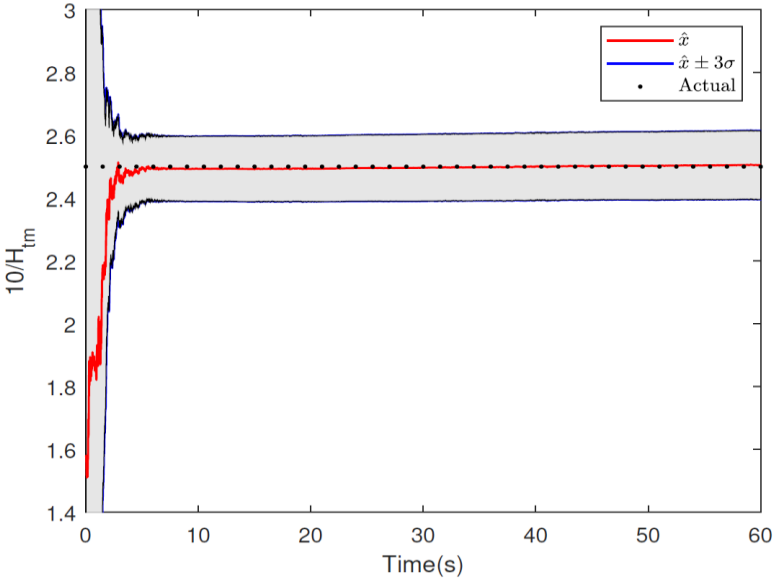
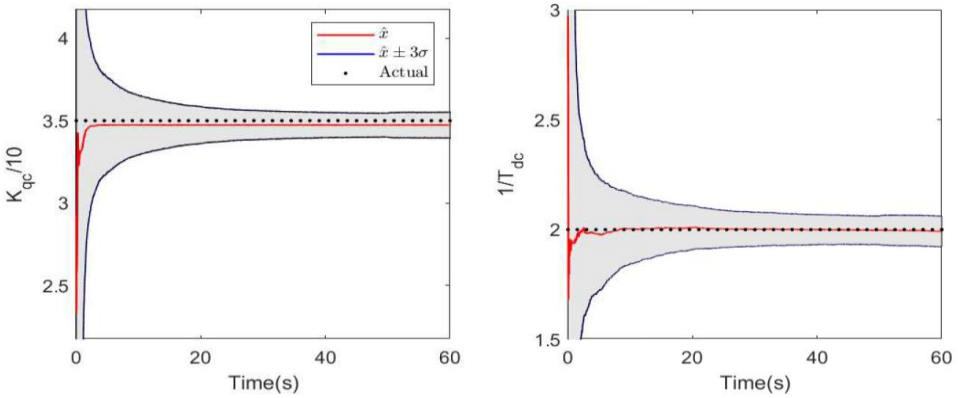
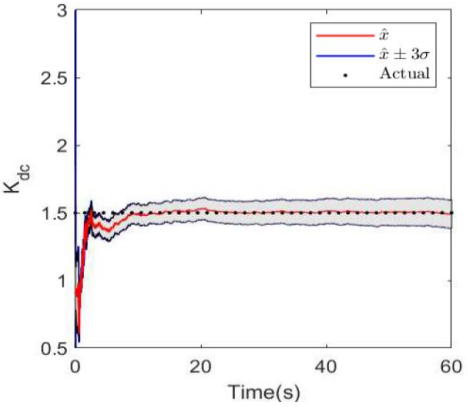
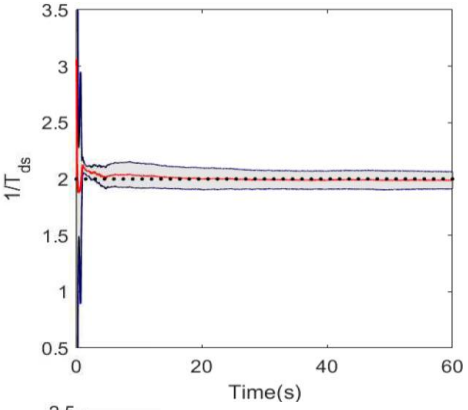
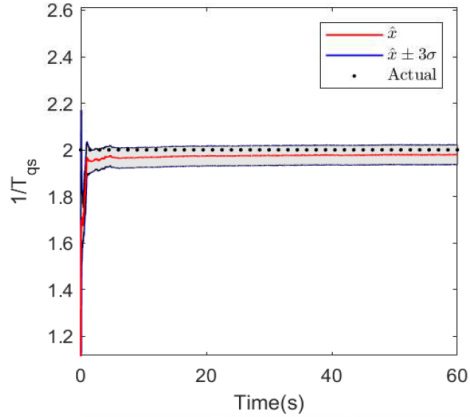


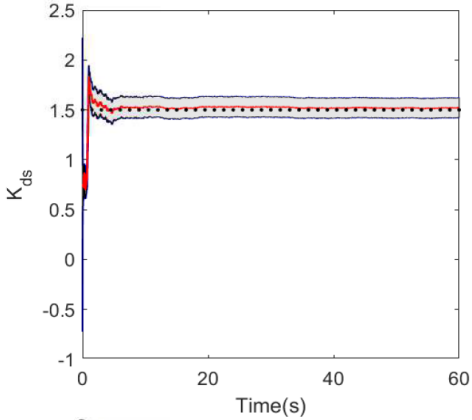
Figure 4-3. Estimation of the modified parameters from the pitch angle controller and the VSC. Base case

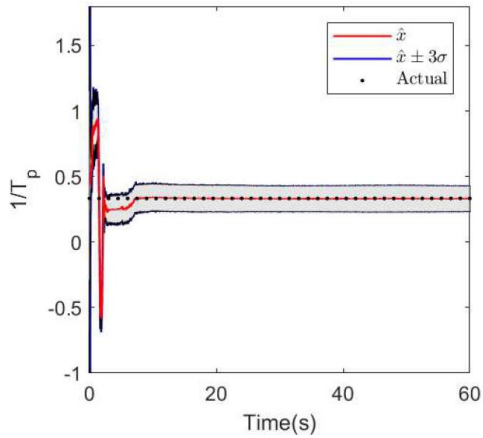



$\hat{x}$
$\hat{x} \pm 3\sigma$
Actual
K_{dc}
Time(s)


$1/T_{ds}$
Time(s)


$\hat{x}$
$\hat{x} \pm 3\sigma$
Actual
$1/T_{qs}$
Time(s)


K_{ds}
Time(s)


$\hat{x}$
$\hat{x} \pm 3\sigma$
Actual
$1/T_{p}$
Time(s)

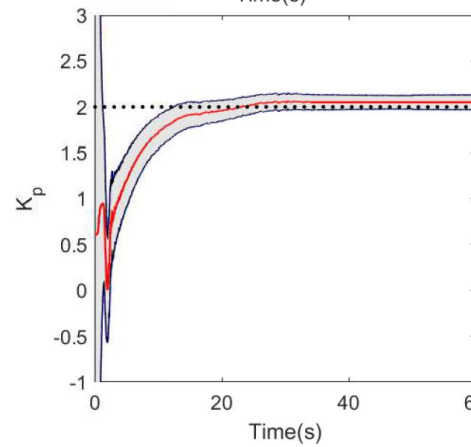

K_{p}
Time(s)

Table 4-2. Relative errors in the parameter estimation. Base case

Parámetro	Valor estimado	Error relativo (%)
H_{tm}	4.012	0.300
K_{qc}	34.426	1.640
T_{dc}	0.503	0.600
K_{dc}	1.513	0.867
T_{ds}	0.504	0.800
T_{qs}	0.511	2.100
K_{ds}	1.510	0.667
T_p	2.991	0.300
K_p	2.023	1.150

4.2.2. Disturbance 1: Wind gust

The first disturbance studied in the system consists of a wind gust, which has been modeled with a Mexican hat wavelet [37]. Additionally, a random walk has been added to the signal with zero mean and a standard deviation of 10^{-2}, yielding the evolution shown in Figure 4-4.

With this wind dynamic, the complete model was simulated in order to obtain the system inputs and measurements to be incorporated to the CKF algorithm. The estimation result for the modified shaft inertia is represented in Figure 4-5, where it has been amplified the transient behavior caused by the considered disturbance. It can be noticed that, after slight variations, the estimated value returns to the same value obtained in the base case, used as initial estimation.

Figure 4-6 shows the evolution experienced by the rest of the modified parameters in the model. The robustness of the proposed methodology can be concluded given the minimum deviations presented by all the parameters from the VSC and the pitch angle controller.

Figure 4-4. Representation of the wind gust modelled as a Mexican hat wavelet

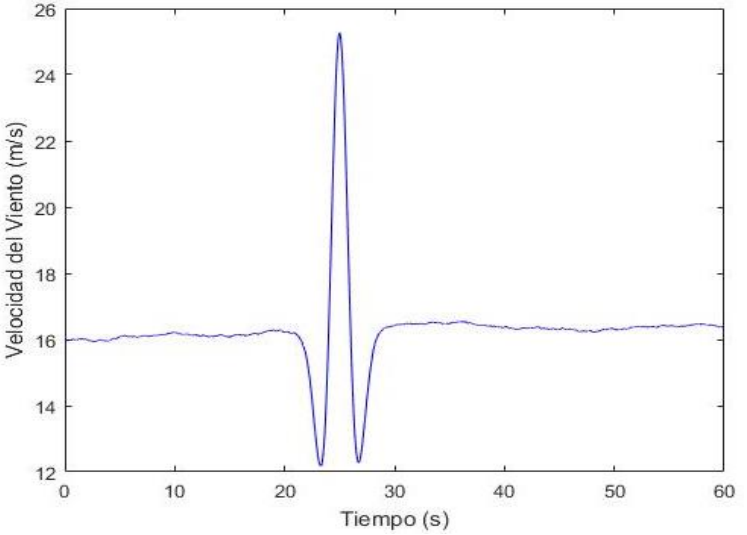


Figure 4-5. Estimation of the modified shaft inertia. Disturbance 1

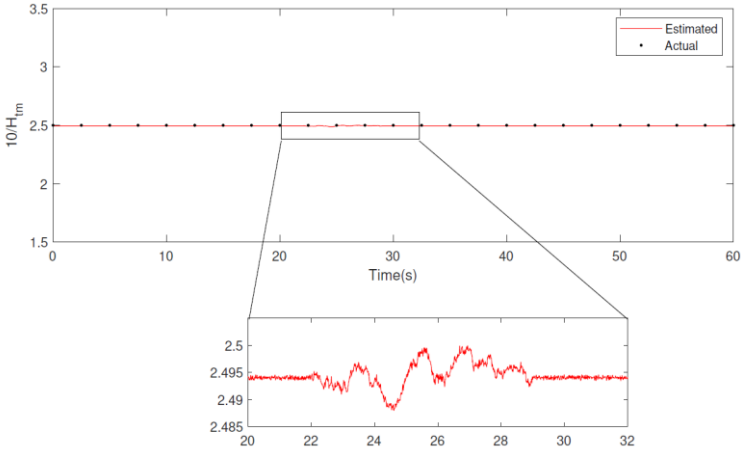
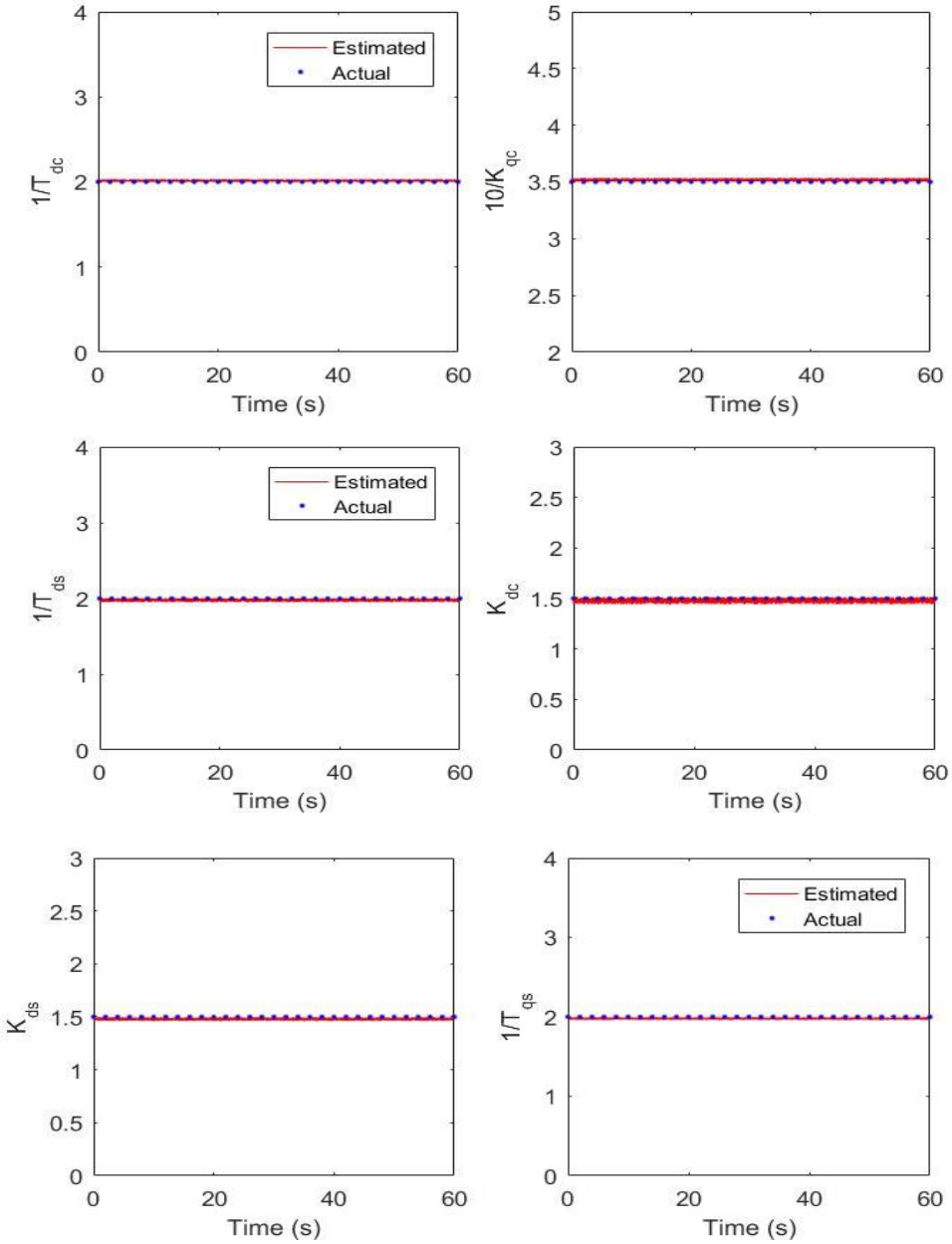
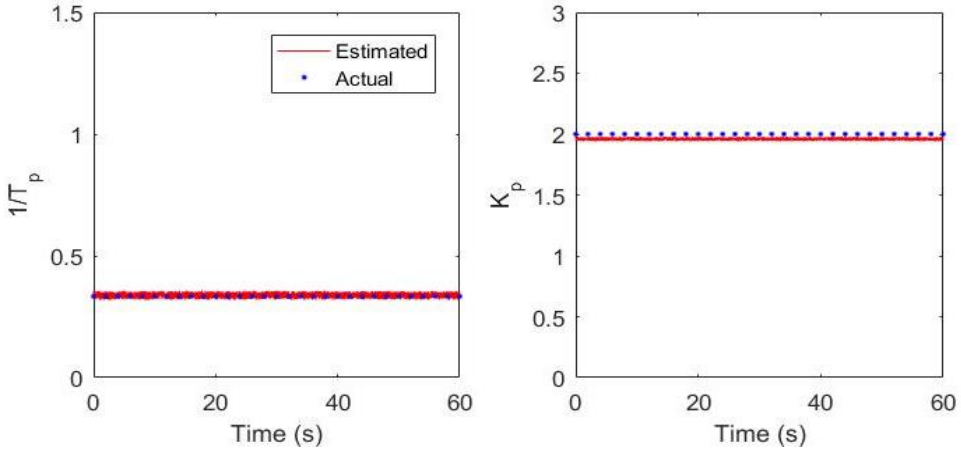


Figure 4-6. Estimation of the modified parameters. Disturbance 1





4.2.3. Disturbance 2: Voltage dip

The second disturbance considered is an abrupt descent in the voltage magnitude at the connection point of the generation set with the external network, which emulates a possible voltage dip caused by a short-circuit close to this bus of the electrical grid.

Specifically, a 70% drop in the voltage is simulated, with a duration of 1 s. The standard deviation of the random walk of this voltage remains at the same value as in the base case. The evolution of the magnitude and angle in this case is represented in Figure 4-7. The wind speed profile in this scenario is that used in the base case.

With the simulated signals used as inputs as measurements in the CKF estimator, the process was carried out using the same initial estimation as in the first disturbance.

As an example, Figure 4-8 includes the evolution of a modified time constant from the VSC. Since the rest of the parameters show no significant variations they are omitted in this document.

Figure 4-7. Voltage dip considerd in the second disturbance

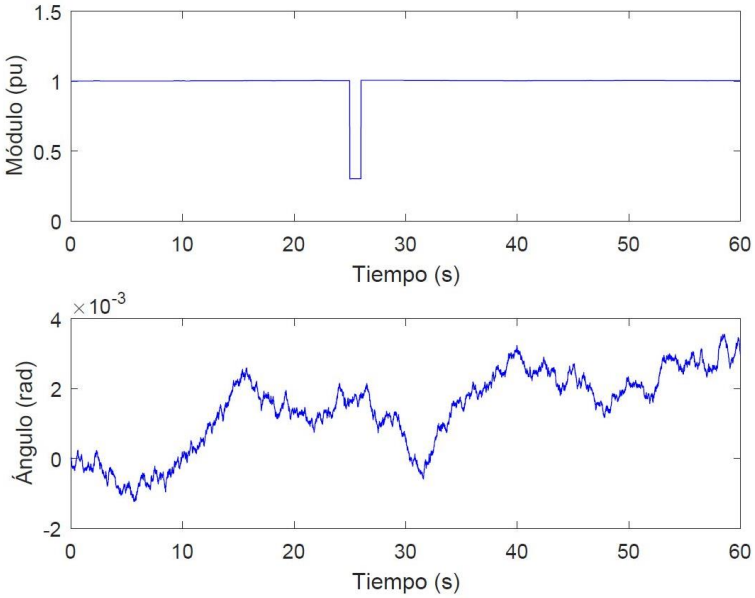
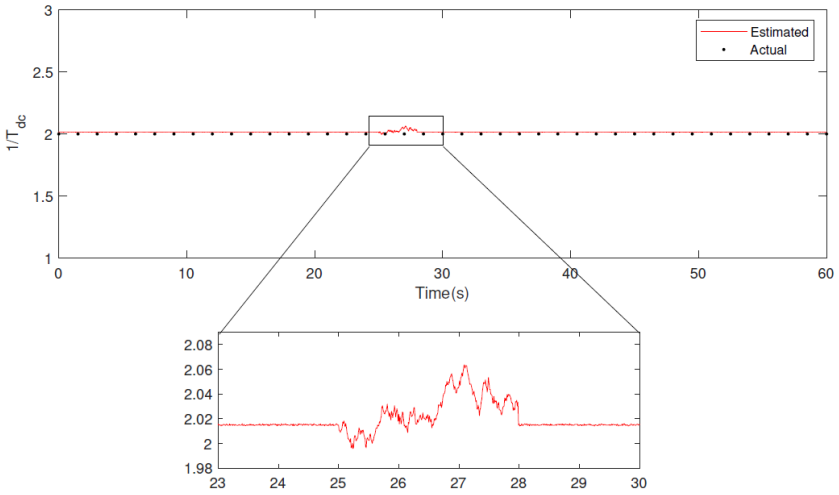


Figure 4-8. Example of the estimation results for the modified parameters. Disturbance 2



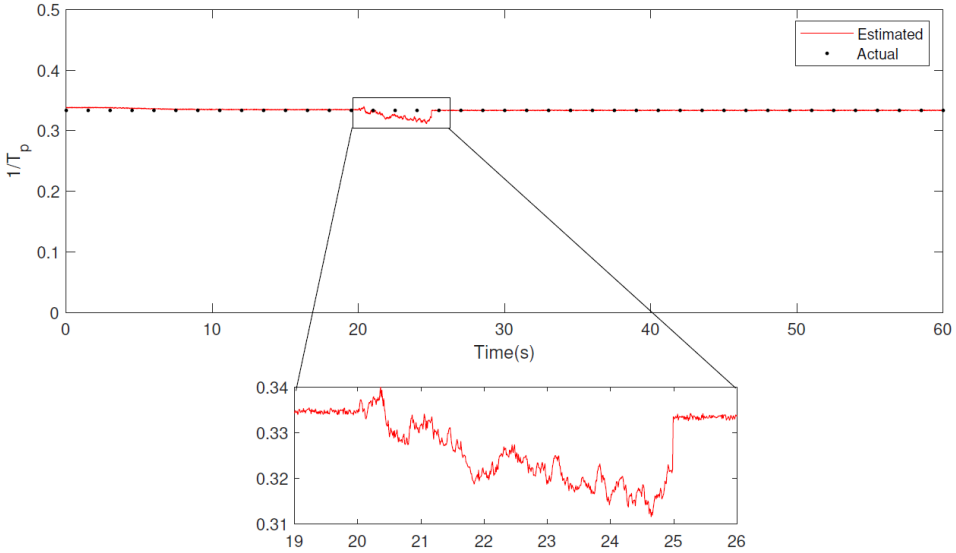
Once again, the transient behavior due to the disturbance is soon vanished and the parameter stabilizes in its initial estimation. The nature of this deviation remarks the robustness of the proposed technique.

4.2.4. Disturbance 3: Topological change

In the last disturbance considered in this part of the thesis, an abrupt and permanent change is simulated in the impedance modeling the tie-line connecting the generation set and the external network. This might represent a sudden change in the external grid configuration. The wind profile and the grid voltage remains as in the base case.

For this scenario, the evolution of the modified time constant of the pitch angle controller, $1/T_p$, is represented in Figure 4-9 as an example.

Figure 4-9. Example of the estimation results for the modified parameters. Disturbance 3



An important point to be highlighted is the fact that the operating point is permanently modified due to this disturbance, meaning that the estimated values of the modified parameters might be modified with respect to their initial estimation. This point leads in some cases to a slight improvement in the accuracy of the estimation.

In light of the presented estimation results provided by the CKF-based method in the base case, and the proven robustness in the disturbances studied, the estimation technique proposed in this thesis is confirmed.

4.2.5. Measurement error impact

Once the accuracy of the estimator has been proven, the influence of the measurement error is tested in this section. The same external conditions as in the base case are considered in this analysis.

With a 1% error in the base case, successive increments in this value have been studied (3%, 5% and 7%). Tables 4-3 to 4-5 collect the estimated values of the original parameters in the model along with their relative errors. To establish a comparison of the different noise levels, Table 4-6 summarizes the maximum relative error in each case.

Note that, while low measurement errors lead to sufficiently accurate estimation results, when the errors in the measured signals are large enough, the CKF-based estimator performance deteriorates, in proportion to those errors.

Table 4-3. Estimation errors for the parameters with a 3% measurement error

Parameter	Estimated value	Relative error (%)
H_{tm}	4.015	0.369
K_{qc}	34.374	1.821
T_{dc}	0.503	0.699
K_{dc}	1.517	1.123
T_{ds}	0.505	0.885
T_{qs}	0.515	2.854
K_{ds}	1.514	0.900
T_p	2.990	0.321
K_p	2.025	1.237

Table 4-4. Estimation errors for the parameters with a 5% measurement error

Parameter	Estimated value	Relative error (%)
H_{tm}	4.022	0.551
K_{qc}	34.13	2.547
T_{dc}	0.510	2.011
K_{dc}	1.589	5.926
T_{ds}	0.515	3.001
T_{qs}	0.526	5.224
K_{ds}	1.518	1.233
T_p	2.941	1.987
K_p	2.047	2.355

Table 4-5. Estimation errors for the parameters with a 7% measurement error

Parameter	Estimated value	Relative error (%)
H_{tm}	4.059	1.466
K_{qc}	33.228	5.333
T_{dc}	0.520	4.021
K_{dc}	1.662	10.800
T_{ds}	0.538	7.698
T_{qs}	0.548	9.577
K_{ds}	1.560	4.000
T_p	2.832	5.912
K_p	2.121	6.041

Table 4-6. Maximum relative error for increasing levels of measurement error

Measurement error	$\lvert E_r^{max} \rvert$ (%)
1%	2.100
3%	2.854
5%	5.926
7%	10.800

4.2.6. Model error impact

In the implemented model for the PMSM coupled to the wind turbine, the dq axis inductances of the machine, L_d and L_q, has been taken as known. The reason is that these values can be obtained from different tests performed on the considered machine. Nevertheless, the accuracy of this information might not be adequate, leading to a source of errors in the estimation technique.

In order to test the performance of the CKF-based methodology with model error, a total of four scenarios have been studied, with increasing levels of error (2%, 5%, 10% and 15%) in the values of L_d and L_q included in the KF model with respect to those in the simulation model.

Tables 4-7 to 4-10 collect the estimated values of the original parameters in the model with their relative errors. To establish a comparison of the different model error levels, Table 4-11 summarizes the maximum relative error in each case.

Note that, while low model errors lead to sufficiently accurate estimation results, when the errors in the known system parameters are large enough, the CKF-based estimator performance deteriorates, in proportion to those errors.

Table 4-7. Estimation errors for the parameters with a 2% model error

Parameter	Estimated value	Relative error (%)
H_{tm}	4.015	0.307
K_{qc}	34.42	1.651
T_{dc}	0.503	0.611
K_{dc}	1.514	0.902
T_{ds}	0.504	0.824
T_{qs}	0.512	2.407
K_{ds}	1.511	0.697
T_p	2.989	0.352
K_p	2.024	1.163

Table 4-8. Estimation errors for the parameters with a 5% model error

Parameter	Estimated value	Relative error (%)
H_{tm}	4.015	0.377
K_{qc}	34.360	1.863
T_{dc}	0.505	0.907
K_{dc}	1.519	1.244
T_{ds}	0.507	1.369
T_{qs}	0.515	3.011
K_{ds}	1.515	1.002
T_p	2.979	0.715
K_p	2.031	1.533

Table 4-9. Estimation errors for the parameters with a 10% model error

Parameter	Estimated value	Relative error (%)
H_{tm}	4.021	0.521
K_{qc}	34.309	2.014
T_{dc}	0.507	1.387
K_{dc}	1.535	2.354
T_{ds}	0.511	2.101
T_{qs}	0.533	6.667
K_{ds}	1.545	2.997
T_p	2.942	1.984
K_p	2.060	3.022

Table 4-10. Estimation errors for the parameters with a 15% model error

Parameter	Estimated value	Relative error (%)
H_{tm}	4.066	1.641
K_{qc}	32.881	6.443
T_{dc}	0.526	5.211
K_{dc}	1.545	3.028
T_{ds}	0.522	4.333
T_{qs}	0.560	12.002
K_{ds}	1.625	8.357
T_p	2.932	2.314
K_p	2.187	9.333

Table 4-11. Maximum relative error for increasing levels of model error

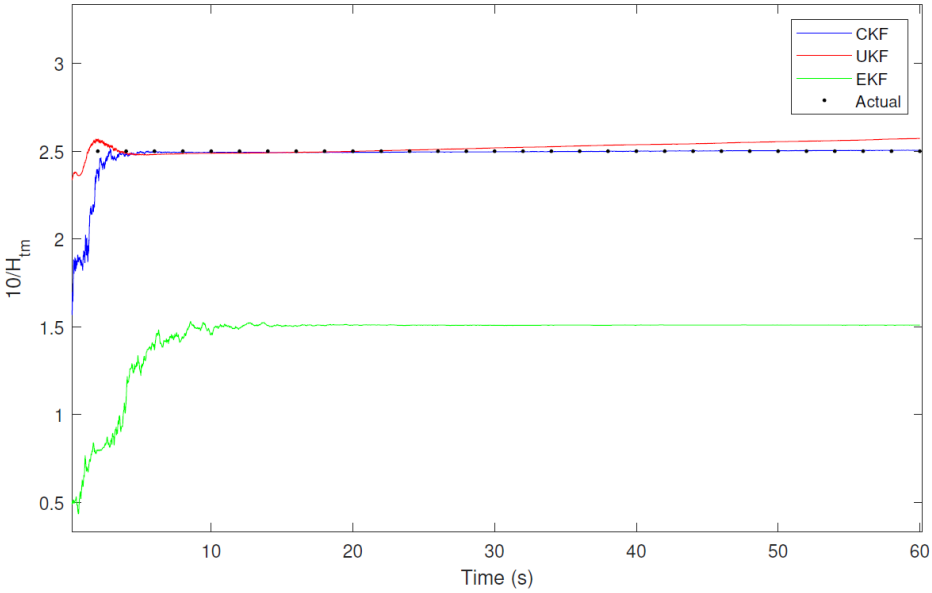
Model error	$\lvert E_r^{max} \rvert$ (%)
2%	2.407
5%	3.011
10%	6.667
15%	12.002

4.2.7. Comparison with other formulations

The last study made in the second part of this thesis consists in the comparison of the performance of the CKF formulation with that shown by other nonlinear KF schemes. Particularly, the UKF and the EKF are considered.

Operating conditions of the base case are taken to establish a fair comparison for the mentioned formulations. As an illustrative example, Figure 4-10 presents the evolution of the modified shaft inertia for the three estimators.

Figure 4-10. Performance comparison in the estimation of shaft inertia by CKF, UKF and EKF



The following considerations can be pointed out from the figure:

- The EKF formulation presents accuracy problems, with an unacceptable value for the modified parameter.
- On the other hand, the UKF presents convergence problems in the long term.

It can be concluded that the CKF scheme is the most adequate for the joint state-parameter estimation of the considered model for the variable-speed wind turbine, the back-to-back VSC and the pitch angle controller.

5. APPLICATION 3: CUSTOMER-PHASE IDENTIFICATION

This section presents the application of KF-based estimation techniques applied to the identification of the electrical phase of single-phase clients in distribution grids, denominated as CPI problem.

First, the proposed KF methodology is described, as well as the tuning of each of the three formulations. The considered case studies are presented subsequently, showing the corresponding estimation results.

5.1. Kalman filter implementation

For the identification of the electrical phase in distribution grids, a comparative analysis is made for three formulations of the KF, the UKF, the CKF and the EnKF. The algorithms of these schemes have been coded using MATLAB framework.

To approach the CPI problem with the three mentioned estimators, a novel methodology is proposed, consisting of a sequential selection and assignment of the single-phase clients, using a statistical criteria based on the calculation of a set of coefficients from the estimated values of the state variables and the covariance of the error in those estimations. This methodology is illustrated in the block diagram presented in Figure 5-1 and described in detail in section IV of the publication in annex II.3.

The use of this selective method is one of the major contributions of this work,

avoiding the inclusion of non-convex equality constraints of the form $x(x-1)=0$ in the estimator formulation, which can lead to convergence problems. The proposed technique is novel to the author's knowledge and can be applied to other binary problems.

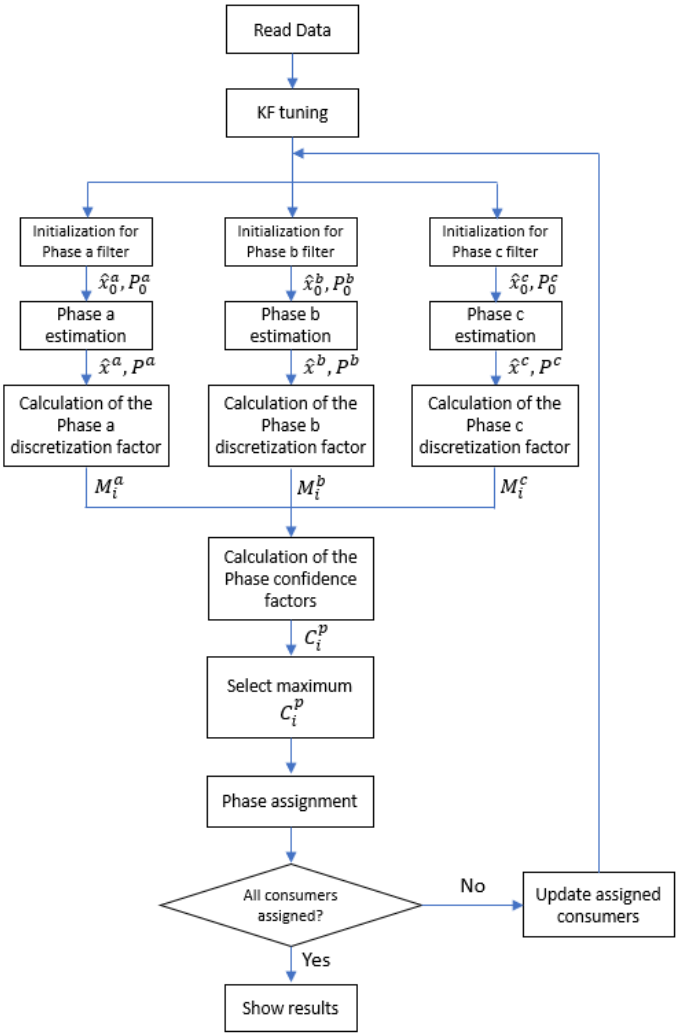
Regarding the tuning of the KF-based estimators, the covariance matrices Q and R have been taken as diagonal matrices with $Q_{ii}=10^{-4}$ and $R_{ii}=9\cdot 10^{-4}$ in the three formulations. For the initial estimation, zero values have been used for the state vector, $\hat{x}_0$, in the UKF and CKF schemes, whereas in the case of EnKF, a binary random initialization has been considered for the samples in the ensemble. In the three cases, the covariance matrix P_0 has been taken as diagonal with $P_{ii}=10$.

The measurements taken from the simulated distribution grid are:

- The hourly active energy consumption for each client in the network considered as the system input vector, $u(t)$.
- The hourly active energy consumption for each of the three phases at the substation level, taken as the measurement vector, $z(t)$, for the correction stage of the KF algorithm.

Additionally, while for the simulated distribution grid a load flow has been solved in order to calculate the network losses accurately, a simplified loss model has been incorporated in the filter equations in order to improve the performance of the estimation. This model assumes unitary power factor and nominal voltage for each consumer.

Figure 5-1. Block diagram of the proposed methodology for the CPI problem



5.2. Results

A total of four scenarios have been considered in order to test and compare the accuracy and robustness of the different KF formulation applied to the CPI problem with the methodology developed in this work.

Finally, the performance of the KF is compared with that of other existing techniques approaching the electrical phase identification.

5.2.1. Scenario 1: Original measurements

In the first scenario, the original measurements from the simulated grid are taken, being the objective to compare the accuracy of the different formulation with increasing number of clients. The hourly energy measurements are obtained from [39], where real hourly data from a European distribution company, comprising smart meter readings for 20 days, are provided.

Tables 5-1, 5-2 and 5-3 show the estimation results for the UKF, the CKF and the EnKF, respectively.

Table 5-1. Estimation results for the UKF. Scenario 1

Consumption curves	Correct assignments	Wrong assignments	Percentage
50	50	0	100%
100	95	5	95%
200	181	19	90.5%
300	216	84	72%
400	233	167	58.25%

Table 5-2. Estimation results for the CKF. Scenario 1

Consumption curves	Correct assignments	Wrong assignments	Percentage
50	50	0	100%
100	100	0	100%
200	195	5	97.5%
300	262	38	87.33%
400	313	87	78.25%

Table 5-3.Estimation results for the EnKF. Scenario 1

Consumption curves	Correct assignments	Wrong assignments	Percentage
50	50	0	100%
100	100	0	100%
200	200	0	100%
300	283	17	94.33%
400	346	54	86.5%

In light of the performance of UKF and CKF with increasing number of loads shown in Tables 5-1 and 5-2, respectively, it can be noticed that, while both estimators correctly assign 100% of customers for a reduced number of loads (50 clients), the behavior of the UKF deteriorates faster as the number of loads increases. As shown in Table 5-3, the EnKF response to increasing system sizes is better than that of the other formulations, which confirms the expected behavior of this KF scheme for high-dimensional problems.

5.2.2. Scenario 2: Analysis of the required amount of data

For the results shown above, hourly smart meter readings for 20 days are used, leading to 480 total energy measurements for each consumer. This section analyzes the quality of the estimation as the number of available measurements decreases. Tables 5-4, 5-5 and 5-6 show the estimation results for the UKF, the CKF and the EnKF, respectively, using 200 consumption curves in each case.

Table 5-4. Estimation results for the UKF. Scenario 2

Available hourly measurements	Hit rate
480	90.5%
400	90.5%
300	90%
200	85%
100	67%

Table 5-5. Estimation results for the CKF. Scenario 2

Available hourly measurements	Hit rate
480	97.5%
400	97.5%
300	96%
200	90%
100	70%

Table 5-6. Estimation results for the EnKF. Scenario 2

Available hourly measurements	Hit rate
480	100%
400	100%
300	100%
200	95%
100	72%

Those results suggest a deterioration of the performance of the KF-based estimation techniques when the number of available measurements is lower than 200 (around 8 days) in this particular scenario. Further tests with feeders comprising different numbers of customers show that the required number of measurement snapshots increases with the number of loads, as intuitively expected. For instance, for 100 clients, the success rate of the three estimators does not deteriorate substantially, even when only 100 energy measurements (around 4 days) are considered for each load.

5.2.3. Scenario 3: Noisy measurements

The performance of the different KF formulations is evaluated in a realistic scenario where errors in the measurements are considered. As the objective of these case studies is to determine the robustness of the KF schemes against measurement errors, a relatively low number of loads is considered, namely 100 consumption curves.

Gaussian noise is artificially added to each measurement after the load flow is computed. Tables 5-7, 5-8 and 5-9 show the estimation results for the UKF, the CKF and the EnKF, respectively.

Table 5-7. Estimation results for the UKF. Scenario 3

Noise level (%)	Hit rate
1	95%
2	92%
3	89%
5	83%

Table 5-8. Estimation results for the CKF. Scenario 3

Noise level (%)	Hit rate
1	100%
2	96%
3	93%
5	89%

Table 5-9. Estimation results for the EnKF. Scenario 3

Noise level (%)	Hit rate
1	100%
2	100%
3	98%
5	92%

In light of those results, it can be concluded, also as expected, that the number of correct assignments decreases with increasing measurement noise, for every KF formulation, being the robustness of the CKF and EnKF similar, superior in any case to that of the UKF formulation. Nevertheless, all formulations show acceptable results when typical noise levels are considered in the measurements.

5.2.4. Scenario 4: Model errors

In the proposed implementation of the KF for the CPI problem, a simplified loss model is considered for which the value of the conductor resistance per unit length, r. In this scenario, the performance of the different KF formulations is evaluated when errors in r are considered.

For the same number of customers as in Scenario 3, Tables 5-10, 5-11 and 5-12 show the estimation results for the UKF, the CKF and the EnKF, respectively.

Table 5-10. Estimation results for the UKF. Scenario 4

Error in r (%)	Hit rate
5	95%
10	91%
15	86%
20	80%

Table 5-11. Estimation results for the CKF. Scenario 4

Error in r (%)	Hit rate
5	100%
10	95%
15	91%
20	84%

Table 5-12. Estimation results for the EnKF. Scenario 4

Error in r (%)	Hit rate
5	100%
10	98%
15	95%
20	89%

It can be concluded that the results remain acceptable, at least for the EnKF, when the assumed resistance error does not exceed 10%.

5.2.5. Comparison with existing methods

Finally, the proposed methodology has been compared to other existing techniques which use the same information from the distribution grid to approach the CPI problem. Specifically, the following studies are selected:

- The method presented in [4], where a LASSO-based technique is considered using hourly energy measurements exclusively.
- The technique described in [5], implementing a PCA analysis with the same information extracted from the network.

Table 5-13 summarizes the success rates for each technique with increasing number of clients. A 1% s.d. error has been considered for all the measurements.

In light of the presented results, it can be concluded that:

- The performance of the PCA is similar to that shown by the UKF in all the scenarios considered.
- The LASSO-based technique has proven to suffer from higher sensitivity to the number of clients than the CKF and EnKF, being the success rates of the three methods similar for a number of clients lower than 200.
- The obtained results for the CKF and EnKF formulations are substantially better in large networks, as it can be observed in the scenario with 400 consumers.

Table 5-13. Rate of correct assignments for the PCA, the LASSO-based and the proposed KF-based methods

Number of clients	PCA	LASSO	UKF	CKF	EnKF
50	100	100	100	100	100
100	100	100	95	100	100
200	92	100	90.5	97	100
300	72.67	87.67	71	86.33	94
400	60.25	71.75	57.75	78.25	86.5

6. FINAL CONCLUSIONS

The conclusions derived from the results shown in chapters 3-5 are divided in accordance with the three parts of this thesis, the fully regulated synchronous generator, the variable speed-wind turbine and the CPI problem. Additionally, the possible lines of investigation are proposed for each study field. Finally, the most relevant publications made by the Ph.D. student with his thesis advisors are presented.

6.1. Conclusions

6.1.1. Regulated synchronous generator

An estimation technique has been presented using the UKF formulation to identify the state and parameters of a synchronous generator with frequency and voltage regulation, conducted by the SG, the AVR and the PSS. In light of the results described in this document, it can be concluded that:

- The two-stages implementation proposed for the UKF allows, with a simple tuning of the estimator, the accurate identification of the parameters involved in the system model, being the maximum relative error under 3.2%, giving evidence of the good performance of the algorithm.
- The presented parameter modification methodology has successfully solved the convergence issues observed with the original parameters

involved in the model equations.

- The difference between the simulation and the estimator models, jointly with the use of external measurements for the UKF, enable the application of the proposed technique to real systems, with a set of measurements obtained from PMUs or a local data logger.
- The inclusion of the PSS in the estimator model constitutes an improvement with respect to other related studies.
- The estimation technique presented has not only shown good accuracy, but also robustness when modification in the Random Walk of the voltage in the external grid is considered.

6.1.2. Wind turbine

In the second part of this thesis, the CKF formulation is considered for the joint estimation of the state and parameters of a wind turbine coupled to a PMSM with full converter control and pitch angle regulation.

The following conclusions are obtained from the results presented in previous sections:

- As it was remarked for the synchronous generator, the parameter modification methodology has successfully solved the convergence problems observed when the original parameters were included in the estimator model.
- The estimation process based on CKF has provided a high degree of accuracy in the identification of the model parameters considering smooth variations in the system for a typical operating point, with the corresponding wind variability.
- Three additional case studies were included to prove the robustness of the algorithm, which presented a good performance with the considered disturbances, recovering the steady state estimation for the parameters after a light transient behavior.
- It was also observed that the CKF based estimator shows acceptable results when the model and measurement noises are increased.

- Finally, the CKF formulation has shown to be superior to other KF schemes, such as the UKF and the EKF.

6.1.3. Customer-phase identification

In the third part of this thesis, a technique based on Kalman filtering has been presented, applied to the identification of the electrical phase which individual loads are connected to in distribution grids, using exclusively hourly energy measurements obtained from smart meters.

A total of 4 scenarios have been considered as case studies to prove the good performance of the proposed methodology in different situation. Additionally, the obtained results for the different KF formulations have been compared to those from two previously published methods.

From the results presented in the chapter 5 of this document, the following conclusions are drawn:

- The UKF scheme is less accurate as the number of loads in the grid increases, showing unacceptable hit rates (<60%) when 400 customers are considered.
- The performance of the CKF is better than that of the UKF, with a lower sensitivity to the size of the state vector and measurement and model errors.
- The EnKF formulation has proven to be the best of the considered KF schemes for the consumer-phase identification problem, both in terms of success rates and sensitivity to noise.
- As it was expected, the proposed methodology is affected by the amount of energy measurements available, relative to the number of clients in the distribution grid under study. As the number of loads in a certain feeder increases, the number of required measurements for a correct phase identification also raises.
- Finally, from the comparative study considered it can be concluded that the CKF and the EnKF schemes present better results than those of the

methods based on PCA and LASSO, particularly when the number of loads is high, being the success rates similar when the grid size is small.

6.2. Future lines of investigation

6.2.1. Regulated synchronous generator

Regarding the possible future investigations related to the parameter estimation in fully regulated synchronous generators, several lines are proposed:

- A more complex model included in the UKF for the generator and its regulators would reflect in more detail the real behavior of the system under study, contributing to the correct operation and control of the grid.
- Other promising future work would be the use of the proposed estimation technique with measurements obtained from real generation sets, given the good performance observed in the simulations.
- The proposed estimation technique might be applied to more complex dynamic phenomena with unknown parameters, such as the natural frequency in high dimensional electric power systems.

6.2.2. Wind turbine

Related to the parameter estimation in variable-speed wind turbines, the possible future lines of investigation are similar to those presented for the synchronous generator, i.e. the use of more complex models for the elements of the studied system, and the application of the proposed technique to real wind turbines, given the good performance in the simulations and the direct application of the proposed method, which uses exclusively external measurements.

Additionally, the proposed methodology is appropriate for the parameter identification in other types of systems using power electronic converters, as might be photovoltaic generation plants or battery energy storage systems.

6.2.3. Customer-phase identification

Finally, the following point are remarked regarding the future lines of investigation related to the phase identification problem:

- The proposed methodology based on Kalman filtering could be modified in order to include three-phase consumers for which only aggregated hourly energy measurements are available.
- Given the potential of the existing measuremet equipments, a possible upgrade of the presented algorithm might be based on the addition of information related to the voltage or the reactive energy consumption, in order to improve the success rates of the implemented estimator.

6.3. Publications

First, the publications related to the three parts of this thesis are highlighted:

- [40] M.A. González-Cagigal, J.A. Rosendo-Macías, A. Gómez-Expósito, "Parameter estimation of fully regulated synchronous generators using Unscented Kalman Filters", in: Electric Power Systems Research, Volume 168, 2019, Pages 210-217, November, 2018.
- [41] M.A. González-Cagigal, J.A. Rosendo-Macías, A. Gómez-Expósito, "Parameter Estimation of Wind Turbines with PMSM using Cubature Kalman Filters," in: IEEE Transactions on Power Systems, Volume .35, 2020, pp. 1796-1804, May 2020.
- [42] M.A. González-Cagigal, J.A. Rosendo-Macías, A. Gómez-Expósito, "Application of nonlinear Kalman filters to the identification of customer phase connection in distribution grids," in: International Journal of Electrical Power and Energy Systems, Volume 125, 2021.

Related to publications [40] and [41], the following presentation at a conference special session was made:

- [43] J.A. Rosendo-Macías, M.A. González-Cagigal, A. Gómez-Expósito, "PMU-Based Estimation of Renewable Power Plants Parameters." In: IEEE Powertech Conference. Milán (Italia). 2019

Finally, as mentioned in the introduction of this document, an additional contribution was made outside the field of power systems, by applying estimation techniques based on KF for the monitorization and possible prediction of the evolution of the Covid-19. We believe some of the novel ideas proposed in this work to model and estimate slowly changing parameters, could be as well of application in the future to related power system problems.

A daily report of the situation of the pandemic in Spain was published in the repository idUS, property of the University of Seville:

- [44] A. Gómez-Expósito, J.A. Rosendo-Macías and M.A. González-Cagigal, "Modelo y análisis de la evolucion de una pandemia vírica mediante filtros de Kalman: el caso del Covid-19 en España", In: https://idus.us.es/handle/11441/94508

Additionally, a preliminary version of the paper was submitted to the repository Medrxiv:

- [45] A. Gómez-Expósito, J.A. Rosendo-Macías and M.A. González-Cagigal, "Monitoring and tracking the evolution of a viral epidemic through nonlinear Kalman filtering: Application to the Covid-19 case", In: https://www.medrxiv.org/content/10.1101/2020.05.11.20098087v1

While the final version, included in annex II.4, and which has been submitted to "IEEE Journal of Biomedical and Health Informatics" is currently under minor revision.

ANNEX I. KALMAN FILTER APPLIED TO THE MONITORIZATION OF A VIRUS SPREAD

In this annex, the motivation and the most relevant results are presented for the estimation technique proposed by the Ph. D. student and his thesis advisors for the monitorization and tracking of the evolution of a viral epidemic. Finally, the conclusions derived from the obtained results are mentioned.

I.1 Motivation

Despite the spectacular medical advances of the 20th century, and the practical eradication of viral diseases that in the past caused great mortality (e.g., smallpox), modern societies are still very vulnerable to the sudden appearance of new viruses, such as the SARS-CoV-2, cause of the Covid-19, for which, in the summer of 2020, there is still no vaccine. In addition, once a viral outbreak originates in a region of a country (in the case of the Covid-19, the Chinese region of Hubei, where the first reported case was dated on December 2019), the globalization of the economy and mass tourism spread it almost inevitably and quickly to the rest of the world.

In the absence of effective treatments, once a certain threshold has been passed, the main and almost sole remedy against the spread of the disease to the entire population is social distancing, the objective of which is to minimize the contact between people, and therefore morbidity [46]. In extreme cases, when the speed of propagation of the outbreak is very high, massive lockdowns of entire countries may be needed, which cannot last indefinitely owing to their drastic impact on the economic activity.

For this reason, all the agents involved (governments, international organizations, institutions, companies and individuals) have the greatest interest in knowing how the number of affected and deceased people will evolve over time, with a view, on the one hand, on verifying the beneficial effects of social distancing, and on the other to scheduling the already saturated health resources and taking the economic measures intended to mitigate as far as possible the devastating effects of an epidemic like that of Covid-19.

Scientists, engineers, economists, etc. are acquainted with several mathematical and statistical toolkits (recently renamed collectively as "data analytics") for the treatment and filtering of time series, with a view on extracting useful information from the available data, uncertain by definition, such as trends, patterns, average values, expected variances, etc. In the specific case of a viral epidemic, such as that of Covid-19, there are basically two categories of models for processing the information:

- Models that try to characterize the "physical" reality explaining the observed data., as those in [47] and [48].

- Models that try to determine explanatory parameters or variables from a purely mathematical point of view, the so-called "black box" approach, used in [49] and [50].

The methodology proposed, which belongs to the second category, uses KF to process both the assumed dynamic model and the information available throughout the outbreak. Reference [51] applies the KF for the estimation of the evolution of AIDS, while several recent studies related to the Covid-19 have arisen. In [52] the KF is used to deal with the estimation of the reproductive number of the virus. A short-term prediction model is proposed in [53], where the time update equations of the estimator are used for future forecasts of the pandemic spread. ARIMA models are combined with a KF in [54] to track the evolution of the Covid-19 in Pakistan. Unlike in those references, where the parameters involved in the state estimation process are supposed to be known, in this work such assumptions are not required. This is the major distinguishing feature of the proposed methodology, compared to the state of the art, and the main contribution of the paper included in annex II.4.

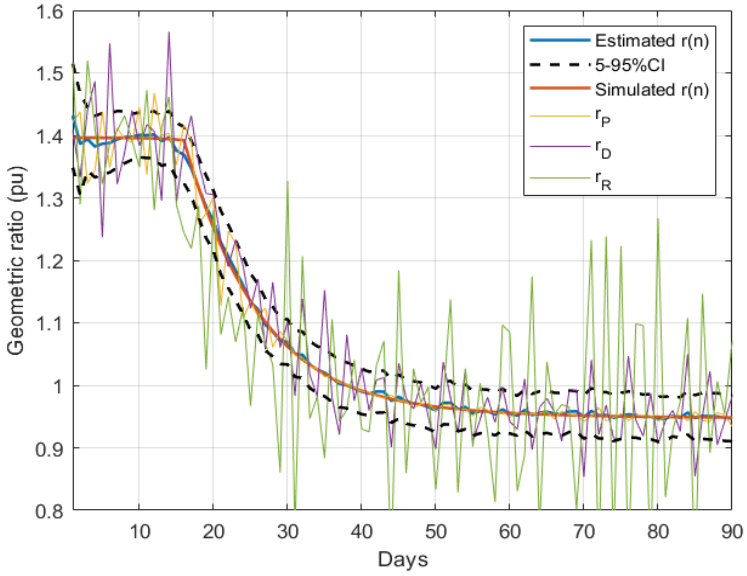
I.2 Main results

In this section, the most relevant results are presented. The reader may refer to the last version of the paper in annex II.4 for the detailed description of the proposed KF-based estimation technique.

II.2.1 Simulation results

First, a set of simulated scenarios, where the SIRD model [55-57] is considered, are used to test the performance of the proposed implementation of the KF. Figure I-1 represents the time evolution of the estimated geometric ratio, $r(n)$, along with the raw noisy measurements provided by the simulation and the actual value of $r(n)$.

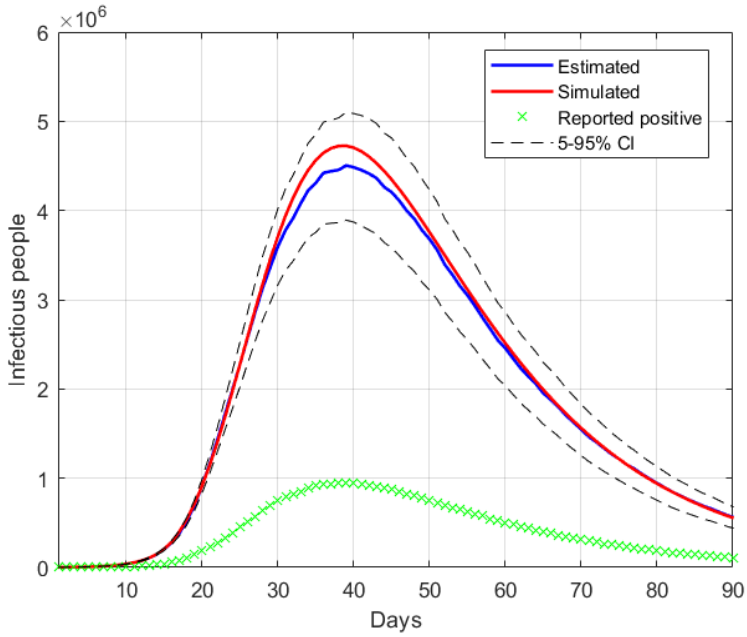
Figure I-1. Estimation of r(n) in the simulated scenario



It can be noticed that the simulated value of $r(n)$ remains always within the 5-95% CI, giving evidence of the good performance of the proposed technique. From the estimated $r(n)$ and the initial testing

ratio considered, $t(0)$, an estimation is obtained for the evolution of the infectious people, $I(n)$, compared in Figure I-2 with the simulated value.

Figure I-2. Estimation of I(n) in the simulated scenario



It is observed how the maximum estimation error (around 4.5%) takes place, as expected, at the peak of the epidemic.

Additionally, the effect of two error sources is studied:

- An abrupt change in the testing ratio, t(n).
- An error in the initial guess of the testing ratio, t(0).

Both analysis confirmed the robustness of the estimation technique, which is subsequently applied to real data from different countries.

II.2.2 Case studies

The case studies are divided in two subsections: 1) the time period when massive lockdowns occurred in most countries, denoted in the media as the "first wave" of the pandemic [58], and 2) the subsequent transient period, once the lockdowns are relaxed, usually through several de-escalation phases, towards the so-called "new normality".

1. Lockdown period (first wave)

A total of four countries have been considered in this period: China, South Korea, Spain and the UK. At the early stage of the pandemic, the information provided by these countries was sufficient to allow the application of the proposed methodology.

As an example, Figures. I-3 and I-4 represent for the Spanish territory the estimated sequence of the geometric ratio, $r(n)$, and the number of infected people, $I(n)$, respectively. In this particular case, and regarding the parameter $t(0)$, it has been taken into account the results of a massive seroprevalence test performed by the government in the first half of May [59], from which it was concluded that the total number of infected people was around 5.2% of the population (approximately 2.3 million people). In view of this valuable information, the initial value $t(0)$ has been adjusted so that the cumulative number of infected people matches the result of the survey on the date it was released (May 13), leading to $t(0)$=0.12. This provides the estimation of $I(n)$ shown in Fig. I-4, where a maximum value of the active infectious people of around 1.3 million can be noticed by mid-April.

Figure I-3. Estimation of r(n) in Spain

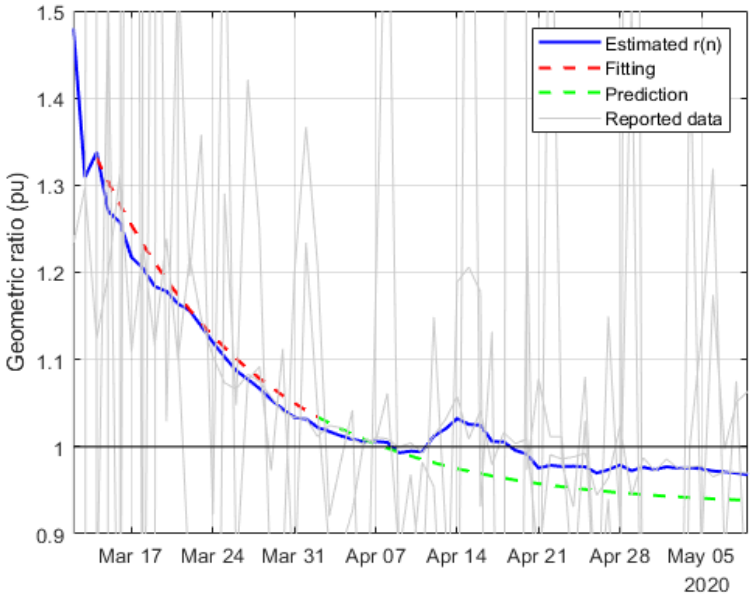
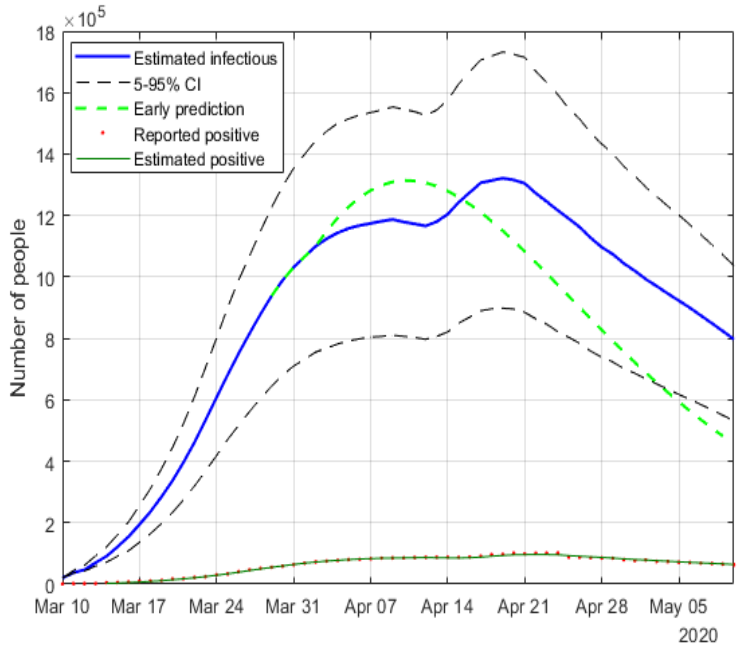


Figure I-4. Estimation of I(n) in Spain



2. Post-lockdown period (second wave)

As the pandemic evolves, it becomes more difficult to properly report on a regular basis all the information involved in the estimation of active positives. Many countries (notably Spain) stopped reporting the number of recovered people, probably owing to the remarkable increase in the number of asymptomatic positive cases, which never entered a hospital and hence never counted as recovered or dead. For this reason, it is not possible to accurately update the estimations of the geometric ratios of active positives, $r(n)$, for some of the countries considered in the early stages. Instead, four countries (USA, Italy, India and Brazil), all of them specially affected by the pandemic and still reporting the information required by the proposed estimation technique, are considered for the estimation of the geometric ratio $r(n)$ and the number of infectious people, $I(n)$. As an example of the obtained results, Figures I-5 and I-6 represent the case of USA.

It is observed how the number of infectious people in this country briefly reached a peak by the end of May ($r(n) \approx 1$). However, Figure I-5 also shows that, afterwards, $r(n)$ has remained somewhat around or above one, which means that the outbreak is not still under full control, and that additional actions should be taken in this country in order to substantially reduce the number of infectious people.

I.3 Conclusion

The proposed methodology has been satisfactorily tested on a simulated case, in the presence of Gaussian noise and other sources of uncertainty, the main one being the number of infectious people at the onset of the outbreak. The estimation technique has also been applied to a set of countries, and the results obtained are divided in two periods:

- A first period, when most of the countries imposed a lockdown. Four territories are reported in this scenario, namely: China, South Korea, Spain, and the UK. The evolution of $r(n)$ reflects in all cases the severity of the lockdown, allowing the first peak of the epidemic to be forecasted well in

advance. In some cases, a slightly increasing trend is apparent in the evolution of this ratio once the lockdown is removed, suggesting that additional mobility restrictions might be necessary.

- For the countries that have continued reporting the required information, the estimation of $r(n)$ is extended up to the moment of writing this manuscript, reflecting the panoply of post-lockdown measures taken by most of them. In this case, four countries are reported: the USA, Italy, India and Brazil. The results show how the geometric ratio $r(n)$ keeps rather close to 1, or slightly above, which explains the onset of the second wave many countries were facing in early November 2020.

In light of the presented results, it can be concluded that the proposed methodology can effectively characterize, by means of the ratio $r(n)$, the evolution of the virus spread, when adequate information of active positives, recovered and deceased people is available. This information on the state and dynamics of the epidemic can be used by the governing authorities in order to take the corresponding actions:

- An increasing trend of the geometric ratio represents a virus spread which might turn out of control, especially when $r(n) > 1$, leading to more restrictive policies.

Figure I-5. Estimation of r(n) in USA

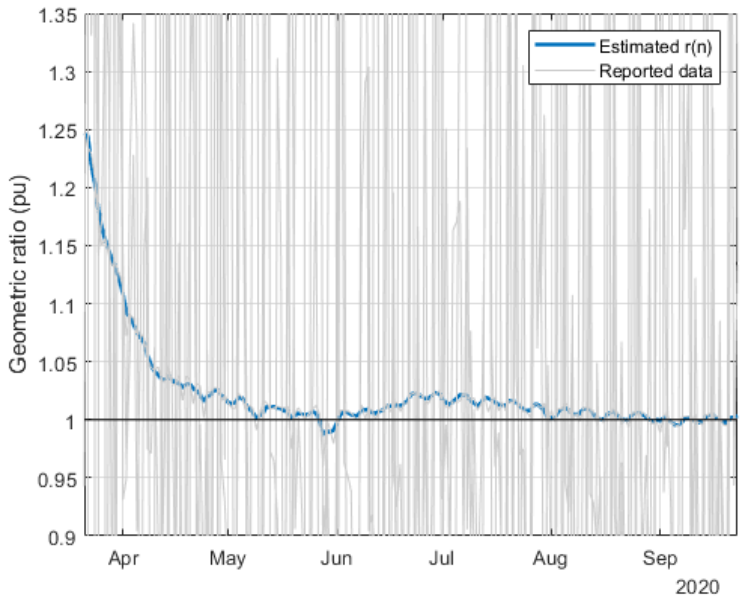
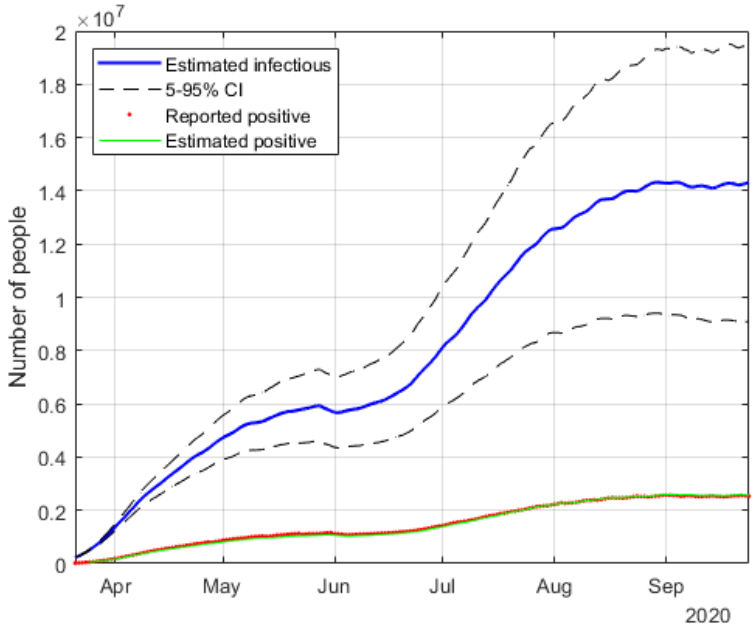


Figure I-5. Estimation of I(n) in USA



- On the contrary, values of $r(n) < 1$ with decreasing trend indicate a situation where the severity of the social distancing measures can be alleviated.

As shown in the simulated scenario, the proposed methodology is not only suitable for the Covid-19, but also for other pandemics that can be characterized using the SIRD model, and for which the required information is available. Future work is aimed to the application of KF-based estimators to new models that can arise with less informative scenarios.

ANNEX II. PUBLICATIONS

In this annex, the most relevant publications of the Ph. D. students and his thesis advisors are included. While annex II.1-II.3 comprise the publications approached in the body of the document, the last version of the paper on the methodology for the monitoring of the Covid-19 evolution, described in annex I, is included in annex II.4.

Annex II.1

M.A. González-Cagigal, J.A. Rosendo-Macías, A. Gómez-Expósito, "Parameter estimation of fully regulated synchronous generators using Unscented Kalman Filters", in: Electric Power Systems Research, Volume 168, 2019, Pages 210-217, November, 2018.

Electric Power Systems Research 168 (2019) 210–217

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Parameter estimation of fully regulated synchronous generators using Unscented Kalman Filters

M.A. González-Cagigal, J.A. Rosendo-Macías*, A. Gómez-Expósito

Department of Electrical Engineering, University of Seville, Spain

ARTICLE INFO

Keywords:
Unscented Kalman Filter
Synchronous generator
Speed governor
Automatic voltage regulator
Power system stabilizer
Parameter estimation

ABSTRACT

This paper presents a parameter estimation technique for generation sets including the synchronous machine-turbine pair, along with the customary controllers: speed governor, automatic voltage regulator and power system stabilizer. The proposed technique is based on the Unscented Kalman Filter for the joint estimation of the system dynamic state and a modified set of parameters from which the actual model parameters and constants can be computed. To the authors' knowledge, this work is the first attempt to estimate such a full set of dynamic state variables and parameters, using just external measurements taken at the generator terminal bus.

1. Introduction

The operation and control of electric power systems require a complete understanding of their associated dynamics, described by a set of equations. Gaining accurate knowledge of the parameters involved in those equations is essential in order to identify the system state and make adequate control decisions for the power system.

Usually, those parameters are assumed to be constant, i.e., independent of factors such as exchanged power, terminal voltage, etc., characterizing the operating point. However, this assumption may not be accurate enough, owing to nonlinearities and time dependence (aging), and a method must be established to determine the updated values of system parameters taking the mentioned factors into consideration.

A feasibility study is made in [1], which proves the accuracy of dynamic state estimators (DSEs) based on Kalman Filters (KFs) in power systems with nonlinear dynamics. Studies such as [2] use KFs to estimate the state of a synchronous machine, showing inaccurate estimation results in the presence of model uncertainties. Also in this context, [3] presents a comparison of different DSEs based on KFs, concluding that Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF) are computationally the most efficient for the synchronous machine dynamics. In [4], the EKF is used for state estimation in power systems with unknown inputs, but no parameter estimation is included.

Particularly, among the different KF formulations, UKF has been used in [5] for state estimation of single synchronous machines, and in [6–8] for multi-machine power systems. Parameter estimation is addressed in [9,10], using a set of internal measurements which are difficult to obtain in practice. The dependence of the available measurements and the generator parameters to be estimated using UKF is studied in [11].

To avoid these difficulties, [12] proposes a method, using phasor measurement units (PMUs), for the estimation of the parameters of a synchronous machine. The most significant internal parameters of these machines are also estimated in [13], whereas [14] considers the machine under saturation conditions. The method proposed in [15] broadens the problem by including in the model the parameters of the Speed Governor (SG) and Automatic Voltage Regulator (AVR), studying the effect of a variable error (up to ± 30%) in the initialization of the parameters.

The Power System Stabilizer (PSS) model is considered in [16]. Based on measurements from PMUs, the parameters associated with the generator, SG and AVR are estimated, while PSS parameters are supposed to be known in the proposed UKF implementation.

To our knowledge, this work is the first attempt to consider a fully regulated generation set, including the synchronous machine itself along with the whole set of regulators (SG, AVR and PSS). Both state and parameters from these elements are jointly estimated using UKF and measurements obtained at the generator terminal bus.

This work is organized as follows: Section 2 formulates the equations used in the proposed UKF algorithm. Section 3 presents the modeling of the overall generation set. The implementation of the UKF is described in Section 4. Section 5 presents a case study to test the accuracy of the proposed estimation technique. Finally, the conclusions obtained from the estimation results are presented in Section 6.

* Corresponding author.
E-mail address: rosendo@us.es (J.A. Rosendo-Macías).

https://doi.org/10.1016/j.epsr.2018.11.018
Received 17 July 2018; Received in revised form 23 October 2018; Accepted 28 November 2018

2. Unscented Kalman Filter

Kalman Filter implementations involve a set of state equations, including the dynamic and the measurement equations. In the case of continuous-time, discrete-measurement, non-linear systems, these equations can be expressed as

$$\dot{x}(t) = f(x(t), u(t)) + w(t) \tag{1}$$

$$z(t_k) = g(x(t_k), u(t_k)) + v(t_k) \tag{2}$$

where $x(t)$ is the state vector, $u(t)$ the system input, and $z(t_k)$ the available measurements at instant t_k. The model and measurement noises, $w(t)$ and $v(t_k)$, are assumed Gaussian processes with covariance matrices Q and R, respectively.

Considering a time step Δt, the above equations have the following discrete counterparts:

$$x_k = x_{k-1} + \Delta t \cdot f(x_{k-1}, u_{k-1}) + w_k \tag{3}$$

$$z_k = g(x_k, u_k) + v_k \tag{4}$$

which are more appropriate for non-linear Kalman filtering techniques, like the EKF and UKF.

Linearizing in (3) the state function, f, has yielded poor results when combined with the approximations made by the EKF, owing to the nature of the equations that describe the behavior of the synchronous machine and associated regulators. Instead, this work makes use of the UKF, which is based on an iterative process with two different stages, as summarized in the sequel [17].

2.1. Prediction stage

For each iteration at instant k, a cloud of $2L + 1$ vectors, called σ-points, is calculated from the estimated expected value of the state vector, $\hat{x}_{k-1}$ (dimension L), and the covariance matrix of the state estimation error, P_{k-1}, using the following expression:

$$\begin{cases} x_{k-1}^{0} = \hat{x}_{k-1} \\ x_{k-1}^{i} = \hat{x}_{k-1} + [\sqrt{(L+\lambda)P_{k-1}}]_i \\ x_{k-1}^{i+L} = \hat{x}_{k-1} - [\sqrt{(L+\lambda)P_{k-1}}]_{i+L} \\ i = 1, \ldots, L \end{cases} \tag{5}$$

where $[\sqrt{(L+\lambda)P_{k-1}}]_i$ is the ith column of the matrix $\sqrt{(L+\lambda)P_{k-1}}$, and λ is a scaling factor calculated as

$$\lambda = \alpha^2(L + \kappa) - L \tag{6}$$

with α and κ being two filter parameters to be tuned.

Those σ-points are evaluated in (3) obtaining $2L + 1$ resultant vectors, x_k^{i-}, from which to obtain the *a priori* estimations $\hat{x}_k^-$ and P_k^-:

$$\hat{x}_k^- = \sum_{i=0}^{2L} W_{\mathrm{mi}} x_k^{i-} \tag{7}$$

$$P_k^- = \sum_{i=0}^{2L} W_{\mathrm{ci}} (x_k^{i-} - \hat{x}_k^-)(x_k^{i-} - \hat{x}_k^-)^T + Q_k \tag{8}$$

where the weighting vectors W_m and W_c are calculated as follows:

$$\begin{cases} W_{m0} = \frac{\lambda}{L+\lambda} \\ W_{c0} = \frac{\lambda}{L+\lambda} + 1 - \alpha^2 + \beta \\ W_{\mathrm{mi}} = W_{\mathrm{ci}} = \frac{1}{2(L+\lambda)} \quad i = 1, \ldots, 2L \end{cases} \tag{9}$$

and β is another tunable parameter.

2.2. Correction stage

On the basis of the *a priori* estimations, a new cloud of vectors is calculated,

$$\begin{cases} x_k^{0-} = \hat{x}_k^- \\ x_k^{i-} = \hat{x}_k^- + [\sqrt{(L+\lambda)P_k^-}]_i \\ x_k^{(i+L)-} = \hat{x}_k^- - [\sqrt{(L+\lambda)P_k^-}]_{i+L} \\ i = 1, \ldots, L \end{cases} \tag{10}$$

and evaluated with the measurement function g in (4), yielding

$$\gamma_k^{i-} = g(x_k^{i-}, u_k) \quad i = 0, \ldots, 2L \tag{11}$$

Those values are weighted using the vector W_m defined in (9),

$$\hat{z}_k^- = \sum_{i=0}^{2L} W_{\mathrm{mi}} \gamma_k^{i-} \tag{12}$$

Then, the covariance matrix of the measurement estimation error, P_{zk}^-, and the cross-covariance matrix of state and measurements, P_{xzk}^-, are obtained using the vector W_c, as follows:

$$P_{\mathrm{zk}}^- = \sum_{i=0}^{2L} W_{\mathrm{ci}} (\gamma^{i-} - \hat{z}_k^-)(\gamma_k^{i-} - \hat{z}_k^-)^T + R_k \tag{13}$$

$$P_{\mathrm{xzk}}^- = \sum_{i=0}^{2L} W_{\mathrm{ci}} (x_k^{i-} - \hat{x}_k^-)(\gamma_k^{i-} - \hat{z}_k^-)^T \tag{14}$$

By using the *a priori* predictions at instant k, from (7), (8) and (13), (14), both the Kalman gain,

$$K_k = P_{\mathrm{xzk}}^- (P_{\mathrm{zk}}^-)^{-1} \tag{15}$$

and the respective *a posteriori* predictions can be obtained,

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - \hat{z}_k^-)^T \tag{16}$$

$$P_k = P_k^- - K_k P_{\mathrm{zk}}^- K_k^T \tag{17}$$

which are needed for the next algorithm iteration.

3. System dynamic modeling

The joint estimation of parameters and state variables requires the knowledge of the dynamic equations modeling the behavior of the synchronous machine and associated regulators.

3.1. Dynamic model of the rotating machine

A number of models can be used to represent the dynamics of a synchronous generator. Some of them are excessively complex for the purpose of this work while others do not allow the implementation of voltage control. The two-axis model in [18] is widely used in transient stability studies and has proven to be suitable for this application. After neglecting the rotor resistance, the model equations yield:

$$\dot{\delta} = \Omega_b(\omega - \omega_s) \tag{18}$$

$$\dot{\omega} = \frac{1}{2H}(p_m - p_e - D(\omega - \omega_s)) \tag{19}$$

$$\dot{e}_q' = \frac{1}{T_{d0}'}(-e_q' - (x_d - x_d')i_d + v_f) \tag{20}$$

$$\dot{e}_d' = \frac{1}{T_{q0}'}(-e_d' - (x_q - x_q')i_q) \tag{21}$$

where Ω_b is the base frequency of the system and the state variables related to the generator are:

δ: rotor angle.
ω: rotor angular speed.
$\dot{e}_d'$: transient electromotive force (emf) in d-axis.
$\dot{e}_q'$: transient emf in q-axis.

The electric power, p_e, in (19) is calculated as follows:

Table 1
Synchronous machine parameters.

Symbol	Parameter
D	Damping coefficient
H	Inertia constant
x_q	q-axis synchronous reactance
x_d	d-axis synchronous reactance
x'_q	q-axis transient reactance
x'_d	d-axis transient reactance
T'_{q0}	q-axis transient time constant
T'_{d0}	d-axis transient time constant

Table 2
Speed Governor parameters.

Symbol	Parameter
R	Governor droop
T_r	Primary control time constant
K_i	Secondary control gain

Table 3
Automatic voltage regulator parameters.

Symbol	Parameter
T_{avr}	Controller time constant
T_e	Exciter time constant
K_0	AVR gain

Table 4
Power system stabilizer parameters.

Symbol	Parameter
T_w	PSS time constant
K_w	PSS gain

$$p_e = v_d i_d + v_q i_q \tag{22}$$

$$v_d = \mathrm{V}\sin(\delta - \theta) \tag{23}$$

$$v_q = \mathrm{V}\cos(\delta - \theta) \tag{24}$$

V and θ being the magnitude and phase angle of the voltage at the generator bus.

Note that Eqs. (18)–(21) include a number of parameters, listed in Table 1, to be estimated. The goal of this work is to estimate as many of them as possible, using UKF and resorting to the existing knowledge about their values.

3.2. Speed governor

This paper considers primary and secondary controls for the mechanical power produced by the generator turbine, characterized by [19],

$$\dot{p}_m = \frac{1}{T_r}(p_{\mathrm{ref}} - p_c - p_m - \frac{1}{R}(\omega - \omega_s)) \tag{25}$$

$$\dot{p}_c = K_i(\omega - \omega_s) \tag{26}$$

where the turbine actuation delay has been neglected. The parameters involved in these equations, listed in Table 2, will be also included in the UKF model. After integration, updated values of p_m are used in (19).

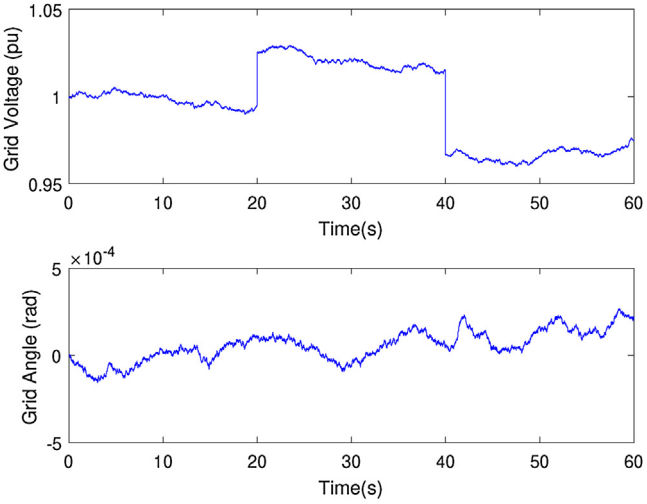


Fig. 2. Evolution of the grid voltage in the base case.

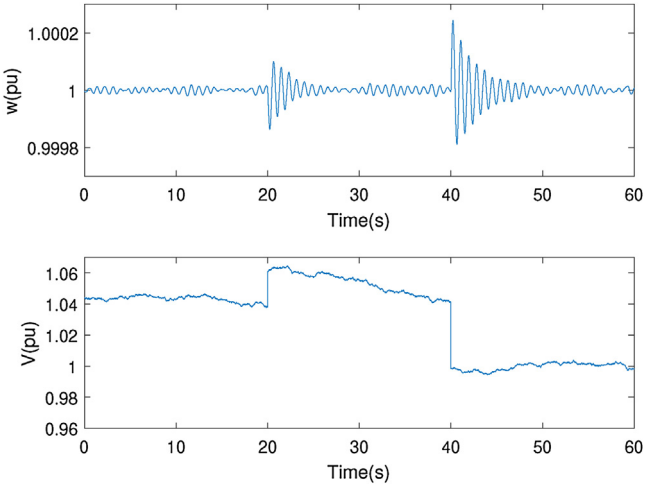


Fig. 3. Evolution of the angular speed and terminal voltage.

3.3. Automatic voltage regulator

The type-III model, providing a reasonable tradeoff between

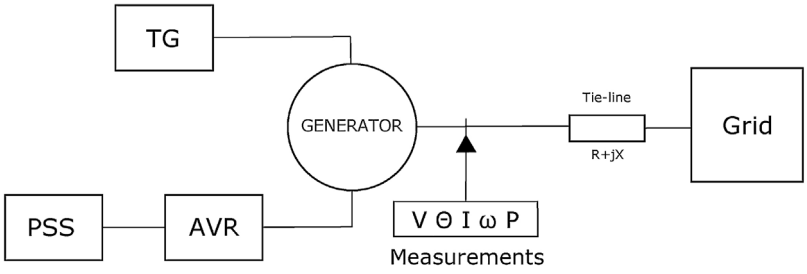


Fig. 1. Representation of the simplified system under study.

Table 5
Parameter simulation values.

Parameter	Unit	Simulated value
x_d	pu	1
x_q	pu	0.65
x'_d	pu	0.3
x'_q	pu	0.55
D	pu	2
H	s	6.5
T'_{q0}	s	0.5
T'_{d0}	s	5
R	pu	0.1
T_r	s	0.1
K_i	pu	50
T_{avr}	s	1
T_e	s	0.2
K_0	pu	2.5
T_w	s	10
K_w	pu	30

Table 6
Initial values of the state variables.

Variable	Initial value
δ	0.427
ω	1
e'_q	1.2
e'_d	0
p_m	0.8
p_c	0
v_f	1.68
v_r	0
v_1	0
v_s	0

accuracy and complexity [20], has been used for the AVR, as follows:

$$\dot{v}_r = \frac{1}{T_{\text{avr}}}(K_0(V_{\text{ref}} + v_s - V) - v_r) \tag{27}$$

$$\dot{v}_f = \frac{1}{T_e}(v_r + v_{f0} - v_f) \tag{28}$$

being v_f the AVR excitement signal, to be used in the emf equation (20), and v_{f0} its initial value. Table 3 collects the AVR parameters, also to be estimated using the UKF.

3.4. Power system stabilizer

For this regulator, which modifies the AVR reference signal, V_{ref}, a simplified type-II model has been implemented, [20]:

$$\dot{v}_1 = -\frac{1}{T_w}(K_w(\omega - \omega_s) + v_1) \tag{29}$$

$$v_s = K_w(\omega - \omega_s) + v_1 \tag{30}$$

Table 4 collects the parameters of this regulator, also to be estimated using UKF.

4. Implementation of UKF

State estimation requires the previous knowledge of the parameters involved in the above equations. However when these parameters are not known, dynamic estimation techniques such as UKF can be used for a joint estimation of state variables and parameters using and augmented state vector [21].

First attempts with a *vanilla* implementation of the joint estimator, directly using the whole set of parameters listed above, led to unacceptable convergence behavior. Therefore, as proposed in [22] for a simpler case with single-axis generator and only frequency regulation, a new modified parameter vector ψ, to be embedded in the augmented state vector, x_a was adopted. This way, $x_a = [x^T, \psi^T]^T$, where x contains the state variables,

$$x^T = [\delta, \omega, e'_q, e'_d, p_m, p_c, v_f, v_r, v_1, v_s]$$

and

$$\psi^T = [\text{dx}_d, \text{dx}_q, x'_d, x'_q, D, H, T'_{d0}, \tfrac{1}{T'_{q0}}, \tfrac{1}{R}, \tfrac{1}{T_r}, \tfrac{100}{K_i}, \tfrac{T_{\text{avr}}}{K_0}, \tfrac{1}{T_e}, \tfrac{1}{T_{\text{avr}}}, \tfrac{100}{K_w}, T_w]$$

where $\text{dx}_d = x_d - x'_d$ and $\text{dx}_q = x_q - x'_q$. So, the size of the augmented state vector is $L = 26$, and the dynamic system model, replacing (3) and (4), is

$$\begin{bmatrix} x_k \\ \psi_k \end{bmatrix} = \begin{bmatrix} x_{k-1} + \Delta t \cdot f(x_{k-1}, u_{k-1}) \\ \psi_{k-1} \end{bmatrix} + w_k \tag{31}$$

$$z_k = g(x_{\text{ak}}, u_k) + v_k \tag{32}$$

where w_k is now the augmented-model noise vector, including not only the state variable components, but also the parameter components.

Five magnitudes can be easily measured at the generator terminal bus, namely: V, θ, ω, I and p_e, where θ is always measured with respect to the synchronous reference frame (for instance, with the help of a PMU) an I is the stator current leaving the generator. Please, note that balanced operating conditions are assumed and, hence, those magnitudes refer to the positive sequence (single-phase model). In the proposed formulation, the five magnitudes are divided into inputs, $u = [V, \theta]$, and measurements, $z = [\omega, I, p_e]$, [23].

The vector z needs to be formulated as a function of the augmented state and input vectors, i.e., the function $g(x_a, u)$ in (32). The first z component, the rotor speed ω, is already a state variable. For the rms value of the stator current, I, and the electric power, p_e, the following expressions are used:

$$I = \sqrt{i_d^2 + i_q^2} \tag{33}$$

$$p_e = v_d i_d + v_q i_q \tag{34}$$

where

$$v_d = \text{Vsin}(\delta - \theta) \tag{35}$$

$$v_q = \text{Vcos}(\delta - \theta) \tag{36}$$

$$i_d = \frac{e'_q - v_q}{x'_d} \tag{37}$$

$$i_q = \frac{-e'_d + v_d}{x'_q} \tag{38}$$

As explained above, the first estimator model, based on the regular set of state variables and parameters, was prone to convergence problems during the initial tests. Trying to circumvent this problem, this work proposes a two-stage estimation technique. In the first stage, the generator parameters are estimated using a simpler model, that considers only the synchronous machine, with no regulators. In other words, the initial estimator comprises the following reduced model,

$$x^T = [\delta, \omega, e'_q, e'_d]$$

and

$$\psi^T = [\text{dx}_d, \text{dx}_q, x'_d, x'_q, D, H, T'_{d0}, \frac{1}{T'_{q0}}]$$

Then, in the second stage, the generador parameters provided by the simpler model are reintroduced in the fully regulated model above, as an initial estimation.

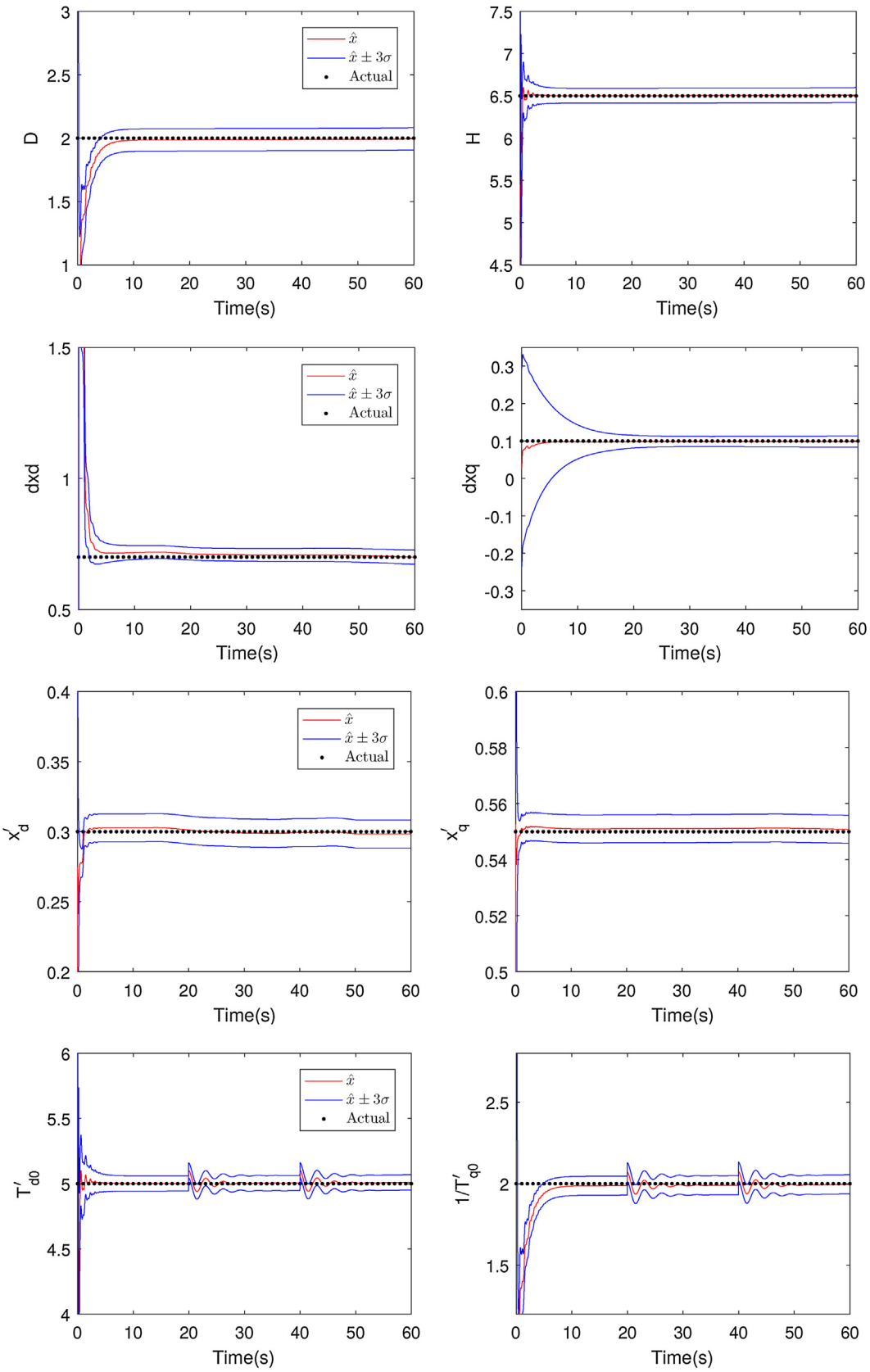


Fig. 4. Estimated parameters for the synchronous generator at the first stage.

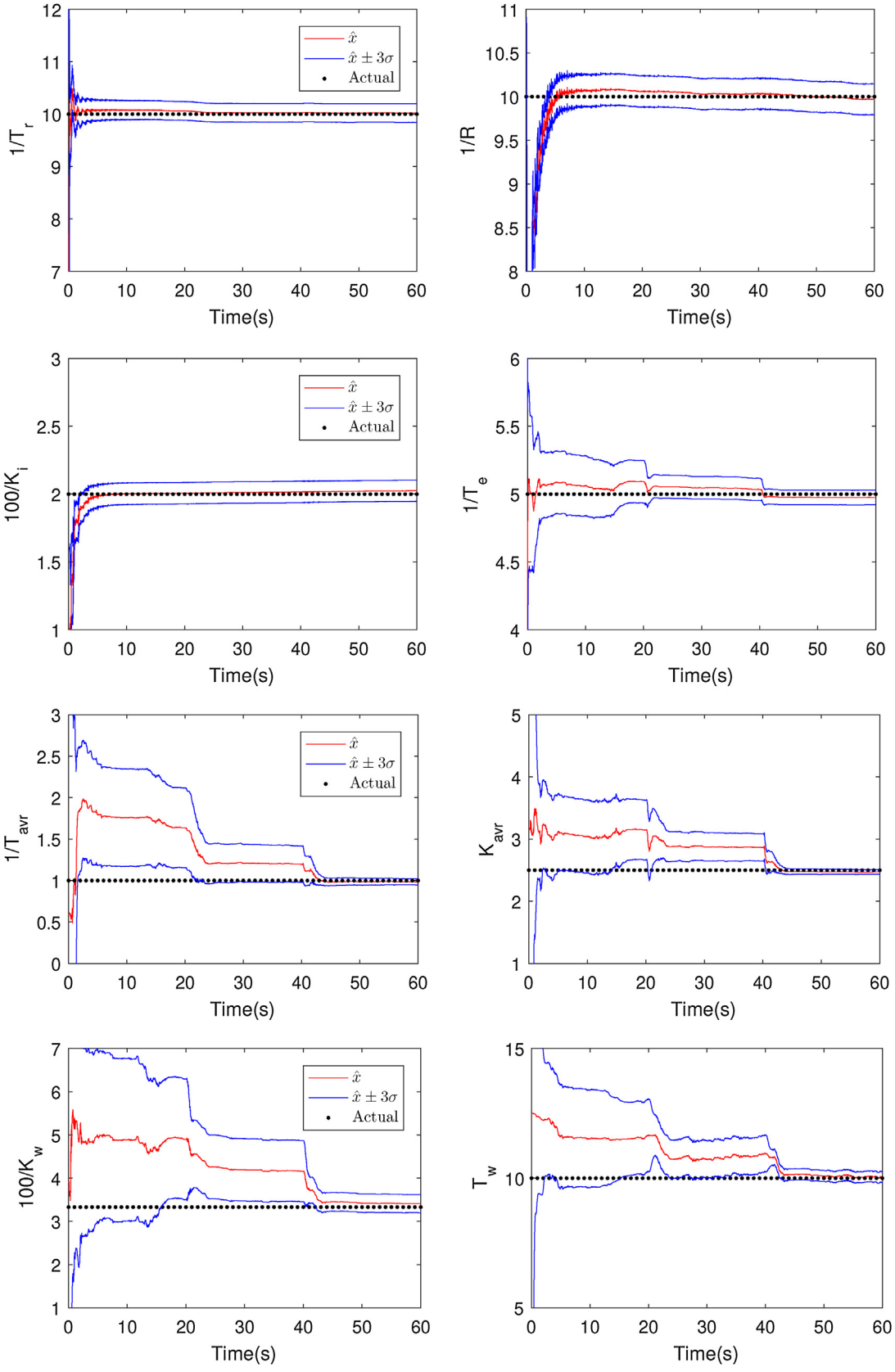


Fig. 5. Estimated parameters for SG, AVR and PSS.

Table 7
Relative error in parameter estimation.

Parameter	Estimated value	Relative error (%)
x_d	0.987	1.30
x_q	0.6583	1.28
x'_d	0.298	0.67
x'_q	0.559	1.43
D	1.999	0.05
H	6.341	2.45
T'_{q0}	0.502	0.40
T'_{d0}	5.117	2.34
R	0.1032	3.20
T_r	0.1006	0.64
K_i	49.75	0.498
T_{avr}	0.988	1.186
T_e	0.2002	0.101
K_0	2.479	0.84
T_w	10.135	1.35
K_w	30.78	2.596

Table 8
Maximum relative error for each standard deviation R_w.

R_w	10^{-6}	10^{-5}	10^{-4}	10^{-3}
$\lvert E_r^{\max}\rvert(\%)$	3.05	3.20	4.15	5.01

5. Case study

In this section, the performance of the proposed method is tested with the reduced power system shown in Fig. 1. The simulation has been carried out using Matlab Simulink. The model considered for the generation of measurements, including the rotor resistance, the turbine actuation delay and a set of control parameters in the PSS model, is more accurate than that adopted by the UKF algorithm [20].

In this case study, the generator is connected to an infinite busbar through a tie-line with impedance $z_L = 0.01 + j0.1$ pu. The voltage at the infinite busbar is modeled with magnitude and angle evolving smoothly, as Gaussian random walks with standard deviations $R_w = 10^{-5}$. Two abrupt voltage magnitude steps (0.03 pu upwards and 0.05 downwards) are also applied, as shown in Fig. 2. Moreover, initial values of active and reactive power contributions to the infinite busbar are assumed, with complex power $s = 0.8 + j0.4$ pu, so that p_{ref} and V_{ref} can be accordingly obtained.

As an example, the evolution of ω and V under these abrupt voltage magnitude variations is represented in Fig. 3.

The system base frequency is 50 Hz, so $\Omega_b = 100\pi$ rad/s and $\omega_s = 1$. The parameters adopted in the simulation models are listed in Table 5, and the time step for simulation and UKF is $\Delta t = 0.01$ s.

The UKF requires an initial state for x_a. The initial values for the state variables, x, are obtained by solving the steady-state case, which is given in Table 6.

For the modified parameter vector adopted, ψ, random initial values are considered in the range ± 20% to ± 40% of their real values. In the second stage of the proposed two-stage estimation technique, the generator parameters are initialized with their estimated value from the first stage.

The covariance of the initial estimation error is defined by a diagonal matrix $P_0 = \mathrm{diag}([P_{x0}^T, P_{\psi 0}^T])$, where the vector P_{x0} corresponds to the state and $P_{\psi 0}$ to the parameters. For the elements of P_{x0}, increasing values are adopted, from the generator variables to those of the PSS, resulting the vector:

$$P_{x0}^T = [10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-3}, 10^{-3}, 10^{-2}, 10^{-2}, 1, 1]$$

The elements of vector $P_{\psi 0}$ for the parameters of the machine and its regulators are tuned in order to achieve a better estimation, obtaining the following result:

$$P_{\psi 0}^T = [10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 5, 5, 5, 1, 1, 1, 10, 10]$$

In this work, the UKF has been implemented with $\alpha = 10^{-4}$, $\kappa = 0$, $\beta = 2$, following the values commonly found in works such as [24], where an analysis of the influence of these scaling parameters in the estimation process is made. The covariance matrix R has been taken as a diagonal matrix with $R_{ii} = 10^{-4}$, corresponding to an error with 1% s.d. (lower accuracy than that typically provided by PMUs).

The covariance matrix Q is assumed to be of the form $Q = \mathrm{diag}([Q_x^T, Q_\psi^T])$, where the vector Q_x corresponds to the state and Q_ψ to the parameters. For the elements of Q_x, a common value of 10^{-6} is considered, while Q_ψ is proposed to have increasing values, from the generator parameters to those of the PSS, resulting the vector:

$$Q_\psi^T = [10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-5}, 10^{-4}, 10^{-4}]$$

The proposed algorithm presents a consistent performance in its ability to properly estimate the model modified parameters. Fig. 4 shows the estimation results for the generator modified parameters, as obtained from the first stage of the proposed technique. The graphics include the evolution of the estimation error covariance. For each parameter, the estimated value, $\hat{x}_i$, is represented jointly with a deviation $\hat{x}_i \pm 3\sqrt{P_{ii}}$. Note that the covariances tend to Q_{ii}, showing the accuracy of the converged estimation.

In the second stage, the synchronous generator parameters show no significant evolution from their initial values, obtained at the first stage, but the model modified parameters corresponding to the SG, the AVR and the PSS evolve successfully to the real values, as shown in Fig. 5.

Hence, the method achieves a correct estimation of the modified parameters, allowing the final computation of the original parameters in (18)–(30).

The relative estimation error of the original parameters is summarized in Table 7.

The robustness of the proposed technique for parameter estimation has been tested with different standard deviations of the Gaussian random walks on the grid voltage, R_w. For each case, the maximum relative error in the estimated parameters, $\lvert E_r^{\max}\rvert$, is shown in Table 8.

6. Conclusions

In this paper, an UKF implementation is proposed to perform the joint estimation of the state variables and parameters of a fully regulated two-axis synchronous machine, including SG for the frequency control and AVR and PSS for voltage control.

Although many techniques, including UKF, have already been used for the estimation of generator parameters, the contribution of this work lies in the use of a modified set of parameters, comprising up to 16 parameters, which are estimated using only 5 external measurements taken at the generator terminal bus. Also, the use of a two-stage cascaded estimation process has made it possible to deal with bad convergence behavior affecting some generator parameters.

A case study has shown that the proposed estimator yields accurate enough results when using different initialization strategies. Four performance tests with different standard deviations of the random walks in the grid voltage show similar relative errors on the estimated values.

Acknowledgement

The first author thanks the Ministry of Education and Professional Training of Spain for the financial support (grant FPU17/06380).

References

[1] Z. Huang, K. Schneider, J. Nieplocha, Feasibility studies of applying Kalman Filter techniques to power system dynamic state estimation, International Power Engineering Conference (IPEC), December, 2007.

[2] P. Ren, H. Lev-Ari, A. Abur, Robust continuous-discrete extended Kalman filter for estimating machine states with model uncertainties, Power Systems Computation Conference (PSCC), June, 2016.

[3] Z. Ning, M. Da, H. Zhenyu, G. Welch, Dynamic state estimation of a synchronous machine using PMU data: a comparative study, IEEE Power & Energy Society General Meeting, January, 2015.

[4] E. Ghahremani, I. Kamwa, Dynamic state estimation in power system by applying the extended Kalman filter with unknown inputs to phasor measurements, IEEE Trans. Power Syst. 26 (November (4)) (2011).

[5] S. Wang, W. Gao, A.P. Sakis Meliopoulos, An alternative method for power system dynamic state estimation based on unscented transform, IEEE Trans. Power Syst. 27 (May (2)) (2012).

[6] J. Qi, K. Sun, J. Wang, H. Liu, Dynamic state estimation for multi-machine power system by unscented Kalman filter with enhanced numerical stability, IEEE Trans. Smart Grid (March) (2016).

[7] F.M. Taimah, A.N. Merzah, Power system dynamic state estimation based on Kalman filter, Int. J. Comput. Appl. (0975-8887) 154 (November (11)) (2016).

[8] J. Zhao, L. Mili, Robust unscented Kalman filter for power system dynamic state estimation with unknown noise statistics, IEEE Trans. Smart Grid (October) (2017).

[9] A. Rouhani, A. Abur, Constrained iterated unscented Kalman filter for dynamic state and parameter estimation, IEEE Trans. Power Syst. 33 (May (99)) (2018) 1, https://doi.org/10.1109/TPWRS.2017.2764005.

[10] M. Huang, W. Li, W. Yan, Estimating parameters of synchronous generators using square-root unscented Kalman filter, Electr. Power Syst. Res. 80 (September (9)) (2010) 1137–1144.

[11] G. Valverde, E. Kyriakides, V. Terzija, A non-linear approach for on-line parameter estimation of synchronous machines, 17th Power Systems Computation Conference, August, 2011.

[12] Y. Wehbe, L. Fan, UKF based estimation of synchronous generator electromechanical parameters from phasor measurements, North American Power Symposium (NAPS), September, 2012.

[13] E.L. Geraldi, T.C.C. Fernandes, R.A. Ramos, An approach to estimate parameters of a three-phase synchronous generator model, 2016 IEEE Power and Energy Society General Meeting (PESGM), November, 2016.

[14] G. Valverde, E. Kyriakides, G.T. Heydt, V. Terzija, On-line parameter estimation of saturated synchronous machines, 2011 IEEE Power and Energy Society General Meeting, October, 2011.

[15] J.C.N. Pantoja, A. Olarte, H. Daz, Simultaneous estimation of exciter governor and synchronous generator parameters using phasor measurements, Electric Power Quality and Supply Reliability Conference (PQ), June, 2014, pp. 43–49.

[16] S. Dutta Chowdhury, N. Senroy, PMU data based online parameter estimation of synchronous generator, International Conference on Power Systems (ICPS), March, 2016.

[17] S Simon D., Optimal State Estimation: Kalman H Infinity and Nonlinear Approaches. ISBN: 13 978-0-471-70858-2.

[18] Glover J., Sarma M., Overbye T., Power Systems Analysis and Design. ISBN-13: 978-1-111-42579-1.

[19] Sauer P.W., Pai M.A., Chow J.H., Power System Dynamics and Stability: With Synchrophasor Measurement and Power System Toolbox 2e. ISBN: 9781119355779.

[20] Milano F., Power System modelling and scripting. ISBN: 978-3-642-13668-9.

[21] E.A. Wan, R. Van der Merwe, The unscented Kalman filter for nonlinear estimation, Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000, October, 2000, pp. 153–158.

[22] H. Ghassempour, Z. Miao, L. Fan, W. Jiang, D. Manjure, Identification of synchronous generator model with frequency control using unscented Kalman filter, Electr. Power Syst. Res. 126 (2015) 45–55.

[23] S. Julier, The scaled unscented transformation, American Control Conference, 2002. Proceedings of the 2002, IEEE Press, 2002, pp. 4555–4559.

[24] J.J. Hyun, L. Hyung-Chul, Analysis of scaling parameters of the batch unscented transformation for precision orbit determination using satellite laser rangingdata, J. Astron. Space Sci. 28 (2011) 183–192, https://doi.org/10.5140/JASS.2011.28.3.183.

Annex II.2

M.A. González-Cagigal, J.A. Rosendo-Macías, A. Gómez-Expósito, "Parameter Estimation of Wind Turbines with PMSM using Cubature Kalman Filters," in: IEEE Transactions on Power Systems, Volume .35, 2020, pp. 1796-1804, May 2020.

Parameter Estimation of Wind Turbines With PMSM Using Cubature Kalman Filters

M. A. González-Cagigal, José A. Rosendo-Macías, *Senior Member, IEEE*, and A. Gómez-Expósito, *Fellow, IEEE*

***Abstract*—This paper presents a parameter estimation technique for a variable-speed wind turbine with permanent magnet synchronous generator and back-to-back voltage source converter. The proposed technique applies the cubature Kalman filter for the joint estimation of the system dynamic state and a modified set of parameters from which the original model parameters can be algebraically recovered. To the authors knowledge, this work is the first attempt to apply such an innovative technique to a PMSM detailed estimation model including the control parameters of the voltage source converter.**

***Index Terms*—Cubature Kalman filter, variable-speed wind turbine, direct-drive synchronous generator, back-to-back voltage source converter, parameter estimation.**

I. Introduction

In a decarbonized and more electrified future, power systems must be able to cope with increasing amounts of intermittent renewable energy. Excluding hydro, wind energy is so far the dominant renewable source worldwide, both in terms of production share (5.6%) and cumulative installed power (550 GW), of which only about 20 GW (less than 4%) correspond to offshore farms. This means that there exists a huge growth potential for offshore wind technology, still in its infancy, while the best places for onshore farms are being quickly occupied. The average offshore turbine size in Europe (5.9 MW) has nearly doubled in the last decade and some vendors are already announcing machine designs of over 12 MW rated power [1]. The most promising topology for offshore wind energy is the direct-driven, multi-pole Permanent Magnet Synchronous Machine (PMSM) with fully-rated back-to-back Voltage Source Converters (VSC), a turbine concept which lacks the gearbox.

Among the upcoming technical challenges raised by the massive integration of renewables, the need for wind and PV sources to contribute to ancillary services stands out. These include voltage and frequency regulation, according to increasingly demanding grid codes, provision of synthetic inertia to keep current stability margins, fast response (flexibility) in the presence of more frequent and deeper net demand gradients, etc.

In this context, it is most important for grid operators to adopt accurate enough PMSG-based wind turbine models, including the fast acting VSC, in order to evaluate their dynamic behavior. This involves a detailed knowledge of the associated components and their defining equations. Ultimately, the system state evolution is determined by the parameters involved in those equations, which are therefore crucial for the correct operation and control of a power system with a high penetration of wind power plants.

System parameters are customarily considered constant, even under changing operating conditions. However, when this assumption is not accurate enough, or when the control parameters provided by the manufacturer are suspected to be inaccurate or outdated, a model validation is needed to obtain their values considering the actual operating point. This need has been recognized, for instance, in the US, where the NERC has released reliability guidelines [2] regarding the validation of generator models, including synchronous machines and other inverter-based ones exceeding a certain rated power. NERC suggests two ways for generator model validation and calibration: 1) taking the generator out of service and performing specific tests; 2) measurement-based methods based on disturbance recordings from synchrophasors (PMUs), data loggers, fault recorders, etc. This work lies in the second category.

Dynamic state estimators (DSEs) based on Kalman filters (KFs) have been used for state estimation of power systems with nonlinear dynamics, as proposed in [3], which includes a comparison study. A particular formulation of KF, the so-called unscented KF (UKF) has proven to be reliable and accurate for this particular application [4]. For instance, in [5] a method to deal with disturbances in the system is proposed, whereas [6] includes the system parameters in the estimation process.

The UKF technique has also been applied in state estimation of synchronous generators, [7], and joint state and parameter estimation of these machines, [8]. Regarding wind turbines, [9] studies the application of UKF in fault diagnosis. For this type of machines, [10] uses extended KF (EKF) in the estimation of the electrical parameters of doubly-fed induction generators (DFIGs), while [11] proves that the performance of the UKF is superior to that of the EKF in the parameter estimation of DFIGs.

EKF is also used in different studies on permanent magnets synchronous motors. Rotor initial position is estimated in [12]

Manuscript received November 28, 2018; revised July 31, 2019; accepted September 21, 2019. Date of publication October 7, 2019; date of current version April 22, 2020. This work was supported in part by the Spanish Ministry of Economy and Competitiveness under Grant PCIN-2015-043 and in part by the Ministry of Education and Professional Training under Grant FPU17/06380. Paper no. TPWRS-01796-2018. *(Corresponding author: José A. Rosendo-Macías.)*

The authors are with the Department of Electrical Engineering, University of Seville, Seville 41092, Spain (e-mail: mgcagigal@us.es; rosendo@us.es; age@us.es).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2019.2945778

for a sensorless direct torque controlled machine. Flux-linkage is included in the estimation process in [13] for a vector control and in [14] when demagnetization situations are considered. On these machines, fault detection is necessary for a correct operation of the power system. A method using EKF is established in [15] and [16] for stator winding inter-turn short circuit identification.

A real implementation of the studied estimation techniques requires noninvasive measurements taken from the system considered. In this respect, [17] proposes a temperature estimation for PMSMs using noninvasive Kalman filters.

PMSGs are considered in a number of studies dealing with EKF. For instance, in [18] an estimation of the rotor speed is performed within a sensorless maximum power point tracker. Control systems without encoder are also considered in [19], where EKF is used to estimate the state variables of a back-to-back converter. Measurements obtained from PMUs are introduced in the EKF algorithm in [20].

A recent formulation of KFs, the so-called cubature Kalman filter (CKF), has shown good performance in problems such as state estimation of synchronous generators [21]. Theoretical aspects of this formulation are addressed in [22], where some limitations of other KF schemes, such as UKF, which are not suffered by the CKF, are highlighted. Joint state and parameter estimation using CKF is studied in [23], applied to a vehicle model, and in [24], with permanent magnet synchronous motors.

This work considers a variable-speed PMSM wind turbine, connected to the grid through a back-to-back pair of VSCs, which is capable of controlling its active and reactive power within specified limits. Based on measurement snapshots taken at the point of connection, the state and parameters of the wind turbine and the VSC are jointly estimated using CKF.

This work is organized as follows. Section II formulates the equations used in the proposed CKF algorithm. Section III analyzes the modeling system under study, with the wind turbine, the synchronous generator and the back-to-back voltage source converter (VSC). The implementation of the CKF is described in Section IV. Section V presents simulation results corresponding to the base case and several scenarios facing different disturbances. Section VI includes a comparison of the performance presented by the CKF estimation technique when increasing measurement and model errors are considered. Finally, the conclusions obtained are presented in Section VII.

II. Cubature Kalman Filter

Kalman Filter implementations require a set of state equations, including the dynamic and the measurement equations. In the case of continuous-time, discrete-measurement non-linear systems, these equations can be expressed as

$$\dot{x}(t) = f(x(t), u(t)) + w(t) \tag{1}$$

$$z(t_k) = g(x(t_k), u(t_k)) + v(t_k) \tag{2}$$

where $x(t)$ is the state vector, $u(t)$ the system input, and $z(t_k)$ the available measurements at instant t_k. The model and measurement noises, $w(t)$ and $v(t_k)$, are assumed to be Gaussian processes with covariance matrices Q and R, respectively.

Considering a time step Δt, the above equations have the following discrete counterparts:

$$x_k = x_{k-1} + \Delta t \cdot f(x_{k-1}, u_{k-1}) + w_k \tag{3}$$

$$z_k = g(x_k, u_k) + v_k \tag{4}$$

which are more appropriate for non-linear Kalman filtering techniques, such as the CKF. This involves an iterative process composed of two different stages, as follows [25].

A. Time Update

At each time k, an estimated state vector $\hat{x}_{k-1}$ of size L, along with the covariance matrix associated to its estimation error, P_{k-1}, are available from the previous step. On the basis of these values, a set of $2L$ cubature points are calculated as follows:

$$S_{k-1}S_{k-1}^T = P_{k-1} \tag{5}$$

$$x_{k-1}^i = S_{k-1}\xi_i\sqrt{L} + \hat{x}_{k-1} \quad i = 1, \ldots, 2L \tag{6}$$

where S is a positive-definite square root of matrix P (in this paper the Cholesky factorization of matrix P will be used), and ξ_i is the i^{th} cubature node, obtained as the intersection of the unit sphere and the $\mathbb{R}^L$ axis. Compared to other KF formulations, the cubature points are less prone than the σ-points calculated in the UKF to numerical inaccuracy or filter instability, [22]. This issue is further discussed in Section V.

The state function $f(\cdot)$ in (3) is evaluated at these cubature points, yielding a set of $2L$ vectors x_k^{i-}, from which an *a priori* estimation of $\hat{x}_k^-$ and P_k^- is in turn computed as follows:

$$\hat{x}_k^- = \frac{1}{2L}\sum_{i=1}^{2L} x_k^{i-} \tag{7}$$

$$P_k^- = \frac{1}{2L}\sum_{i=1}^{2L} x_k^{i-}x_k^{i-T} - \hat{x}_k^-\hat{x}_k^{-T} + Q_k \tag{8}$$

B. Measurement Update

Once the *a priori* estimation is obtained, the covariance matrix P_k^- is factorized in order to calculate the matrix S_k^-,

$$S_k^- S_{k-1}^{-T} = P_k^- \tag{9}$$

and a new set of $2L$ cubature points,

$$x_k^{i-} = S_k^-\xi_i\sqrt{L} + \hat{x}_k^- \quad i = 1, \ldots, 2L \tag{10}$$

at which function $g(\cdot)$ in (4) is evaluated, yielding:

$$\gamma_k^{i-} = g(x_k^{i-}, u_k) \quad i = 1, \ldots, 2L \tag{11}$$

Then the measurement estimation, $\hat{z}_k^-$, its covariance matrix, P_{zk}^-, and the cross-covariance matrix of state and measurements, P_{xzk}^-, are calculated as follows:

$$\hat{z}_k^- = \frac{1}{2L}\sum_{i=1}^{2L} \gamma_k^{i-} \tag{12}$$

$$P_{zk}^- = \frac{1}{2L}\sum_{i=1}^{2L} \gamma_k^{i-}\gamma_k^{i-T} - \hat{z}_k^-\hat{z}_k^{-T} + R_k \tag{13}$$

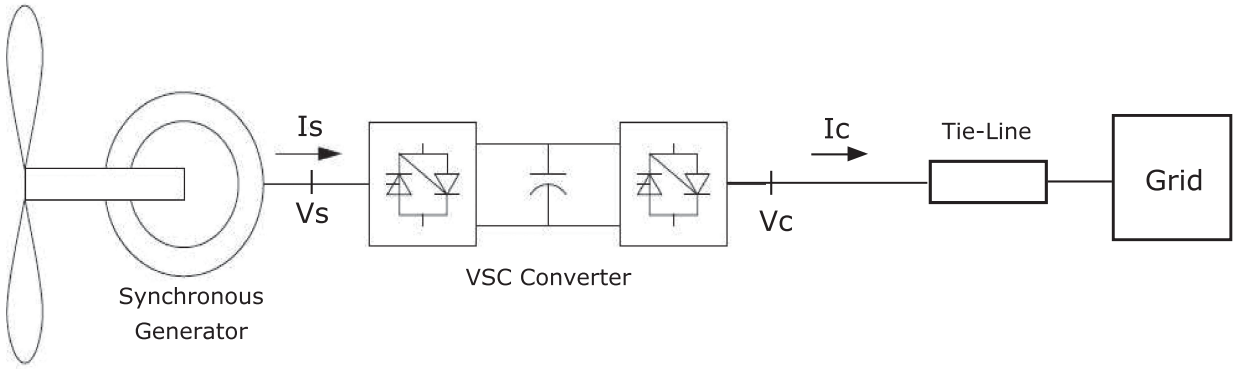


Fig. 1. Components of the simplified system under study.

$$P_{xzk}^{-} = \frac{1}{2L}\sum_{i=1}^{2L} x_k^{i-}\gamma_k^{i-T} - \hat{x}_k^{-}\hat{z}_k^{-T} \tag{14}$$

allowing the cubature Kalman gain to be obtained from,

$$K_k = P_{xzk}^{-}(P_{zk}^{-})^{-1} \tag{15}$$

Finally, the *a posteriori* predictions of the state vector, $\hat{x}_k$, and the covariance P_k are obtained as

$$\hat{x}_k = \hat{x}_k^{-} + K_k(z_k - \hat{z}_k) \tag{16}$$

$$P_k = P_k^{-} - K_k P_{zk}^{-} K_k^T \tag{17}$$

which enter the next iteration of the algorithm.

III. System Modeling

This section presents the equations describing the dynamics of the system under study, composed of the wind turbine with pitch angle control, the synchronous generator and the back-to-back VSC (Fig. 1).

A. Wind Turbine

The mechanical power produced by the turbine at wind speed v_w is given by, [26]:

$$p_w = \frac{1}{2S_n} c_p(\lambda)\pi R^2 v_w^3 \tag{18}$$

where R is the rotor radius, S_n the system rated power and $c_p(\lambda)$ the performance coefficient. For a turbine model with control of the pitch angle, θ_p, and a given tip speed ratio, λ, this coefficient is obtained from,

$$c_p(\lambda) = 0.22\left(\frac{116}{\lambda_i} - 0.04\theta_p - 5\right)e^{-\frac{12.5}{\lambda_i}} \tag{19}$$

$$\frac{1}{\lambda_i} = \frac{1}{\lambda + 0.08\theta_p} - \frac{0.035}{\theta_p^3 + 1} \tag{20}$$

$$\lambda = \frac{2\omega R}{n_{pole} v_w} \tag{21}$$

ω is the shaft angular speed and n_{pole} is the number of poles of the generator.

The pitch angle control is aimed at maximizing the power production for a specified angular speed. In this work, the following optimal power, p_w^{opt}, characteristic for each wind turbine, is considered,

$$p_w^{opt} = \begin{cases} 0 & \text{if} \quad \omega < 0 \\ 2\omega - 1 & \text{if} \quad 0 \le \omega \le 1 \\ 1 & \text{if} \quad \omega > 1 \end{cases} \tag{22}$$

while the control of θ_p is represented with first order dynamics:

$$\dot{\theta}_p = \frac{1}{T_p}\left(K_p(\omega - \omega^{ref}) - \theta_p\right) \tag{23}$$

ω^{ref} being the reference angular speed.

To connect the turbine and the generator, a rigid shaft is considered, so that the equation describing the angular speed dynamics can be expressed as follows:

$$\dot{\omega} = \frac{1}{2H_{tm}}\left(\frac{p_m - p_e}{\omega}\right) \tag{24}$$

where p_e is the electric power comsumption.

Equations (23)–(24) include a set of dynamic parameters, listed in Table I, whose values are unknown. The goal of this work is to estimate those parameters using a CKF.

TABLE I
Wind Turbine Parameters

Symbol	Parameter
T_p	Pitch angle control Time Constant
K_p	Pitch angle control Gain
H_{tm}	Turbine-rotor inertia constant

B. Synchronous Generator

In this work, a permanent magnet synchronous generator is considered, where the voltage (v_s), and current (i_s), at the machine terminals are related through the electromagnetic equations in dq axis, [26],

$$v_{sd} = \omega L_q i_{sq} \tag{25}$$

$$v_{sq} = -\omega(L_d i_{sd} - \psi_p) \tag{26}$$

L_d and L_q being the generator inductances in dq axis, and ψ_p the rotor permanent field flux.

These magnitudes are used to calculate the active and reactive power delivered by the synchronous generator,

$$p_s = v_{sd} i_{sd} + v_{sq} i_{sq} \tag{27}$$

TABLE II
BACK-TO-BACK VSC PARAMETERS

Symbol	Parameter
T_{qs}	Generator side q-axis Time Constant
K_{ds}	Generator side d-axis Gain
T_{ds}	Generator side d-axis Time Constant
K_{qc}	Grid side q-axis Gain
K_{dc}	Grid side d-axis Gain
T_{dc}	Grid side d-axis Time Constant

$$q_s = v_{sq} i_{sd} - v_{sd} i_{sq} \tag{28}$$

As the machine is supposed to be rigidly connected to the turbine, p_s replaces p_e in equation (24).

In this work, the value of the generator inductances, L_d and L_q, which can be obtained with different factory tests, are assumed to be known and therefore are not included in the parameter estimation problem.

C. Back-to-back VSC

The converter allows controlling the active power and either the reactive power produced by the generator or the voltage at the VSC external bus. The control variables are the currents on the VSC generator and grid sides, i_s and i_c, respectively, in $d-q$ axis. The dynamic equations are represented as follows [26]:

$$\frac{di_{sq}}{dt} = \frac{1}{T_{qs}} \left(\frac{p_w^{opt}}{\omega(\psi_p - L_d i_{sd})} - i_{sq} \right) \tag{29}$$

$$\frac{di_{sd}}{dt} = \frac{1}{T_{ds}} \left(K_{ds}(q_{s0} - q_s) - i_{sd} \right) \tag{30}$$

$$\frac{di_{cq}}{dt} = K_{qc}(p_s - p_c) \tag{31}$$

$$\frac{di_{cd}}{dt} = \frac{1}{T_{dc}} \left(K_{dc}(v^{ref} - v_h) - i_{cd} \right) \tag{32}$$

Table II collects the parameters of the converter to be estimated using CKF.

Equations (29)–(32) represent the VSC model as implemented by the CKF algorithm, which constitutes a simplification of the simulation model adopted for the sole purpose of obtaining measured magnitudes, [28]. This latter model includes other components which are present in real wind power plants, such as the AC filter and the DC-link voltage controller, so that the measurements used in the estimation process are as close as possible to those that would be captured from the actual system.

IV. IMPLEMENTATION OF THE CKF

In the previous section, the parameters involved in the dynamic model (18)–(32) have been presented. As most of those parameters are not precisely known, or are simply unknown, they should be included in an augmented state vector for a joint state-parameter estimation using CKF.

An initial attempt to estimate the original raw parameters was unsuccessful, owing to convergence problems. Therefore, a parameter modification technique is proposed, similar to the one adopted in [27] for synchronous machine parameter estimation. With this technique, the augmented state vector becomes $x_a^T = [x^T, \psi^T]$, where the state vector, x, contains the following state variables,

$$x^T = [i_{sd}, i_{sq}, \omega, \theta_p, i_{cd}, i_{cq}]$$

and ψ is the modified parameter vector, as proposed in this work,

$$\psi^T = \left[\frac{10}{H_{tm}}, \frac{K_{qc}}{10}, \frac{1}{T_{dc}}, K_{dc}, \frac{1}{T_{ds}}, \frac{1}{T_{qs}}, K_{ds}, \frac{1}{T_p}, K_p \right]$$

Therefore, the size of the augmented vector is $L = 15$.

Equations (3) and (4) can be rewritten in terms of x_a and an augmented-model noise vector, w_k, involving the state and parameter components, yielding:

$$\begin{bmatrix} x_k \\ \psi_k \end{bmatrix} = \begin{bmatrix} x_{k-1} + \Delta t \cdot f(x_{k-1}, u_{k-1}) \\ \psi_{k-1} \end{bmatrix} + w_k \tag{33}$$

$$z_k = g(x_{ak}, u_k) + v_k \tag{34}$$

Regarding the measurement update stage in the CKF algorithm, seven easily measurable magnitudes are considered: the magnitude and phase angle of the voltage (V, θ_V) and current (I, θ_I) at the VSC external bus, which can be provided by a PMU unit, the shaft angular speed ω, the pitch angle θ_p and the wind speed v_w. In turn, it is customary to divide this set into inputs, $u^T = [v_w, V, \theta_V]$, and measurements, $z^T = [\omega, I, \theta_I, \theta_p]$, [29].

Then, the function $g(x_{ak}, u_k)$ in (34) reduces in this case to,

$$I = \sqrt{i_{cd}^2 + i_{cq}^2} \tag{35}$$

$$\theta_I = \arctan\left(\frac{i_{cd}}{i_{cq}} \right) \tag{36}$$

Note that ω and θ_p lead to trivial expressions, as they are state variables.

V. CASE STUDIES

In this section, the results of three performance tests are presented to confirm the accuracy and robustness of the proposed estimation method, when the reduced power system shown in Fig. 1 faces different disturbances. The magnitudes defining the steady-state starting point (v^{ref} and ω^{ref}) are obtained by assuming an initial complex power injection at the connection point ($S_c = 0.7 + j0.5$ pu), the base power being equal to the rated power of the wind turbine ($S_B =; 2$ MW). The simulations have been carried out using Matlab Simulink. Note that the simulation model considered for the generation of exact magnitudes, [26], is more detailed and accurate than that adopted by the CKF algorithm. Specifically, the stator resistance and a set of control parameters involved in the VSC and the pitch angle control, are neglected in the CKF system model described in Section III. This is intentionally done in order to consider the more realistic situation in which the model assumed by the CKF does not necessarily reflect the exact model, but a simplified one.

The exact parameter values adopted in the simulation are listed in Table III, excluding those that are not considered in the estimation process, such as the stator resistance. The time step for the simulations and the CKF process is $\Delta t = 0.01$ s.

The CKF estimator requires an initial state for x_a. For the state variables, x, an initialization process is used, based on the

TABLE III
EXACT PARAMETER VALUES

Parameter	Unit	Exact Value
H_{tm}	s	4
K_{qc}	pu	35
T_{dc}	s	0.5
K_{dc}	pu	1.5
T_{ds}	s	0.5
T_{qs}	s	0.5
K_{ds}	pu	1.5
T_p	s	3
K_p	pu	2

TABLE IV
INITIAL VALUES FOR THE STATE VARIABLES

Variable	Initial Value
i_{sd}	1
i_{sq}	1.07
ω	0.7
θ_p	0
i_{cd}	0.4
i_{cd}	0.75

steady-state point [30], leading to the initial values shown in Table IV.

In order to test worst case conditions, regarding the accuracy of the available parameter values (from manufacturer data, similar machines or other sources), the proposed modified parameter vector, ψ, is initialized with a random value in the range $\pm 30\% \div \pm 50\%$ of their real values. This way, the proposed estimation technique becomes a method for the validation of the parameter values.

The initial covariance of the state estimation error is defined as a diagonal matrix,

$$P_0 = diag([P_{x0}^T, P_{\psi 0}^T])$$

where

$$P_{x0}^T = [10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}, 10^{-4}]$$

corresponds to the state variables, and

$$P_{\psi 0}^T = [1, 1, 1, 1, 1, 1, 1, 1]$$

to the modified parameters.

The covariance matrix Q has been assumed as a diagonal matrix with $Q_{ii} = 10^{-8}$, while the covariance matrix R has been taken as a diagonal matrix with $R_{ii} = 10^{-4}$, corresponding to an error with 1% s.d. (lower accuracy than that typically provided by PMUs).

A. Base Case: Smooth Operating Point Variations

In this scenario, the impedance of the line connecting the PMSM generator to the infinite busbar is $z_L = 0.01 + j0.1$ pu. The magnitude and angle of the busbar voltage evolve as Gaussian random walks with standard deviation $R_w = 10^{-4}$, as represented in Fig. 2.

The system under study is supposed to be in normal operating conditions. The evolution considered for the wind speed, v_w, is taken at a mean value of 16 m/s with a standard deviation of

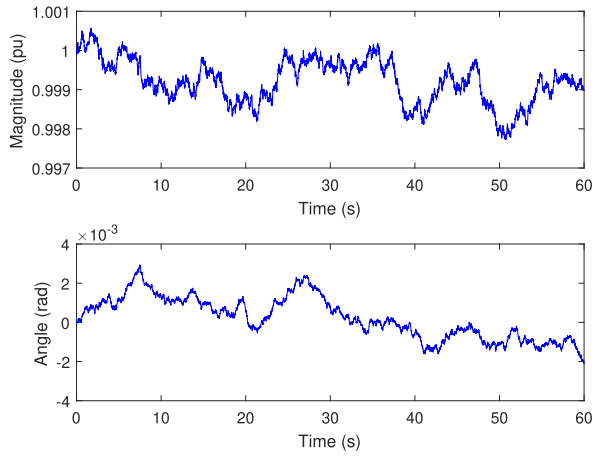


Fig. 2. Gaussian random walks applied to the connection point bus voltage.

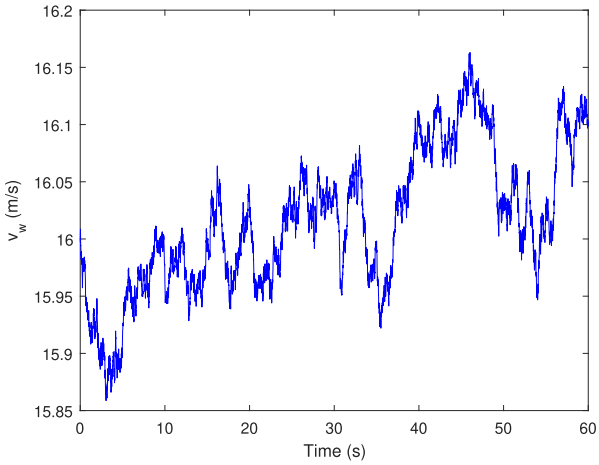


Fig. 3. Wind speed evolution for the base case scenario.

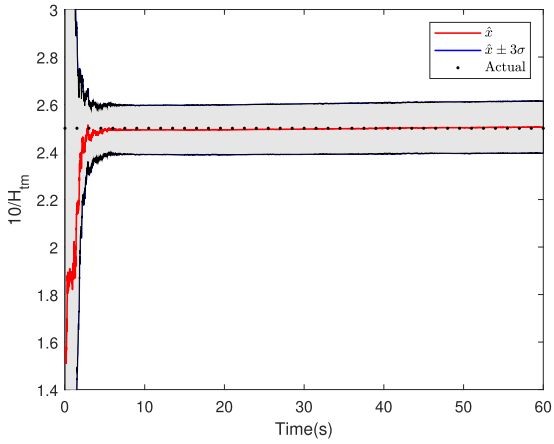


Fig. 4. Estimation result for the turbine-motor inertia constant.

10^{-2} simulating the real variability, [31]. This signal (Fig. 3), is used as an input for the CKF algorithm.

The estimation technique proposed in this work has proven to be accurate when small and random variations of the operating point are considered. Fig. 4 shows the results of the CKF estimation process throughout 60 s for the modified parameter

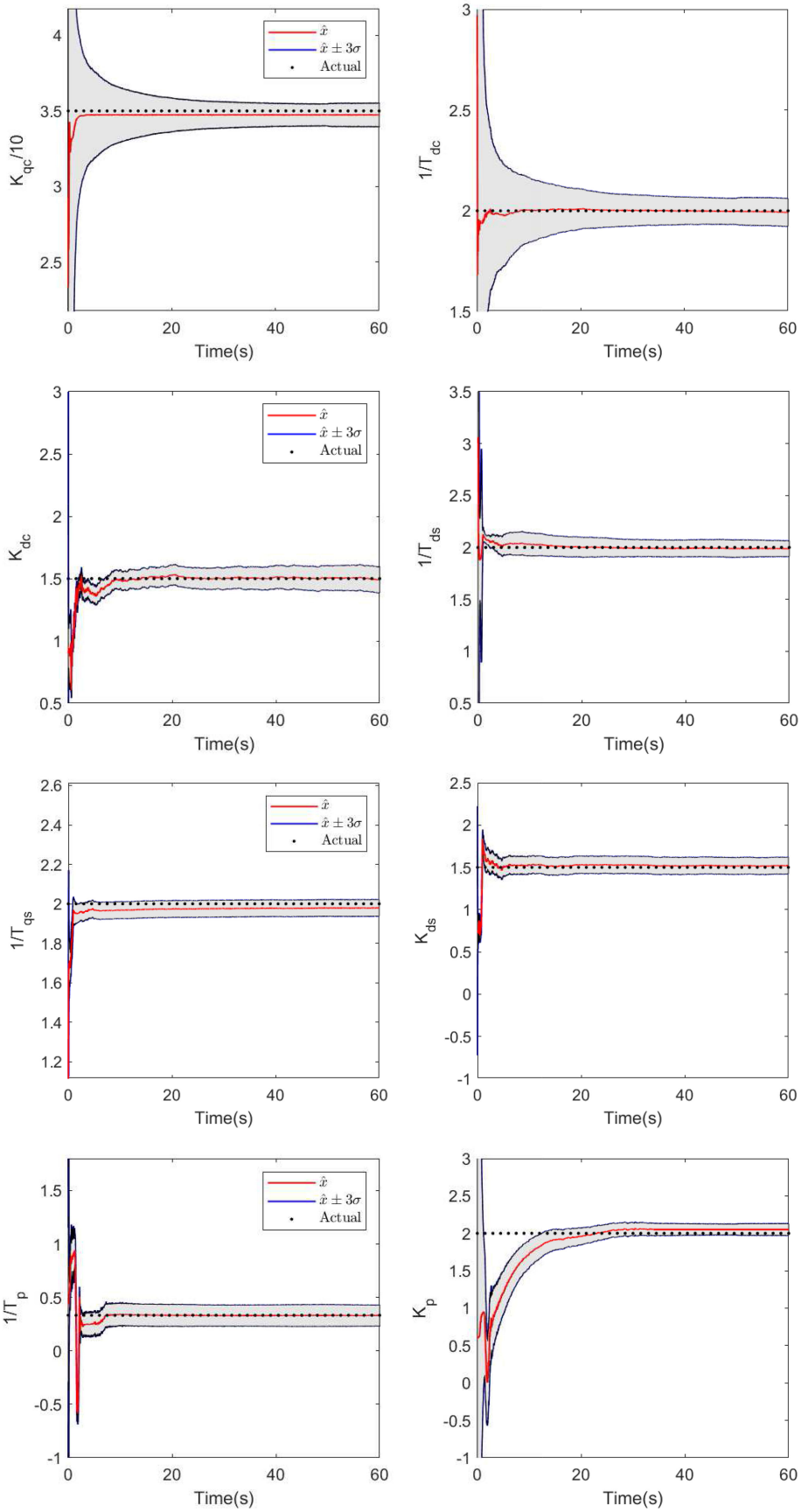


Fig. 5. Estimated parameters for VSC and pitch angle control with smooth variations.

TABLE V
RELATIVE ERROR IN PARAMETER ESTIMATION WITH SMOOTH VARIATIONS

Parameter	Estimated Value	Relative Error (%)
H_{tm}	4.012	0.300
K_{qc}	34.426	1.640
T_{dc}	0.503	0.600
K_{dc}	1.513	0.867
T_{ds}	0.504	0.800
T_{qs}	0.511	2.100
K_{ds}	1.510	0.667
T_p	2.991	0.300
K_p	2.023	1.150

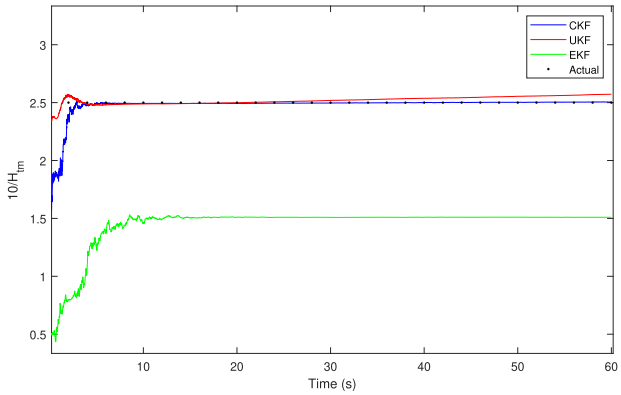


Fig. 6. Performance comparison in the estimation of shaft inertia by CKF, UKF and EKF.

representing the turbine-motor inertia constant H_{tm}, while the VSC and the pitch angle control parameters are represented in Fig. 5. For each parameter, the estimated value, $\hat{x}_i$, is represented along with the $\hat{x}_i \pm 3S_{ii}$ bounds. Note that the covariances tend to Q_{ii}, showing the accuracy of the converged estimation.

As can be seen, the method eventually converges to the neighborhood of the correct modified parameters, from which the original parameters can be recovered, as per equations (18)–(32). The estimated value of the original parameters and the relative estimation error are summarized in Table V. Note that the largest relative error is 2.1%.

The performance of the proposed estimation technique, based on CKF, has been compared with that shown by other KF filter schemes, such as UKF and EKF, as illustrated in Fig. 6, where the estimation of the shaft inertia under soft variations in the system is shown. Note that the EKF provides an unacceptable estimated value for the modified parameter and that the UKF presents convergence problems in the long term.

Once the accuracy of the parameter estimation process has been shown under mild operating conditions, the robustness of the proposed estimation technique is tested with three large disturbances separately arising in the system under study.

B. Disturbance 1: Wind Gust

This test considers an abrupt change in the wind speed, which is modeled as a *Mexican Hat Wavelet* [31], as shown in Fig. 7.

The evolution of the voltage at the point of connection is the same as in the base case.

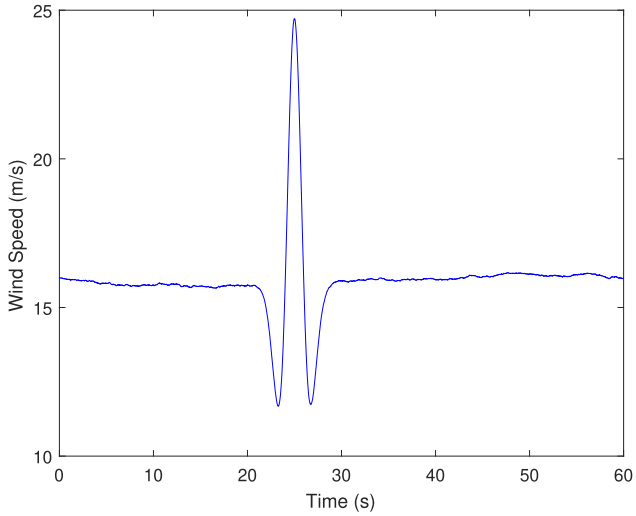


Fig. 7. Representation of the *Mexican Hat Wavelet*.

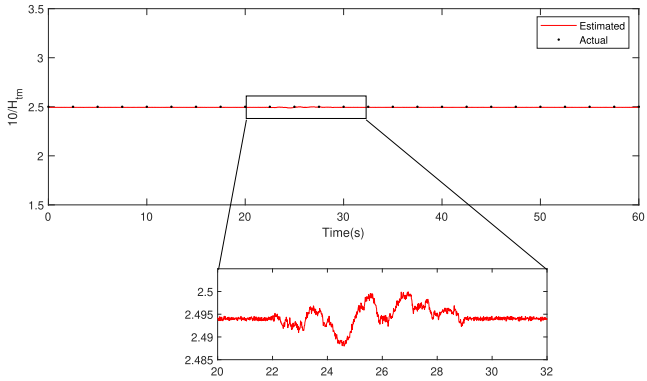


Fig. 8. Estimation result for $10/H_{tm}$ with wind gust.

In this case, the parameters are initialized with their estimated values, as provided by the base case. As the initial estimation is more accurate, the elements of $P_{\psi 0}$ are reduced to 10^{-4}. The value of Q_{ii} has also been modified to 10^{-6}.

The estimated parameters showed no significant variations during the wind gust, reaching the same values as in the base case after a small transient, once the disturbance vanishes. As an example, the evolution of $10/H_{tm}$ is provided in Fig. 8, where a zoomed view is added to better visualize the transient behavior.

C. Disturbance 2: Voltage Dip

The second disturbance considered is a 70% voltage dip at the point of connection that elapses for 1 s (Fig. 9). The tie-line impedance and the standard deviation of the Gaussian random walks are the same as in the previous case. The wind speed is assumed to remain as in the base case.

The initial estimation for the parameters and the elements of $P_{\psi 0}$ and Q remain the same as in the previous case (Disturbance 1).

Except for a small transient arising during the disturbance, the modified parameters showed no significant variations in the presence of the voltage dip, reaching the same estimated values as in the base case. As an example, the evolution of $1/T_{dc}$ in

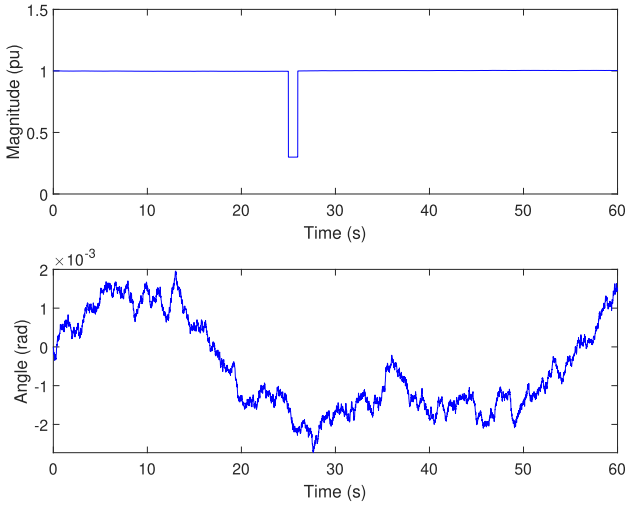


Fig. 9. Voltage dip at the point of connection.

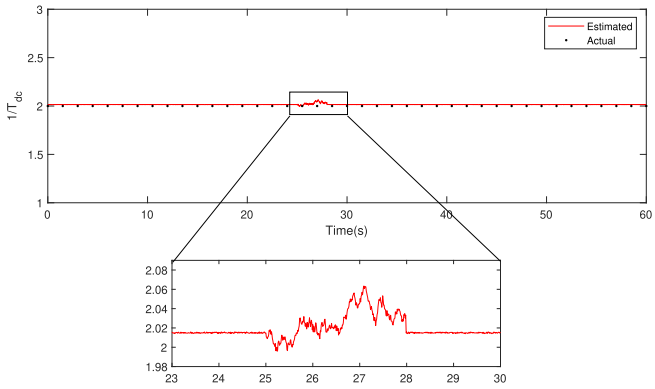


Fig. 10. Estimation result for $1/T_{dc}$ with voltage dip.

shown in Fig. 10, where the transient behavior can be noticed in the zoomed view.

D. Disturbance 3: Topological Change

In the last case study considered in this work, an abrupt variation in the tie-line impedance is assumed, representing a sudden topological change in the external system. For this purpose, at instant $t = 20\,\text{s}$, the value of z_L is modified to $z_L = 0.01 + j0.5$ pu permanently. The value of the voltage at the connection point and the wind speed remain the same as in the base case.

The initial estimation for the parameters and the elements of the matrices $P_{\psi 0}$ and Q remain the same as in the disturbances 1 and 2.

As the disturbance considered in this case is permanent, the estimated values are slightly different to those prior to the disturbance. See the evolution of $1/T_p$ in Fig. 11 as an example, where the scale of the transient period is augmented. Note that the accuracy of the estimated parameter in this case has improved minimally.

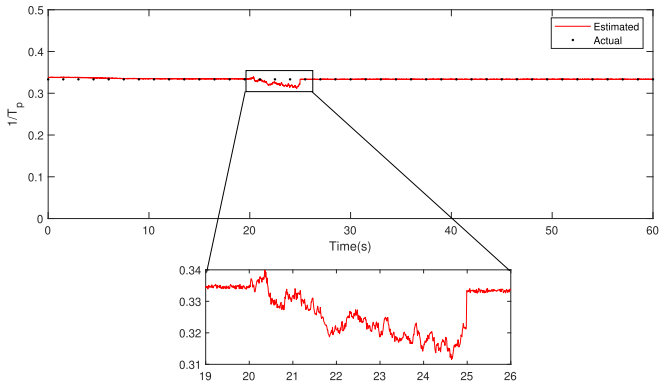


Fig. 11. Estimation result for K_p with topological change.

TABLE VI
MAXIMUM RELATIVE PARAMETER ERROR FOR INCREASING MEASUREMENT ERRORS

Measurement error	$\lvert E_r^{max}\rvert(\%)$
1%	2.100
3%	2.854
5%	5.926
7%	10.800

TABLE VII
MAXIMUM RELATIVE PARAMETER ERROR FOR INCREASING ERRORS IN L_d AND L_q VALUES

Model error	$\lvert E_r^{max}\rvert(\%)$
2%	2.107
5%	3.011
10%	6.667
15%	12.002

VI. MEASUREMENT AND MODEL ERROR IMPACT

Once the accuracy of the proposed estimation technique has been proved, the influence of the measurement error on the CKF performance is tested. For this purpose, the smooth system perturbations considered in the base case are repeated by gradually increasing the s.d. of the measurement errors. Table VI summarizes the results obtained in terms of maximum relative error in the parameter estimation. As expected, the lower the measurement accuracy the poorer the estimation results. Note that, if PMUs are used as the source of measurements, with errors typically lower than 1%, then the maximum parameter error will be less than 2%.

On the other hand, the synchronous generator model presented in Section III, as given by (25) and (26), includes the generator inductances in dq axis, L_d and L_q, which are considered to be known for the CKF implementation. The performance of the estimator with increasing error in those parameters is compared in Table VII, where the variation of the maximum relative error is shown.

Note that, while low model errors lead to sufficiently accurate estimation results, when the errors in the known system parameters are large enough, the CKF-based estimator performance deteriorates, in proportion to those errors.

VII. Conclusion

In this paper, a CKF is used to perform a joint estimation of the state variables and parameters of a variable-speed wind turbine with a direct-drive synchronous generator and a back-to-back VSC.

The proposed method includes a set of modified parameters, providing accurate estimation results (2.1% maximum relative error) under normal small variations in the operating point. A comparison of three KF schemes (CKF, UKF and EKF) has shown that the CKF is the most suitable for the parameter estimation of the system under study.

The robustness of the technique has been successfully proven with three different performance tests, simulating typical large disturbances that can occur in real life. Finally, the results provided by the CKF, for increasing values of measurement errors, show that the estimator performance is acceptable as long as the quality of the measurements remains within a range typically achieved by existing synchrophasors.

References

[1] Offshore Wind in Europe: Key trends and statistics 2017, European Wind Energy Association (EWEA). https://windeurope.org/wp-content/uploads/files/about-wind/statistics/WindEurope-Annual-Offshore-Statistics-2017.pdf

[2] NERC, "Reliability Guideline: Power Plant Dynamic Model Verification using PMUs," 2016. Downloadable from: https://www.nerc.com/comm/PC_Reliability_Guidelines_DL/Reliability%20Guideline%20-%20Power%20Plant%20Model%20Verification%20 using%20PMUs%20-%20Resp.pdf

[3] Z. Huang, K. Schneider, and J. Nieplocha, "Feasibility studies of applying Kalman Filter techniques to power system dynamic state estimation," in *Proc. Int. Power Eng. Conf.*, Dec. 2007, pp. 376–382.

[4] A. Kumar and B. C. Pal, "Dynamic estimation and control of power systems," ISBN: 978-0-12-814005-5. San Diego, CA, USA: Academic edition, Oct. 2018.

[5] J. Zhao and L. Mili, "A robust generalized-maximum likelihood unscented Kalman filter for power system dynamic state estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 4, pp. 578–592, Aug. 2018.

[6] A. Rouhani and A. Abur, "Constrained iterated unscented Kalman filter for dynamic state and parameter estimation," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2404–2414, May 2018.

[7] S. Wang, W. Gao, and A. P. Sakis Meliopoulos, "An alternative method for power system dynamic state estimation based on unscented transform," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 942–950, May 2012.

[8] M. A. González-Cagigal, J. A. Rosendo-Macías, and A. Gómez-Expósito, "Parameter estimation of fully regulated synchronous generators using unscented Kalman Filters," *Electric Power Syst. Res.*, vol. 168, 2019, pp. 210–217, Nov. 2018.

[9] C. Mengnan, Q. Yingning, F. Yanhui, W. Hao, and D. Infield, "Wind turbine fault diagnosis based on unscented Kalman Filter," in *Proc. Int. Conf. Renewable Power Gener.*, Oct. 2015, pp. 1–5.

[10] M. Abdelrahem, C. Hackl, and R. Kennel, "Application of extended Kalman filter to parameter estimation of doubly-fed induction generators in variable-speed wind turbine systems," in *Proc. Int. Conf. Clean Electric. Power*, Jun. 2015, pp. 226–233.

[11] S. P. Azad and J. E. Tate, "Parameter estimation of doubly fed induction generator driven by wind turbine," in *Proc. IEEE/PES Power Syst. Conf. Expo.*, Mar. 2011, pp. 1–8.

[12] Q. Jilong, T. Yantao, G. Yimin, and Zhucheng, "A sensorless initial rotor position estimation scheme and an extended Kalman Filter observer for the direct torque controlled permanent magnet synchronous motor drive," in *Proc. Int. Conf. Elect. Mach. Syst.*, Oct. 2008, pp. 3945–3950.

[13] X. Jiang, P. Sun, and Z. Q. Zhu, "Modeling and simulation of parameter identification for PMSM based on EKF," in *Proc. Int. Conf. Comput., Mechatronics, Control Electron. Eng.*, Aug. 2010, pp. 345–348.

[14] X. Xiao and C. Chen, "Reduction of torque ripple due to demagnetization in PMSM using current compensation," *IEEE Trans. Appl. Supercond.*, vol. 20, no. 3, pp. 1068–1071, Jun. 2010.

[15] L. Otava and L. Buchta, "Permanent magnet synchronous motor stator winding fault detection," in *Proc. IECON - 42nd Ann. Conf. IEEE Ind. Electron. Soc.*, Oct. 2016, pp. 1536–1541.

[16] L. Otava and L. Buchta, "Implementation and verification of the PMSM stator interturn short fault detection algorithm," in *Proc. 19th Eur. Conf. Power Electron. Appl.*, Sep. 2017, pp. P.1–P.10.

[17] G. Feng, C. Lai, J. Tjong, and N. Kar, "Non-invasive Kalman filter based permanent magnet temperature estimation for permanent magnet synchronous machines," *IEEE Trans. Power Electron.*, vol. 33, no. 12, pp. 10673–10682, Dec. 2018.

[18] A. Echchaachouai, S. Elhani, A. Hammouch, and S. Guedira, "Extended Kalman filter used to estimate speed rotation for sensorless MPPT of wind conversion chain based on a PMSG," in *Proc. 1st Int. Conf. Elect. Inf. Technologies Proc.*, Jul. 2015, pp. 172–177.

[19] Z. Zhang, F. Wang, G. Si, and R. Kennel, "Predictive encoderless control of back-to-back converter PMSG wind turbine systems with extended Kalman filter," in *Proc. IEEE 2nd Annu. Southern Power Electron. Conf.*, Dec. 2016, pp. 1–6.

[20] S. A. A. Shahriari, M. Raoofat, M. Mohammadi, M. Dehghani, and M. Saad, "Dynamic state estimation of a doubly fed induction generator based on a comprehensive nonlinear model," *Simul. Model. Pract. Theory*, vol. 69, p. 92–112, Dec. 2016.

[21] D. G Pillai, A. Vivek, and V. Srikanth, "Non-linear state estimation of PMSM using derivative-free and square-root Cubature Kalman Filter," in *Proc. Int. Conf. Intell. Comput., Instrum. Control Technologies*, Jul. 2017, pp. 126–131.

[22] I. Arasaratnam, "Cubature Kalman filtering: Theory and applications," Ph.D. Thesis, School of Graduate Studies of McMaster University, Hamilton, Ontario 2009.

[23] Y. Sun and Q. Chen, "Joint estimation of states and parameters of vehicle model using cubature Kalman filter," in *Proc. IEEE Int. Conf. Syst., Man, Cybernet.*, Budapest, Oct. 2016, pp. 000977–000982.

[24] G. R Gopinath and S. P. Das, "Speed and position sensorless control of interior permanent magnet synchronous motor using square-root cubature Kalman filter with joint parameter estimation," in *Proc. IEEE Int. Conf. Power Electron., Drives Energy Syst.*, Trivandrum, Dec. 2016, pp. 1–5.

[25] H. Pesonen and R. Piché, "Cubature-based Kalman filters for positioning," in *7th Workshop Positioning Navigation Commun.*, Mar. 2010, vol. 45, no. 49, pp. 11–12.

[26] F. Milano, "Power system modelling and scripting," ISBN: 978-3-642-13668-9, Springer Publishing Company, Incorporated, 2010.

[27] H. Ghassempour, Z. Miao, L. Fan, W. Jiang, and D. Manjure, "Identification of synchronous generator model with frequency control using unscented Kalman filter," in *Electric Power Syst. Res.*, vol. 126, pp. 45–55, Sep. 2015.

[28] R. Gagnon and J. Brochu, "Wind Farm - Synchronous Generator and Full Scale Converter (Type 4) Detailed Model," Mathworks. https://es.mathworks.com/help/physmod/sps/examples/wind-farm-synchronous-generator-and-full-scale-converter-type-4-detailed-model.html. Last visited: July 2019.

[29] S. Julier, "The scaled unscented transformation," in *American Control Conference, 2002. Proceedings of the 2002*, Piscataway, NJ, USA: IEEE Press, 2002, pp. 4555–4559.

[30] J. G. Slootweg, H. Polinder, and W. L. Kling, "Initialization of wind turbine models in power system dynamics simulations," in *IEEE Porto Power Tech Proc. (Cat. No.01EX502)*, vol. 4, pp. 6, Aug. 2001.

[31] IEC 61400-1, Wind Turbines - Part 1: Design Requirements, International Electrotechnical Commission, Geneva, Switzerland, Tech. Rep., August 2005.

Annex II.3

M.A. González-Cagigal, J.A. Rosendo-Macías, A. Gómez-Expósito, "Application of nonlinear Kalman filters to the identification of customer phase connection in distribution grids," in: International Journal of Electrical Power and Energy Systems, Volume 125, 2021.

Electrical Power and Energy Systems 125 (2021) 106410

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Application of nonlinear Kalman filters to the identification of customer phase connection in distribution grids

M.A. González-Cagigal, J.A. Rosendo-Macías, A. Gómez-Expósito

Department of Electrical Engineering, University of Seville, Spain

ARTICLE INFO

Keywords:
Ensemble Kalman filter
Cubature Kalman Filter
Unscented Kalman Filter
Phase identification
State estimation

ABSTRACT

This paper presents a state estimation approach to address the problem of identifying the phase to which single-phase customers are connected in three-phase distribution grids. The proposed method performs Kalman filtering on the information provided simultaneously by the smart meter of every customer and the aggregated energy consumption measured at each phase of the secondary substation feeding the set of customers. Different nonlinear formulations of the Kalman filter are tested and their performance compared, showing that the ensemble Kalman filter provides better estimation results when the system size increases. The accuracy, robustness and limitations of the estimator are also tested when measurement errors are considered.

1. Introduction

Correct operation monitoring and control in distribution systems are essential in order to assure a good quality of the service provided to the customers. In this context, it is most important for grid operators to unambiguously know which loads are connected to each of the three phases of the system (the European feeder topology is assumed in this work). An accurate connectivity information is a prerequisite to promote the correct phase balance of LV consumers, alleviating in this way the problems derived from feeder unbalances, such as sharper voltage drops, which can even violate the grid codes, and increasing power losses which also affect the lifespan of the equipment due to temperature rise. Moreover, the penetration of renewable energy resources at the distribution level also benefits from the phase identification, since it helps to establish a better production-consumption balance for each phase of the grid.

In this regard, despite the efforts undertaken by distribution companies, they frequently lack enough information about the phase connection of their single-phase customers, owing for instance to network reconfiguration after faults, phase switching derived from improper maintenance, or inaccurate recording of the true load-to-phase connectivity. In these circumstances, a method must be developed to estimate as accurately as possible the actual phase to which a customer is connected in LV feeders, which is known as the customer-phase identification (CPI) problem.

The CPI problem has been approached in several ways by previous works. In [1], a signal processing perspective is applied to voltage observations, which are also used both in [2], for a correlation-based methodology, and in [3], where a spectral clustering technique is proposed. The connecting phase of underground distribution transformers is determined in [4] through phase voltage measurements. Smart meters have improved the communication between the loads and the substations and can be also used for the CPI problem [5,6]. A method based on Least Absolute Shrinkage and Selection Operator (LASSO) is proposed in [7], also using smart meter data from a LV distribution network. In [8], a novel approach for phase identification using graph theory and principal component analysis (PCA) is tested. The possible missing information in smart meter data is dealt with in [9] through a correlation analysis.

In this work, the CPI is addressed by applying a Kalman filtering (KF) state estimation technique to a set of hourly consumption curves obtained from real loads, along with hourly energy measurements taken at each phase of the secondary substation. The proposed method conservatively assumes that other electrical magnitudes potentially provided by smart meters, such as voltage readings or reactive power consumed by each load, are not available. Moreover, a simplified loss model is adopted, allowing the impact of each load on the energy delivered by each phase of the transformer to be estimated for a given topology.

The KF is a dynamic state estimator (DSE) widely used in electric power systems in any of its diverse forms [10–12]. Particularly, the application of KF to parameter estimation in power systems has been successfully tested, e.g. in [13–15], providing evidence of the DSE potential for the CPI problem.

In this work, three nonlinear KF schemes are tested and compared, namely the so-called unscented Kalman filter (UKF), the cubature Kalman filter (CKF), and the ensemble Kalman filter (EnKF). The proposed technique takes advantage of the hourly information provided by smart meters to sequentially assign customers to the most likely phase. The proposed KF implementation does not explicitly enforce, beforehand, binary constraints for the state variables, but rather adopts a novel, statistically-based inference logic successively rounding state

https://doi.org/10.1016/j.ijepes.2020.106410
Received 26 December 2019; Received in revised form 6 July 2020; Accepted 30 July 2020

variables to their nearest binary value. As the input information considered in this work is similar to that assumed in [7,8], the KF-based estimation is also compared to those competing techniques for different scenarios.

The paper is organized as follows: Section 2 provides a brief background on the different KF schemes considered. Next, the feeder used for testing and the simplified loss model adopted are presented in Section 3. The implementation details of the KF based estimation techniques, as applied to the CPI problem, are given in Section 4. In Section 5, the results obtained in two different scenarios are presented and discussed, while the proposed KF technique is compared in Section 6 with other published works dealing with CPI. The conclusions derived from the case studies considered are drawn in Section 7.

2. Kalman filter background

In this section, three different Kalman filter formulations are succinctly reviewed, as applied to continuous-time, discrete-measurement nonlinear systems. In the discrete-time framework, the associated equations may be expressed as follows,

$$x_k = f(x_{k-1}, u_{k-1}) + w_k \tag{1}$$

$$z_k = g(x_k, u_k) + v_k \tag{2}$$

where x_k is the state vector at instant k, u_k the system input, and z_k the vector of available measurements. Gaussian processes are considered for the model and measurement noises, w_k and v_k, with covariance matrices Q and R, respectively.

The iterative processes of the Kalman filter schemes considered in this work, all of them involving prediction and correction stages, are described below.

2.1. Unscented Kalman Filter

At instant k, a cloud of $2L + 1$ vectors, the so-called σ-points, is obtained from the previously estimated state vector, $\hat{x}_{k-1}$ (dimension L), and the covariance matrix of the state estimation error, P_{k-1}, as follows [16]:

$$\begin{cases} x_{k-1}^0 = \hat{x}_{k-1} \\ x_{k-1}^i = \hat{x}_{k-1} + [\sqrt{(L+\lambda)P_{k-1}}]_i \\ x_{k-1}^{i+L} = \hat{x}_{k-1} - [\sqrt{(L+\lambda)P_{k-1}}]_{i+L} \\ i = 1,...,L \end{cases} \tag{3}$$

$[\sqrt{(L+\lambda)P_{k-1}}]_i$ being the i^{th} column of the corresponding matrix, and λ a scaling factor calculated as follows:

$$\lambda = \alpha^2(L + \kappa) - L \tag{4}$$

where α and κ are two filter parameters to be tuned.

These σ-points are evaluated through Eq. (1), yielding $2L + 1$ vectors, x_k^{i-}, from which the *a priori* estimations $\hat{x}_k^-$ and P_k^- are obtained:

$$\hat{x}_k^- = \sum_{i=0}^{2L} W_{mi} x_k^{i-} \tag{5}$$

$$P_k^- = \sum_{i=0}^{2L} W_{ci}\left(x_k^{i-} - \hat{x}_k^-\right)(x_k^{i-} - \hat{x}_k^-)^T + Q_k \tag{6}$$

where the weighting vectors W_m and W_c are calculated from:

$$\begin{cases} W_{m0} = \frac{\lambda}{L+\lambda} \\ W_{c0} = \frac{\lambda}{L+\lambda} + 1 - \alpha^2 + \beta \\ W_{mi} = W_{ci} = \frac{1}{2(L+\lambda)} i = 1,...,2L \end{cases} \tag{7}$$

β being another tunable parameter.

On the basis of the *a priori* information, the correction stage starts with the calculation of a new cloud of vectors, x_k^-, which are evaluated with the measurement function $g(\cdot)$ in Eq. (2), and weighted with the vectors W_m, yielding

$$\gamma_k^{i-} = g(x_k^{i-}, u_k) i = 0, \ ...,2L \tag{8}$$

$$\hat{z}_k^- = \sum_{i=0}^{2L} W_{mi} \gamma_k^{i-} \tag{9}$$

Then, the covariance matrix of the measurement estimation error, P_{zk}^-, and the cross-covariance matrix of state and measurements, P_{xzk}^-, are obtained using the vector W_c as follows:

$$P_{zk}^- = \sum_{i=0}^{2L} W_{ci}\left(\gamma_k^{i-} - \hat{z}_k^-\right)(\gamma_k^{i-} - \hat{z}_k^-)^T + R_k \tag{10}$$

$$P_{xzk}^- = \sum_{i=0}^{2L} W_{ci}\left(x_k^{i-} - \hat{x}_k^-\right)(\gamma_k^{i-} - \hat{z}_k^-)^T \tag{11}$$

The correction stage concludes with the *a posteriori* predictions,

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - \hat{z}_k^-)^T \tag{12}$$

$$P_k = P_k^- - K_k P_{zk}^- K_k^T \tag{13}$$

which are based on the *a priori* predictions at instant k and the so-called Kalman gain, K_k, calculated from

$$K_k = P_{xzk}^-(P_{zk}^-)^{-1} \tag{14}$$

2.2. Cubature Kalman Filter

This KF formulation uses a set of $2L$ cubature points calculated from $\hat{x}_{k-1}$ and P_{k-1} through the following expressions [17]:

$$S_{k-1}S_{k-1}^T = P_{k-1} \tag{15}$$

$$x_{k-1}^i = S_{k-1}\xi_i\sqrt{L} + \hat{x}_{k-1} i = 1, \ ...,2L \tag{16}$$

where S is a positive-definite square root of matrix P (the Cholesky factorization of matrix P is customarily used), and ξ_i is the i^{th} cubature node, obtained as the intersection of the unit sphere and the $\mathbb{R}^L$ axis.

The state function $f(\cdot)$ in (1) is evaluated for the set of cubature points, yielding a set of $2L$ vectors x_k^{i-}, from which the *a priori* estimation is computed,

$$\hat{x}_k^- = \frac{1}{2L}\sum_{i=1}^{2L} x_k^{i-} \tag{17}$$

$$P_k^- = \frac{1}{2L}\sum_{i=1}^{2L} x_k^{i-}x_k^{i-T} - \hat{x}_k^-\hat{x}_k^{-T} + Q_k \tag{18}$$

For the correction stage, the covariance matrix P_k^- is factorized in order to calculate both the matrix S_k^-,

$$S_k^- S_k^{-T} = P_k^- \tag{19}$$

and a new set of $2L$ cubature points, x_k^-, at which function $g(\cdot)$ in (2) is evaluated to obtain γ_k^-.

Then the measurement estimation, $\hat{z}_k^-$, its covariance matrix, P_{zk}^-, and the cross-covariance matrix of state and measurements, P_{xzk}^-, are calculated as follows:

$$\hat{z}_k^- = \frac{1}{2L}\sum_{i=1}^{2L} \gamma_k^{i-} \tag{20}$$

$$P_{zk}^- = \frac{1}{2L}\sum_{i=1}^{2L} \gamma_k^{i-}\gamma_k^{i-T} - \hat{z}_k^-\hat{z}_k^{-T} + R_k \tag{21}$$

$$P_{xz,k}^{-} = \frac{1}{2L}\sum_{i=1}^{2L} x_k^{i-}\gamma_k^{i-T} - \hat{x}_k^{-}\hat{z}_k^{-T} \tag{22}$$

The *a posteriori* predictions of the state vector, $\hat{x}_k$, and the covariance P_k are calculated with the Kalman gain using the same Eqs. (14) and (13) as in the UKF algorithm.

2.3. Ensemble Kalman Filter

The EnKF [18], is a Monte Carlo approximation of the original KF which has proven accurate enough in high-dimensional state-space problems. The ensemble is represented by an *LxN* matrix, *N* being the number of samples considered.

The ensemble is first propagated through the state and measurement functions,

$$\begin{cases} x_k^{i-} = f(x_{k-1}^{i}, u_{k-1}) \\ z_k^{i-} = g(x_k^{i}, u_k) \\ i = 1,...,N \end{cases} \tag{23}$$

and then the mean values are calculated:

$$\overline{x_k^-} = \frac{1}{N}\sum_{i=1}^{N} x_k^{i-} \tag{24}$$

$$\overline{z_k} = \frac{1}{N}\sum_{i=1}^{N} z_k^{i-} \tag{25}$$

The EnKF correction stage is based on the calculation of the intermediate matrices

$$\overline{P_k}H_k^T = \frac{1}{N}\sum_{i=1}^{N}\left(x_k^{i-} - \overline{x_k^-}\right)(z_k^{i-} - \overline{z_k})^T \tag{26}$$

$$H_k\overline{P_k}H_k^T = \frac{1}{N}\sum_{i=1}^{N}\left(z_k^{i-} - \overline{z_k}\right)(z_k^{i-} - \overline{z_k})^T \tag{27}$$

which allow the Kalman gain and the updated values of each sample in the ensemble to be obtained:

$$K_k = \overline{P_k}H_k^T(H_k\overline{P_k}H_k^T + R)^{-1} \tag{28}$$

$$x_k^i = x_k^{i-} + K_k(z_k - z_k^{i-})i = 1,\ ...,N \tag{29}$$

Finally, the corrected covariance matrix, P_k, is calculated as follows:

$$\overline{x_k} = \frac{1}{N}\sum_{i=1}^{N} x_k^i \tag{30}$$

$$P_k = \frac{1}{N}\sum_{i=1}^{N}\left(x_k^i - \overline{x_k}\right)(x_k^i - \overline{x_k})^T \tag{31}$$

3. Problem statement and modeling

In order to assess the ability of the KF-based methodology to address the CPI problem, a typical distribution grid is considered comprising N_s single-phase and N_t three-phase customers, not necessarily balanced, resulting in $N_c = N_s + 3N_t$ total consumption curves in the network. In this work, it is assumed that the number of three-phase customers is 20% of the total. The proposed methodology, as presented in Section 4, is easily extensible to other load connection arrangements, such as two-phase loads, still found in some areas. Fig. 1 shows an example of a distribution network with 100 loads, used in the sequel to test the estimation techniques.

The energy consumption of each load *i* at a certain hour *k*, denoted by $E_{i,k}$, is obtained from [19], where real hourly data from a European distribution company, comprising smart meters readings for 20 days, are provided, leading to a total of 480 energy measurements for each customer. In case of three-phase loads, a single-phase consumption is assigned to each phase. As customers with null consumption cannot be identified (they provide no information), the corresponding curves are removed from the raw data.

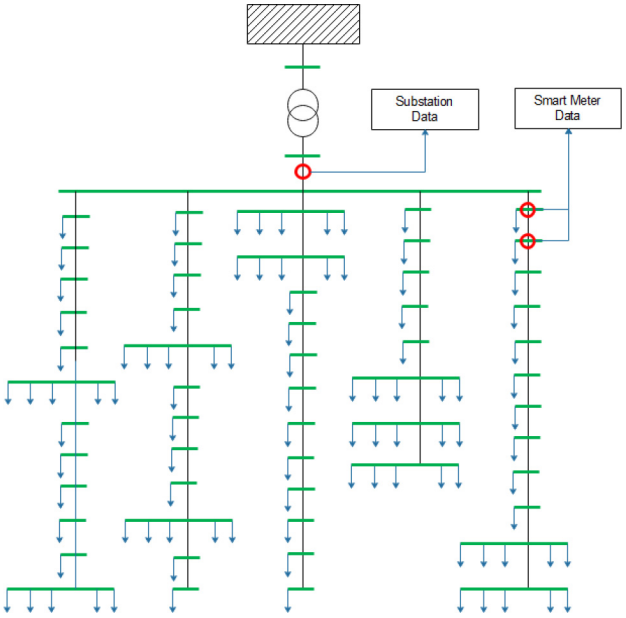


Fig. 1. Single-line diagram of one of the test networks.

To fully characterize the energy consumption of the customers, the reactive energy is also obtained from the raw data in [19]. Finally, the resulting hourly curves are randomly associated to a certain phase (a, b or c).

Along with the previous customer information, a typical distribution feeder topology is considered, each customer being associated to one of the grid nodes. Then, a load flow can be solved at each hour *k* in order to obtain the energy delivered by each phase of the MV/LV secondary substation, $E_{SS,k}^a$, $E_{SS,k}^b$ and $E_{SS,k}^c$, which is used by the proposed KF-based estimation technique as a measurement. This fully defines the distribution grid model involved in the estimation process.

4. Kalman Filter Implementation

The application of the different KF schemes described in Section 2 to the CPI problem is illustrated in the flowchart represented in Fig. 2, as summarized in the sequel.

4.1. Parallel filtering

For every iteration of each tested technique, three independent KF-based estimators run in parallel, one for each phase *p* of the distribution grid, being the state vector composed of N_c variables, x_i^p, with $p = a, b, c$, so that $L = N_c$. Any three-phase customer is characterized by a set of three consecutive state variables, $\{x_i^p, x_{i+1}^p, x_{i+2}^p\}$, for each phase. Those variables indicate if the consumption *i* is associated to the corresponding phase *p*. For these static variables, the state function, $f(\cdot)$, in Eq. (1), is taken as a random walk, independent of the system input, u_k, yielding the following expression:

$$x_k^p = x_{k-1}^p + w_k^p \tag{32}$$

where the model noise, w_k^p, is considered to have the same covariance matrix, *Q*, for each phase.

The system input, u_k, used in the measurement function $g(\cdot)$ in Eq. (2), is determined by the hourly energy consumption of each customer, $u_k^T = E_1, E_2, ...,E_{N_c,k}$. Regarding the network losses, as no measurements about the voltage and power factor of each load are available, a

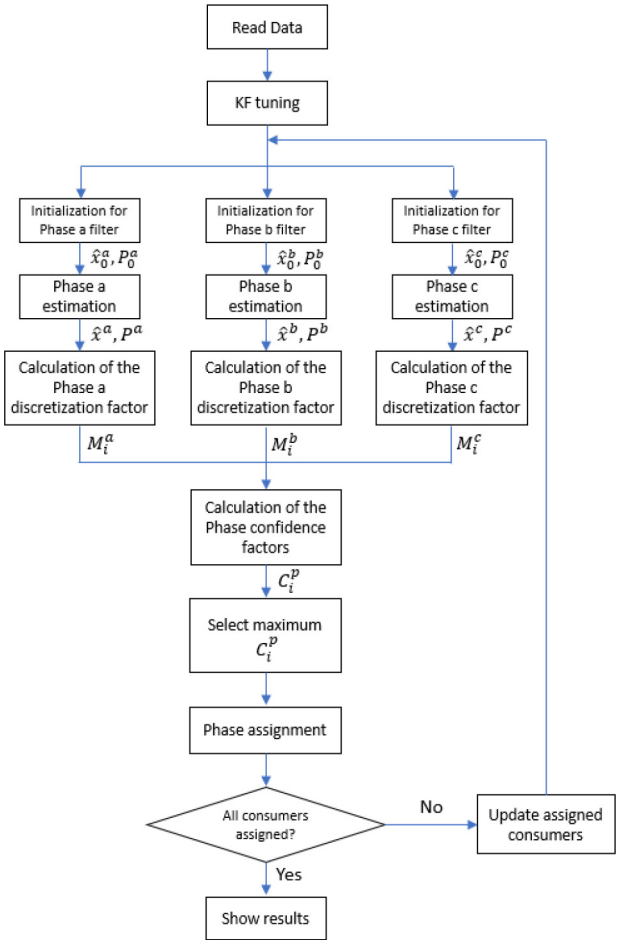


Fig. 2. Flowchart of the CPI methodology.

simplified loss model is adopted by the KF estimators, for which a given rated voltage and power factor $\cos\varphi_{i,k} = 1$ are assumed for all loads points. With these assumptions, an average hourly current, $I_{i,k}$ is calculated for each consumer i from the corresponding hourly energy consumption, $E_{i,k}$:

$$I_{i,k} = \frac{E_{i,k}}{T \cdot U} \tag{33}$$

where U is the voltage magnitude, and T is the energy integration period. In this work, as the energy measurements are supposed to be obtained hourly, $T = 1h$.

Accordingly, the energy loss attributable to customer i is given by:

$$E_{i,k}^{loss} = I_{i,k}^2 \cdot T \cdot r \cdot l_i \tag{34}$$

where r is the conductor resistance per unit length and l_i is the electrical distance from each consumer to the secondary substation, both assumed to be known ($r = 0.223\ \Omega/\text{km}$, has been adopted in this paper).

The vector u_k also includes the variables x_i^p corresponding to customers which have been previously assigned to a phase, as described below.

The measurement of each filter, z_k^p, is taken as the energy delivered at each hour by the corresponding phase of the secondary substation, $z_k^p = E_{SS,k}^p$, so the measurement function $g(\cdot)$ in Eq. (2) reduces to:

$$z_k^p = \sum_{i=1}^{N_c} x_{i,k}^p \cdot \left(E_{i,k} + E_{i,k}^{loss} \right) + v_k^p \tag{35}$$

where the measurement noise, v_k^p, is considered to have the same covariance matrix, R, for each phase.

4.2. Initialization and tuning

As explained formerly, before the KFs are applied, a set of input data are gathered, including the hourly consumption of each customer and the energy delivered by the MV/LV transformer for the time window available. A separation is made between single-phase and three-phase loads. For the last ones, it is assumed that three energy measurements are available, one per phase, but the phase labels (a, b or c) are unknown. At the end of the estimation processes, the actual distribution of customers among the three phases is used to evaluate the performance of each estimation procedure.

The KF formulations require that an initial estimation is adopted, which is determined by the state vector $\hat{x}_0$ and the covariance matrix of the initial estimation error, P_0. For the UKF and CKF implementations, all state variables in the three phases are initialized as 0, assuming a complete lack of knowledge about their real values. In the case of EnKF, the samples are given random binary values as initial estimation, which enhances the convergence of the estimation algorithm. The covariance matrix, P_0 is considered as a diagonal matrix with $P_{ii} = 10$ in the three KF schemes.

The covariance matrices Q and R are defined as diagonal matrices, considering typical values of $Q_{ii} = 10^{-4}$ and $R_{ii} = 9 \cdot 10^{-4}$ respectively, the last one being equivalent to assuming a s.d. of 3% for the measurement errors.

For the estimation based on UKF, it is necessary to define the tunable parameters of the filters introduced in Eqs. (4) and (7). A study is made in [20] over the influence of these parameters in the estimation process, concluding that $\alpha = 10^{-4}$, $\beta = 2$ and $\kappa = 3 - N_c$, are reasonable values for good estimation results.

Regarding the EnKF, the number of samples in the ensemble is taken as $N = 10 \cdot N_c$ so that it can suit the different sizes of the state vector.

4.3. Candidate selection and assignment

Once the three parallel estimators have provided a customer distribution for their corresponding phase, a single consumption is assigned and removed from the subsequent estimation processes. The candidate selection at each iteration of the CPI procedure is based on the estimated values of the state variables, $\hat{x}_i^p$, and the covariance of their estimation error, P_{ii}.

Forcing the state variables to be only 0 or 1, as required by the nature of the problem in hand, would involve equality constraints of the form $\hat{x} \cdot (\hat{x} - 1) = 0$. However, the application of the KF estimation algorithms to such non-convex model is prone to convergence problems. For this reason, the approach proposed in this work to enforce the binary character of the state variables relies on the confidence level of a given consumption being associated to a certain phase and not to the others. This is quantified through the so-called phase confidence factors, C_i^p, calculated with the probability density function (PDF) of the state variables, as follows.

First, the cumulative density function for $\hat{x}_i^p > 0.5$, denoted as M_i^p, is calculated as follows:

$$\begin{cases} M_i^p = \frac{1}{\sqrt{2 \cdot \pi \cdot P_{ii}}} \int_{0.5}^{\infty} e^{-\frac{(x - \hat{x}_i^p)^2}{2 \cdot P_{ii}}} dx \\ p = a, b, c \\ i = 1, \ldots, N_c \end{cases}$$

Since M_i^p provides information on the discrete value associated to the corresponding state variable, this coefficient is called phase discretization factor. Fig. 3 includes a graphic representation of this factor, corresponding with the shaded area in the Gaussian density function, considering 0.5 as lower limit.

Then, for each energy consumption curve i, the phase confidence factor combine the information derived from the three phases of the corresponding load through the following expression:

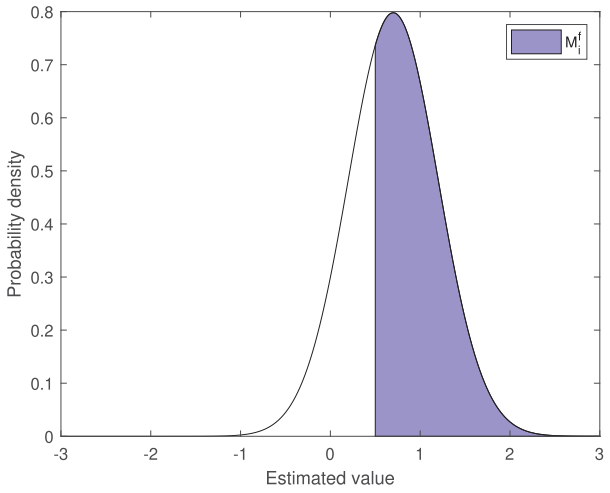


Fig. 3. Representation of M_i^p.

$$\begin{cases} C_i^a = M_i^a \cdot (1 - M_i^b) \cdot (1 - M_i^c) \\ C_i^b = M_i^b \cdot (1 - M_i^c) \cdot (1 - M_i^a) \\ C_i^c = M_i^c \cdot (1 - M_i^a) \cdot (1 - M_i^b) \\ i = 1, \ldots, N_c \end{cases}$$

The maximum value of the phase confidence factors C_i^p is selected, which determines the consumption i with the highest probability of being associated to the phase p and not associated to the other phases.

Finally, once a consumption is selected using the previously defined phase confidence factors, the three variables associated with that customer are assigned integer values as follows:

- If the consumption is related to a single-phase client, $\hat{x}_i^p = 1$ for the phase the client is connected to, and $\hat{x}_i^p = 0$ for the other two phases.
- For three-phase clients, assume without loss of generality that the consumption i is related to the phase a. Then, $\hat{x}_i^a = 1$, $\hat{x}_i^b = 0$ and $\hat{x}_i^c = 0$, as in the case of single-phase customers. Additionally, as two energy consumptions from a certain three-phase load cannot be associated to the same phase, the variables $\hat{x}_{i+1}^a$ and $\hat{x}_{i+2}^a$ are set to 0.

4.4. Update results

If all the loads have been assigned, then the estimation process ends. Otherwise, the state variables and the system input are updated.

For each filter, the state variables of the assigned consumptions are extracted from the state vector x and introduced as part of the system input, so that the number of unassigned consumers N_c is reduced accordingly. The estimation process is then repeated with the remaining consumers until $N_c = 0$, meaning that the final estimation result has been achieved.

5. Case studies

In this section, the proposed KF formulations are compared on several case studies, which can be grouped into four different scenarios.

By performing a series of preliminary tests in which the percentages of three-phase loads range from 10 to 30%, no significant differences have been observed in the performance of the proposed methodology. For this reason, as stated above, 20% of three-phase customers has been assumed in all scenarios.

5.1. Scenario I: Original measurements

In this scenario, the actual consumption is used for the KF-based estimation, assuming that all the measurements are correctly obtained.

Table 1
Estimation results for the UKF. Scenario I.

Consumption curves	Correct assignments	Wrong assignments	Percentage
50	50	0	100%
100	95	5	95%
200	181	19	90.5%
300	216	84	72%
400	233	167	58.25%

Table 2
Estimation results for the CKF. Scenario I.

Consumption curves	Correct assignments	Wrong assignments	Percentage
50	50	0	100%
100	100	0	100%
200	195	5	97.5%
300	262	38	87.33%
400	313	87	78.25%

The performance of UKF and CKF with increasing number of loads is shown in Tables 1 and 2, respectively. Note that, while both estimators correctly assign 100% of customers for $N_c = 50$, the behavior of the UKF deteriorates faster as the number of loads increases.

As shown in Table 3, the EnKF response to increasing system sizes is better than that of the other formulations, which confirms the expected behavior of this KF scheme for high-dimensional problems.

The convergence of the estimation processes is not only determined by the value of $\hat{x}_i^p$, but also by the covariance of the estimation error for each state variable, P_{ii}. An illustrative example is shown in Fig. 4, representing the PDFs obtained for the three state variables of a certain single-phase consumer, given the values of $\hat{x}_i^p$ and P_{ii} from the KF-based estimation process. The phase discretization factors, M_i^p, corresponding to the shaded areas in the graphics, and their numeric values are included in the respective legends. In this particular case the computed values of the phase confidence factors are $C_i^a = 0.014$, $C_i^b = 0.711$ and $C_i^c = 0.001$, meaning a confidence of 0.711 that the consumer is associated to phase b. This value would be compared with those of the rest of the clients in order to obtain the selected candidate in the corresponding iteration of the proposed methodology.

Finally, in order to illustrate the evolution of the phase confidence factor, C_i^p, Fig. 5 represents this coefficient in descending order at different stages of the estimation process for 100 loads (i.e., 300 C_i^p factors), using the EnKF formulation. It can be noticed that the maximum of the confidence factors, used to select the next candidate at the corresponding iteration, is close to 1 in all cases.

In the top graph of Fig. 5, when the whole set of 100 loads is still unassigned, a large number of phase confidence factors remain with a small value ($C_i^p = 0.125$), in accordance to the common initial value $M_i^p = 0.5$ adopted for the three-phase discretization factors, which means a complete lack of information, at this early stage of the iterative process, on the phase to which those consumption curves should be associated. Then, as more loads are assigned, the coefficients of the remaining loads more clearly show a trend towards 1 or 0 (about one third tends to 1 whereas the remaining two thirds tend to 0).

Table 3
Estimation results for the EnKF. Scenario I.

Consumption curves	Correct assignments	Wrong assignments	Percentage
50	50	0	100%
100	100	0	100%
200	200	0	100%
300	283	17	94.33%
400	346	54	86.5%

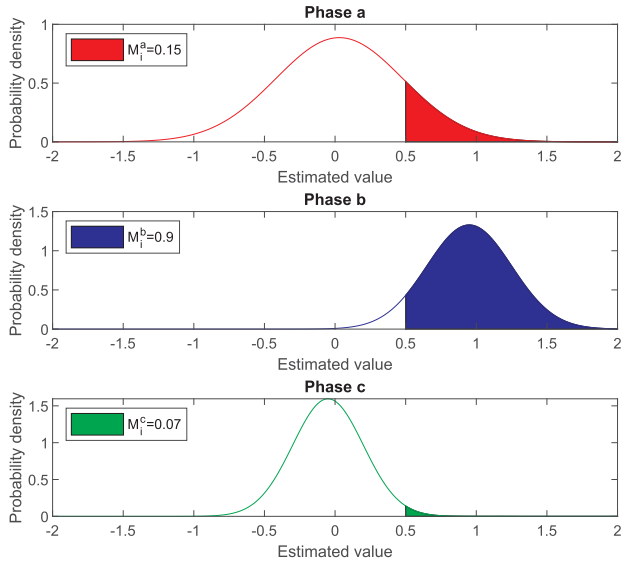


Fig. 4. Probability density functions for a sample consumer.

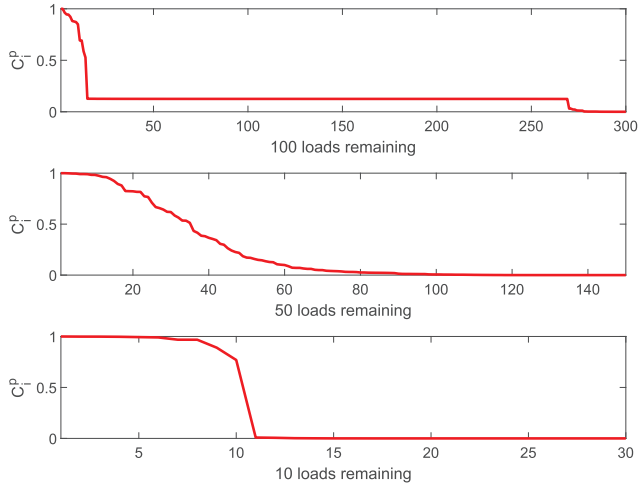


Fig. 5. Evolution of the phase confidence factor at different stages.

5.2. Scenario II: Analysis of the required amount of data

For the results shown above, hourly smart meter readings for 20 days are used, leading to 480 total energy measurements for each consumer. This section analyzes the quality of the estimation as the number of available measurements decreases. Table 4 summarizes the performance of the three estimators for the case in which $N_c = 200$ consumption curves ($N_s = 116$ single-phase and $N_t = 28$ three-phase customers).

Those results suggest a deterioration of the performance of the KF-based estimation techniques when the number of available measurements is lower than 200 (around 8 days) in this particular scenario. Further tests with feeders comprising different numbers of customers show that the required number of measurement snapshots increases with the number of loads, as intuitively expected. For instance, for $N_c = 100$ (one half of that shown in Table 4), the success rate of the three estimators does not deteriorate substantially, even when only 100 energy measurements (around 4 days) are considered for each load.

Table 4
Estimation results for different KF formulations. Scenario II.

Available data	UKF Hit Rate (%)	CKF Hit Rate (%)	EnKF Hit Rate (%)
480	90.5	97.5	100
400	90.5	97.5	100
300	90	96	100
200	85	90	95
100	67	70	72

Table 5
Estimation results for different KF formulations. Scenario III.

Noise level	UKF Hit Rate (%)	CKF Hit Rate (%)	EnKF Hit Rate (%)
1%	95	100	100
2%	92	96	100
3%	89	93	98
5%	83	89	92

5.3. Scenario III: Noisy measurements

The performance of the different KF formulations is evaluated in a realistic scenario where errors in the measurements are considered.. As the objective of these case studies is to determine the robustness of the KF schemes against measurement errors, a relatively low number of loads is considered, namely $N_s = 58$ single-phase and $N_t = 14$ three-phase clients, which leads to a total of $N_c = 100$ consumption curves.

Gaussian noise is artificially added to each measurement after the load flow is computed. The rates of correct phase-to-customer assignments for increasing measurement errors are summarized in Table 5 for the three estimators.

In light of those results, it can be concluded, also as expected, that the number of correct assignments decreases with increasing measurement noise, for every KF formulation, being the robustness of the CKF and EnKF similar, superior in any case to that of the UKF formulation. Nevertheless, all formulations show acceptable results when typical noise levels are considered in the measurements.

5.4. Scenario IV: Model errors

In the proposed implementation of the KF for the CPI problem, a simplified loss model is considered for which the value of the conductor resistance per unit length, r, is required as per Eq. (34). In this scenario, the performance of the different KF formulations is evaluated when errors in r are considered.

For the same number of customers as in Scenario III, Table 6 summarizes the rates of correct phase-to-customer assignments for errors in r ranging from 5 to 20%.

It can be concluded that the results remain acceptable, at least for the EnKF, when the assumed resistance error does not exceed 10%.

6. Comparison with existing methods

Considering the nature of the information on which the proposed KF methods are based, a comparison can be easily made with the performance of the methods proposed in [7], where a LASSO-based technique is applied to the CPI problem using exclusively energy consumption curves obtained from smart meters, and in [8], where PCA is considered for the same purpose.. Different scenarios are considered for the comparison, with increasing number of consumers and 1% s.d. in the

Table 6
Estimation results for different KF formulations. Scenario IV.

Error in r	UKF Hit Rate (%)	CKF Hit Rate (%)	EnKF Hit Rate (%)
5%	95	100	100
10%	91	95	98
15%	86	91	95
20%	80	84	89

Table 7
Rate of correct assignments for the PCA, the LASSO-based and the proposed KF-based methods.

Number of clients	PCA	LASSO	UKF	CKF	EnKF
50	100	100	100	100	100
100	100	100	95	100	100
200	92	100	90.5	97	100
300	72.67	87.67	71	86.33	94
400	60.25	71.75	57.75	78.25	86.5

measurement errors for all cases.

The rate of correct assignments for each method in the different cases studied are summarized in Table 7, where it is observed that:

- The performance of the PCA method is similar to that of the UKF in all the studied scenarios.
- The LASSO-based method has proven to be more sensitive to the number of loads than the CKF and the EnKF formulations, being the success rates of the three techniques very close for $N_c \leqslant 200$.
- The results obtained by the CKF and the EnKF schemes are substantially better when the number of loads increases, as can be noticed particularly for $N_c = 400$.

7. Conclusions

In this paper, a technique based on Kalman filtering and smart meters information is presented to identify the electrical phase which individual loads are connected to in distribution grids. For this purpose, the performance of three KF schemes, UKF, CKF and EnKF, is tested and compared.

The proposed estimation algorithm iteratively selects the customer with the highest probability to be connected to a certain phase, based on the estimated value of the corresponding state variable and the covariance of the estimation error. This way of handling binary variables prevents the computational problems potentially arising by the enforcement of the customary equality constraints, and might find application in other binary-constrained problems.

Four separate scenarios have been considered to test the accuracy and robustness of the KF formulation for feeders with increasing number of loads and increasing measurement and model errors. The amount of measurements required for a good performance of the algorithm is also analyzed. From the results obtained, the following conclusions are drawn:

- The UKF shows a poor behavior as the number of loads increases, achieving unacceptable success rates (<60%) when 400 customers are considered. Additionally, the robustness of this formulation in the presence of wrong measurements is weak.
- The performance of the CKF is better than that of the UKF, showing a lower sensitivity to the state vector size and wrong data.
- The EnKF has proven to be the best of the three KF schemes considered for the CPI problem, both in terms of success rates and sensitivity to noisy measurements.
- As expected, the performance of the KF-based schemes is affected by the number of energy measurements available, relative to the number of loads. The more loads in the same feeder, the more measurements are needed to correctly ascertain the phase connection.

A comparison of the CKF and EnKF with published methods, based on PCA and LASSO, has shown that the latter provide similar results to those of the KF only when a reduced number of loads is considered.

Further research efforts will be devoted to redesign the proposed KF-based procedure, so that it can take advantage of additional electrical quantities, such as P, Q and V, potentially available in the context of advanced DMS.

CRediT authorship contribution statement

M.A. González-Cagigal: Conceptualization, Methodology, Data curation, Writing - original draft, Writing - review & editing. **J.A. Rosendo-Macías:** Conceptualization, Methodology, Supervision, Writing - original draft, Writing - review & editing. **A. Gómez-Expósito:** Conceptualization, Methodology, Supervision, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors thank the Ministry of Education and Professional Training of Spain (grant FPU17/06380) and the project Pastora (PI-1897/12/2019) for the financial support.

References

[1] Pezeshki H, Wolfs PJ. Consumer phase identification in a three phase unbalanced LV distribution network. In: 2012 3rd IEEE PES innovative smart grid technologies Europe (ISGT Europe), Berlin; 2012. p. 1–7.
[2] Pezeshki H, Wolfs P. Correlation based method for phase identification in a three phase LV distribution network. In: 2012 22nd Australasian Universities Power Engineering Conference (AUPEC), Bali; 2012. p. 1–7.
[3] Blakely L, Reno MJ, Feng W. Spectral clustering for customer phase identification using AMI Voltage Timeseries. In: 2019 IEEE power and energy conference at Illinois (PECI), Champaign, IL, USA; 2019. p. 1–7.
[4] Chen CS, Ku TT, Lin CH. Design of phase identification system to support three-phase loading balance of distribution feeders. IEEE Trans Indus Appl 2012;48(1):191–8.
[5] Caird K. "Meter phase identification," U.S. Patent Application 20100164473, Patent No. 12/345702; Jan. 2010.
[6] Arya V, Seetharam D, Kalyanaraman S, Dontas K, Pavloski C, Hoy S, et al. Phase identification in smart grids. In: Smart grid communications (SmartGridComm), 2011 IEEE international conference on; 2011. p. 25–30.
[7] Tang X, Milanovic JV. "Phase Identification of LV Distribution Network with Smart Meter Data," In: 2018 IEEE Power & energy society general meeting (PESGM), Portland, OR; 2018. p. 1–5.
[8] Satya Jayadev P, Rajeswaran A, Bhatt NP, Pasumarthy R. A novel approach for phase identification in smart grids using Graph Theory and Principal Component Analysis. In: 2016 American Control Conference (ACC), Boston, MA; 2016. p. 5026–31.
[9] Xu M, Li R, Li F. Phase identification with incomplete data. IEEE Trans Smart Grid 2018;9(4):2777–85.
[10] Ren Pengxiang, Lev-Ari Hanoch, Abur Ali. Robust continuous-discrete extended Kalman filter for estimating machine states with model uncertainties. In: Power systems computation conference (PSCC); June 2016.
[11] Ning Zhou, Da Meng, Zhenyu Huang, Huang G. Dynamic state estimation of a synchronous machine using PMU data: A comparative study. IEEE Power Energy Soc General Meet 2015.
[12] Qi J, Sun K, Wang J, Liu H. Dynamic state estimation for multi-machine power system by unscented Kalman filter with enhanced numerical stability. In: IEEE Transactions Smart Grid; March 2016.
[13] Rouhani A, Abur A. Constrained Iterated Unscented Kalman Filter for Dynamic State and Parameter Estimation. IEEE Trans Power Syst 2018;33(99):1. https://doi.org/10.1109/TPWRS.2017.2764005.
[14] González-Cagigal MA, Rosendo-Macías JA, Gómez-Expósito A. Parameter estimation of fully regulated synchronous generators using Unscented Kalman Filters. Electric Power Syst Res 2018;168(2019):210–7.
[15] Wehbe Y, Fan L. "UKF based estimation of synchronous generator electromechanical parameters from phasor measurements." In: North American Power Symposium (NAPS); Sept. 2012.
[16] Simon D. Optimal State Estimation: Kalman, H Infinity and Nonlinear Approaches. ISBN: 13 978-0-471-70858-2.
[17] Pesonen H, Piché R. "Cubature-based Kalman filters for positioning". In: Positioning Navigation and Communication (WPNC), 2010 7th Workshop on, vol., no., pp. 45–49, 11–12, March 2010.
[18] Afrasiabi S, Afrasiabi M, Rastegar M, Mohammadi M, Parang B, Ferdowsi F. Ensemble Kalman Filter based Dynamic State Estimation of PMSG-based Wind Turbine. 2019 IEEE Texas Power and Energy Conference (TPEC), College Station, TX, USA. 2019. p. 1–4.
[19] Arpan Koirala, Pablo Arboleya, Bassam Mohamed, Suárez-Ramón Lucía. Non-Synthetic European Low Voltage Test System. Int J Electr Power Energy Sys 2019;118:10.1016.
[20] Hyun JJ, Hyung-Chul L. Analysis of scaling parameters of the batch unscented transformation for precision orbit determination using satellite laser rangingdata. J Astron Space Sci 2011;28:183–92.

Annex II.4

A. Gómez-Expósito, J.A. Rosendo-Macías, M.A. González-Cagigal, "Monitoring and Tracking the Evolution of a Viral Epidemic Through Nonlinear Kalman Filtering: Application to the Covid-19 Case", in: IEEE Journal of Biomedical and Health Informatics, under minor revision.

Monitoring and Tracking the Evolution of a Viral Epidemic Through Nonlinear Kalman Filtering: Application to the Covid-19 Case

A. Gómez-Expósito, *Fellow, IEEE*, J. A. Rosendo-Macías, *Senior Member, IEEE*, and M. A. González-Cagigal

***Abstract*— This work presents a novel methodology for systematically processing the time series that report the number of positive, recovered and deceased cases from a viral epidemic, such as Covid-19. The main objective is to unveil the evolution of the number of real infected people, and consequently to predict the peak of the epidemic and subsequent evolution. For this purpose, an original nonlinear model relating the raw data with the time-varying geometric ratio of infected people is elaborated, and a Kalman Filter is used to estimate the involved state variables. A hypothetical simulated case is used to show the adequacy and limitations of the proposed method. Then, several countries, including China, South Korea, Italy, Spain, UK and the USA, are tested to illustrate its behavior when real-life data are processed. The results obtained clearly show the beneficial effect of the severe lockdowns imposed by many countries worldwide, but also that the softer social distancing measures adopted afterwards have been almost always insufficient to prevent the second virus wave.**

***Index Terms* - Nonlinear Kalman filtering, parameter estimation, Covid-19, geometric series.**

I. Introduction

Despite the spectacular medical advances of the 20th century, and the practical eradication of viral diseases that in the past caused great mortality (e.g., smallpox), modern societies are still very vulnerable to the sudden appearance of new viruses, such as the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), cause of the coronavirus disease 2019 (Covid-19), for which there is still no vaccine. In addition, once a viral outbreak originates in a region of a country (in the case of the Covid-19, the Chinese region of Hubei, where the first reported case was dated on December 2019), the globalization of the economy and mass tourism spread it almost inevitably and quickly to the rest of the world.

In the absence of effective treatments, once a certain threshold has been passed, the main and almost sole remedy against the spread of the disease to the entire population is social distancing, the objective of which is to minimize the contact between people, and therefore morbidity [1]. In extreme cases, when the speed of propagation of the outbreak is very high, massive lockdowns of entire countries may be needed, which cannot last indefinitely owing to their drastic impact on the economic activity.

For this reason, all the agents involved (governments, international organizations, institutions, companies and individuals) have the greatest interest in knowing how the number of affected and deceased people will evolve over time, with a view, on the one hand, to verifying the beneficial effects of social distancing, and on the other to scheduling the already saturated health resources and taking the economic measures intended to mitigate as far as possible the devastating effects of an epidemic like that of Covid-19.

Scientists, engineers, economists, etc. are acquainted with several mathematical and statistical toolkits (recently renamed collectively as "data analytics") for the treatment and filtering of time series, with a view to extracting useful information from the available data, uncertain by definition, such as trends, patterns, average values, expected variances, etc. In the specific case of a viral epidemic, such as that of Covid-19, there are basically two categories of models for processing the information:

1. Models that try to characterize the "physical" reality explaining the observed data. In the case of a viral epidemic, these models [2], [3] consider, for example, what fraction of people are at work, in teaching or travelling, how long it takes for an infected person to manifest symptoms, what is the mortality rate according to age groups, etc. This type of modeling is widely used in engineering, because the dynamics of the underlying systems or devices are generally well characterized, through mathematical relationships obtained from the physical laws that govern them (such is the case, for example, of electrical networks or an artificial satellite).

2. Models that try to determine explanatory parameters or

Manuscript submitted for review on May 11, 2020. The third author thanks the financial support of the Spanish Ministry of Education and Professional Training (grant FPU17/06380).

All authors are with the Department of Electrical Engineering, Laboratory of Engineering for Energy and Environmental Sustainability, University of Seville, Spain (e-mails: age@us.es, rosendo@us.es, mgcagigal@us.es).

variables from a purely mathematical point of view ("black box" approach), without going into the causes or interactions between components that explain the resulting data. Given uncertain data, which enter the system regularly (in our case, every day), the aim is to characterize its temporal evolution by adjusting the parameters of a mathematical model, so that the differences between what is observed and what is estimated are minimized. Two variants can be considered in this category:

2.a) Mathematical models that do not assume *a priori* what the shape of the temporal evolution of the involved magnitudes will be, but rather use a state transition equation, which tries to capture the dynamics of the problem in question by relating the variables in an instant of time to the variables in the previous instant. In this case, it is a matter of determining how the coefficients that define this equation evolve over time. In the case of epidemics, among the most popular models are those derived from the SIR (Susceptible-Infected-Recovered) model, [4], such as the one used for example in [5] to analyze the evolution of Covid-19 in Italy. This model is also considered in [6], where the evolution of the epidemic is forecasted using a novel state filtering algorithm.

2.b) Mathematical models based on the assumption that the evolution of infected people, deceased, etc. obeys a predetermined curve (based on the experience of previous epidemics), whose coefficients are estimated based on the time series of reported data. For example, the evolution of the accumulated number of infected people can be satisfactorily approximated by means of a sigmoid curve, as assumed in [7], where the curve proposed by Gompertz [8] is used.

The methodology proposed in this work belongs to the second category. As explained in the next section, we depart from the basic SIR model, by considering that the number of susceptible people, being large enough and changing relatively slowly, does not have to be explicitly considered in the model, but can be rather embedded in other equally significant parameters, such as the time-varying ratio of the geometric series characterizing the progression of affected people. Moreover, the proposed model explicitly distinguishes between people who have proved positive in a test, and people actually infectious, who are many more and for whom there is no reliable information available.

In this work, a Kalman filter (KF) is used to process both the assumed dynamic model and the information available throughout the outbreak. The KF, proposed for linear dynamic systems in the early 1960s, is considered one of the fundamental tools that allowed men to walk on the moon, as it was successfully used in guiding the Apollo program space missions [9]. This filter, which constitutes a generalization of the technique known as "recursive least squares", estimates the maximum likelihood evolution (that is, the most statistically probable, according to the assumed uncertainties and the observed samples) of the state of a dynamic system, and can be generalized to the non-linear case, including situations where the model parameters are also to be estimated.

Reference [10] applies the KF for the estimation of the evolution of AIDS, while several recent studies related to the Covid-19 have arisen. In [11] the KF is used to deal with the estimation of the reproductive number of the virus. A short-term prediction model is proposed in [12], where the time update equations of the estimator are used for future forecasts of the pandemic spread. ARIMA models are combined with a KF in [13] to track the evolution of the Covid-19 in Pakistan. Unlike in those references, where the parameters involved in the state estimation process are supposed to be known, in this work such assumptions are not required. This is the major distinguishing feature of the proposed methodology, compared to the state of the art, and the main contribution of the paper.

II. Proposed Model

We start from the well-known and simple SIR model [4], mathematically described by:

$$\dot{S}_c(t) = -\beta \cdot S_c(t) I(t)/N$$
$$\dot{I}(t) = \beta \cdot S_c(t) I(t)/N - \gamma \cdot I(t)$$
$$\dot{R}_c(t) = \gamma \cdot I(t)$$

where $S_c(t)$ and $R_c(t)$ are, respectively, the cumulative or total susceptible and recovered people, $I(t)$ represents the *active* infectious (not to be confused with cumulative infectious), β and γ are the transmission and recovery rates, and N is the total population of the studied region, satisfying $N = S_c(t) + I(t) + R_c(t)$. Note that, in this compact model, the deceased cases are paradoxically included in $R_c(t)$ (alternatively, they could be subtracted from N).

For practical purposes, the discrete counterparts obtained by numerical integration (forward Euler) are rather of interest. Moreover, as dead people are separately reported, they can be explicitly modeled, leading to a discrete-time SIRD (Susceptible-Infected-Recovered-Deceased) model, as used in [14-16]:

$$S_c(n+1) = S_c(n) - \beta \cdot S_c(n) I(n)/N \quad (1)$$
$$I(n+1) = I(n) + \beta \cdot S_c(n) I(n)/N - (\gamma + \mu) I(n) \quad (2)$$
$$R_c(n+1) = R_c(n) + \gamma \cdot I(n) \quad (3)$$
$$D_c(n+1) = D_c(n) + \mu \cdot I(n) \quad (4)$$

where n is the elapsed time (in days) from a given origin, $D_c(n)$ is the cumulative dead and μ is a mortality ratio.

The data publicly reported (available in references such as [17],[18]), typically comprise the following three items:

- Fraction of infectious people who, subject to a test, yield a positive outcome. This considers the fact that there may be many more infected than those reported positives, as happens with a large number of asymptomatic people. The cumulative positives will be denoted $P_c(n)$.

- Fraction of recovered people who have been previously identified as positive. For simplicity of notation, the same symbol as in the basic SIR model, $R_c(n)$, will be used in the sequel, even though we are referring here to a subset of R_c.
- Cumulative number of deceased, $D_c(n)$, which is assumed to be the same as in the SIRD model, even though the actual number of dead by the virus may differ from the reported figures.

Some sources directly provide the *active* positive cases, $P(n)$, defined as: $P(n) = P_c(n) - R_c(n) - D_c(n)$. Note that both $I(n)$ and $P(n)$ tend to zero as n increases sufficiently (end of the viral outbreak), while the remaining cumulative magnitudes asymptotically reach a maximum or steady-state value.

Epidemiologists use the so-called basic reproductive number, R_0 (average number of people infected by a single infectious person during the infective period at the onset of the outbreak) to characterize whether and how fast an epidemic spreads at the very beginning. If R_0>1, then the epidemic will progress exponentially. As time elapses, though, the number of susceptible people decreases, either by the virus evolution itself or as a consequence of social distancing measures, and R_0 is replaced by the effective reproductive number, R_t. In terms of SIRD coefficients, R_0 and R_t are given by:

$$R_0 = \beta/(\gamma + \mu) \quad ; \quad R_t(n) = R_0 S_c(n)/N$$

However, as thoroughly discussed in [19], the basic reproductive number R_0 is not free from ambiguity and controversy. For instance, it is stated in [19] that "using R_0 as a threshold parameter for a population-level model could produce misleading estimates of the infectiousness of the pathogen, the severity of an outbreak, and the strength of the medical and/or behavioral interventions necessary for control". Moreover, if R_0 is estimated from time series of reported data, as in [11], then there is no way, at least for a new virus such as Covid-19, to subsequently check or contrast the accuracy of the estimates. This probably explains the wide confidence intervals so far reported for R_0 values [20]. Similar arguments apply to R_t.

For this reason, instead of or in addition to R_0, we postulate in this work the use of a more intuitive and measurable index, related with the growth rate of the infected class, to duly and unambiguously characterize a viral epidemic. Let the daily evolution of the active infectious be expressed as a geometric time series:

$$I(n+1) = \boldsymbol{r}(\boldsymbol{n}) \cdot I(n) \tag{5}$$

where $r(n)$ is the time-varying ratio of the series. Then the daily growth rate is obtained from:

$$\text{Growth rate (p.u.)} = \Delta I(n)/I(n) = r(n) - 1$$

Clearly, as long as $r(n)$>1, the viral outbreak will continue its expansion, whereas the disease extinguishes when $r(n)$<1. There is no ambiguity in using $r(n)$ as a threshold, when referred to a whole population. Note however that, if (5) were expressed in terms of cumulative magnitudes, rather than daily or active cases, then $r(n)$ would tend asymptotically to 1.

By direct comparison of (5) with (2), the following relation is obtained,

$$r(n) = 1 + \beta \cdot S_c(n)/N - (\gamma + \mu)$$

or, in terms of R_t:

$$r(n) = 1 + (\gamma + \mu)[R_t(n) - 1]$$

Given $r(n)$, one still would have to guess the values of the parameters involved in the SIRD model (1)-(4), to obtain R_0. We contend that there is no need to worry in the short term about R_0, as $r(n)$ suffices to duly track the epidemic evolution on a daily basis.

This work is aimed at estimating, from the daily reported data, the evolution of $r(n)$ and, as a consequence, the growth rate of the infectious people. Note that, if $r(n)$ can be somehow estimated, then equation (1) becomes unnecessary. In our approach, the impact of susceptible people, a factor which varies smoothly, is also embedded into $r(n)$.

In order to take advantage of the reported numbers of positive, deceased, and recovered cases, the following relationships are considered, taking into account (3)-(4):

$$P(n) = t(n) \cdot I(n) \tag{6}$$
$$D(n) = \mu(n) \cdot I(n) \tag{7}$$
$$R(n) = t(n) \cdot \gamma(n) \cdot I(n) = \gamma_t(n) \cdot I(n) \tag{8}$$

where $t(n)$ is a testing or reporting ratio that models the fraction of those infectious who are subject to tests and yield positive, $D(n) = D_c(n) - D_c(n-1)$ is the daily increase in the number of deaths, and $R(n) = R_c(n) - R_c(n-1)$ is the daily variation in the number of recovered cases.

From (7) at two consecutive instants, keeping (5) in mind:

$$D(n+1) = \mu(n+1) \cdot r(n) \cdot I(n)$$
$$D(n) = \mu(n) \cdot I(n)$$

and dividing:

$$r_D(n) = r(n) \cdot K_\mu(n) \tag{9}$$

where $r_D(n) = D(n+1)/D(n)$ is the ratio of consecutive daily deaths and $K_\mu(n) = \mu(n+1)/\mu(n)$ is in turn a ratio of consecutive mortality ratios.

Similarly, from (6) and (5):

$$P(n+1) = t(n+1) \cdot r(n) \cdot I(n)$$
$$P(n) = t(n) \cdot I(n)$$

and dividing:

$$r_P(n) = r(n) \cdot K_t(n) \quad (10)$$

where $r_P(n) = P(n+1)/P(n)$ is the ratio of consecutive daily positives and $K_t(n) = t(n+1)/t(n)$ is a ratio of consecutive $t(n)$ coefficients.

Finally, from (8) and (5):

$$R(n+1) = \gamma_t(n+1) \cdot r(n) \cdot I(n)$$
$$R(n) = \gamma_t(n) \cdot I(n)$$

and dividing:

$$r_R(n) = r(n) \cdot K_\gamma(n) \quad (11)$$

where $r_R(n) = R(n+1)/R(n)$ is the ratio of daily recovered cases and $K_\gamma(n) = \gamma_t(n+1)/\gamma_t(n)$ is a ratio of consecutive $\gamma_t(n)$ coefficients.

III. Kalman Filter Application and Implementation

From the above simple model, given by (9)-(11), a system of state equations can be formulated allowing the sequence $r(n)$ to be estimated by means of a non-linear KF, such as the Extended KF (EKF), Unscented KF (UKF) or Ensemble KF (EnKF) [9], [21]. This type of filter is capable of estimating the dynamic evolution of both parameters and state variables. Even though its maximum likelihood has only been proven for linear problems, it is applied successfully in nonlinear problems, such as the one in hand.

In our proposal, the state vector is composed of the infectious geometric ratio, $r(n)$, not directly measured, and the gains $K_\mu(n)$, $K_t(n)$ and $K_\gamma(n)$, all of them initially assumed to evolve with a random walk. Therefore, the state vector is

$$x = [K_\mu \quad K_t \quad K_\gamma \quad r]^t$$

and the state transition equation:

$$x(n) = x(n-1) + w(n)$$

where $w = [w_\mu \quad w_t \quad w_\gamma \quad w_r]^t$ is a Gaussian noise vector with a covariance matrix Q(n), which accounts for model errors such as possible time lags not duly considered in the model.

The state is to be estimated with the help of a measurement vector, composed of $r_D(n)$, $r_P(n)$ and $r_R(n)$, along with the pseudo-measurements of $K_\mu(n)$, $K_t(n)$ and $K_\gamma(n)$. Thus, assuming that the coefficients $\mu(n)$, $t(n)$ and $\gamma_t(n)$ change slowly, their variation ratios can be considered to lie around the unity. Therefore, the measurement vector is,

$$z(n) = [r_D(n) \quad r_P(n) \quad r_R(n) \quad 1 \quad 1 \quad 1]^t$$

and the measurement equation,

$$z(n) = h[x(n)] + v(n)$$

where $v = [v_D \quad v_P \quad v_R \quad v_\mu \quad v_t \quad v_\gamma]^t$ is a Gaussian noise vector that models the measurement error with a covariance matrix R(n), and the nonlinear measurement function is

$$h(x) = [r \cdot K_\mu \quad r \cdot K_t \quad r \cdot K_\gamma \quad K_\mu \quad K_t \quad K_\gamma]^t$$

With this formulation, the EKF is able to deal properly with the non-linearity arising in the measurement equation, performing at each iteration a linear prediction step, followed by the non-linear correction step. So, the KF provides the sequence of states estimates $\hat{x}(n)$, whose last component, the estimate of $r(n)$, will be denoted as $r_K(n)$.

The sequence $r_K(n)$ incorporates, in a statistically optimal fashion, the noisy information provided by raw data ratios, such as $r_D(n)$, $r_P(n)$ or $r_R(n)$, which results in a more reliable estimation than the raw data ratios themselves.

After several initial trials, an enhanced model has been finally implemented, considering two improvements:

1) The random walk in $r(n)$ can be advantageously replaced, while keeping the linearity of the state equation, by a linear prediction based on the recent past history, e.g., based on $r(n-1)$, $r(n-2)$ and $r(n-3)$. In this case,

$$x = [K_\mu \quad K_t \quad K_\gamma \quad r \quad r_{n-1} \quad r_{n-2} \quad r_{n-3}]^t$$

and

$$x(n) = A \cdot x(n-1) + w(n) \quad (12)$$

where:

$$A = \begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & 1 & & & & \\ & & & \frac{4}{3} & & & \frac{-1}{3} \\ & & & 1 & & & \\ & & & & 1 & & \\ & & & & & 1 & \end{bmatrix}, \quad w = \begin{bmatrix} w_\mu \\ w_t \\ w_\gamma \\ w_r \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

2) The information available at instant n can be used to improve previous estimations. For this purpose, the Rauch-Tung-Striebel (RTS) smoother [9], is implemented in two steps:

- Forward pass performed with EKF.
- Backward recursion smoother based on the linear state

transition equation.

Regarding the tuning of the filter, the diagonal elements of $R(n)$ related to v_D, v_P and v_R are self-tuned according to the respective sample variances of the last available days of $r_D(n)$, $r_P(n)$ and $r_R(n)$. The terms related to the pseudo-measurements are set to 10^{-4} to increase their weights. The diagonal elements of $Q(n)$ are set to constant values $Q_{i,i} = 10^{-3}$, according to the observed variance of $r_K(n) - r_K(n-1)$, except in the case of history variables, where $Q_{i,i} = 0$. The initial values of $x(0)$ and $P(0)$, required by the filter, are not relevant as their effect quickly vanishes. The proposed values are: $r_K(0)$ and history variables initialized to the mean of the raw ratios, K_i gains initialized to 1, and $P(0)$ initialized to an identity matrix.

Additionally, the results of the KF allow the estimation or prediction of two magnitudes of great interest:

1) The day in the future when the infection peak will be reached, n_p, which will occur when $r_k(n_p) = 1$. For this purpose, the sequence $r_K(n)$ can be fitted to a predetermined evolution in order to predict its future behavior from the past history. According to what can be observed empirically, we have fitted the sequence $r_K(n)$ to a decreasing exponential, characteristic of first-order systems. It is worth stressing that, at this point, we are talking about the peak of infected, the peak of positives being a proxy for it.

2) The estimated number of active infectious people, $I_K(n)$. This magnitude can be computed from those infected at a given reference day, n_0, as follows:

$$I_K(n) = \prod_{i=n_0}^{n-1} r_k(i) \cdot I(n_0) \qquad (13)$$

This means that, besides the errors of the sequence $r_K(n)$, the uncertainty of $I_K(n)$ inherits that of the initial guess $I(n_0)$. Therefore, as happens also with the SIR model, and in fact with any other model based on differential equations, precisely estimating $I_K(n)$ based on the above expression requires that an accurate number of infectious people be known on a given day, which can be very challenging. For our results, we will obtain the initial guess from $I_K(n_0) = P(n_0)/t(n_0)$, for an assumed value of $t(n_0)$. For instance, $t(n_0) = 0.2$ if we believe the first day there were 5 infectious people per reported positive.

IV. Simulation Results

In this section, the performance of the proposed implementation of the KF is tested on a set of simulated scenarios, where the SIR model described in Section II is considered for the propagation of a virtual virus.

While the total simulation time is 90 days, a lockdown is assumed to take place at day 15. This restrictive policy is aimed to completely stop non-essential public mobility, resulting in a quick reduction of the transmission rate, β (which is assumed to evolve exponentially, from an initial value 0.5 to 0.09) and, therefore, also of the ratio $r(n)$. The remaining simulation parameters are given in Table I.

In all simulated cases presented in the sequel, artificial Gaussian noise has been added to the measurements used by the KF algorithm in order to represent a more realistic scenario where the reported information presents inaccuracies.

TABLE I
Value of the Parameters in the Simulation

Parameter	Definition	Simulation value
μ	Mortality ratio	0.004
γ	Recovery rate	0.1
$I(0)$	Initial infectious	1000
N	Total population	47 Million
$t(n)$	Testing ratio	0.2

A. Base case

In the base case, the testing ratio, $t(n)$, is assumed to remain constant (except for the noise). The time evolution of the estimated geometric ratio, $r(n)$, is represented in Fig. 1 along with the raw noisy measurements provided by the simulation and the actual value of $r(n)$. Note that the estimation provided by the KF is very close to the simulated value, giving evidence of the good performance of the proposed method. Fig. 2 shows the benefit attained from the incorporation of the smoothing filter mentioned above to the basic EKF algorithm, in terms of more damped oscillations.

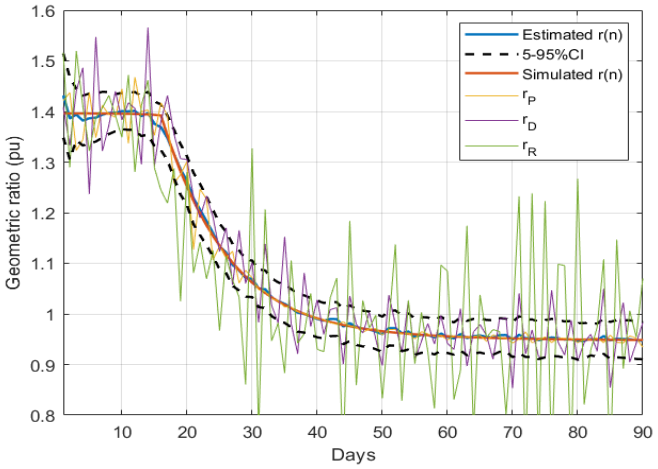


Fig. 1. Estimation of r(n) in the base case

To compare the proposed KF implementation with other methods customarily employed as filters in these cases, Fig. 3 shows the estimated value of $r(n)$ along with the results provided by three moving-average filters respectively applied to each of the noisy measurements, r_P, r_D and r_R. As can be seen, the KF more closely tracks the evolution of $r(n)$.

From the estimated $r(n)$ sequence, and the initial testing ratio, $t(0)$, an estimation is obtained for the evolution of infectious people, which is compared in Fig. 4 with the simulated value of $I(n)$ and the reported positives. The maximum estimation error

(around 4.5%) takes place, as expected, at the peak of the epidemic.

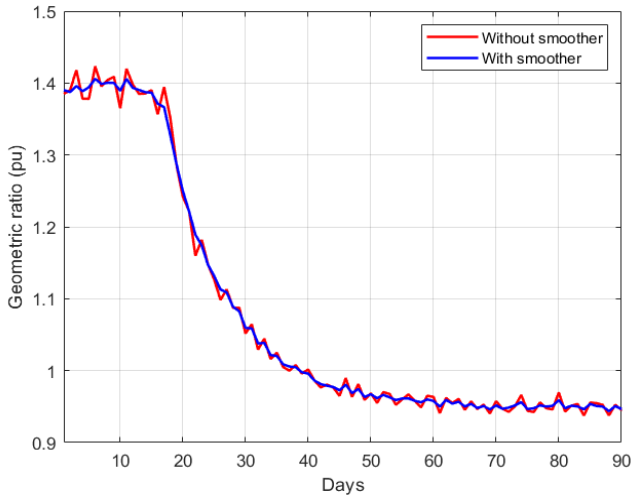


Fig. 2. Estimation of r(n) with and without the smoother

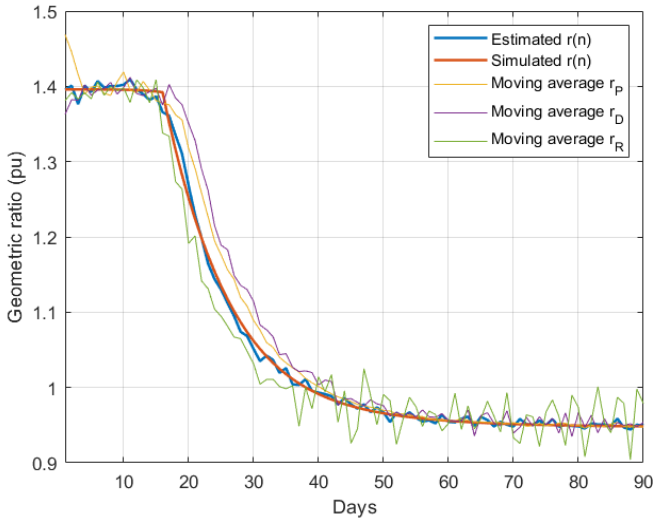


Fig. 3. Comparison of the proposed method with moving average filters

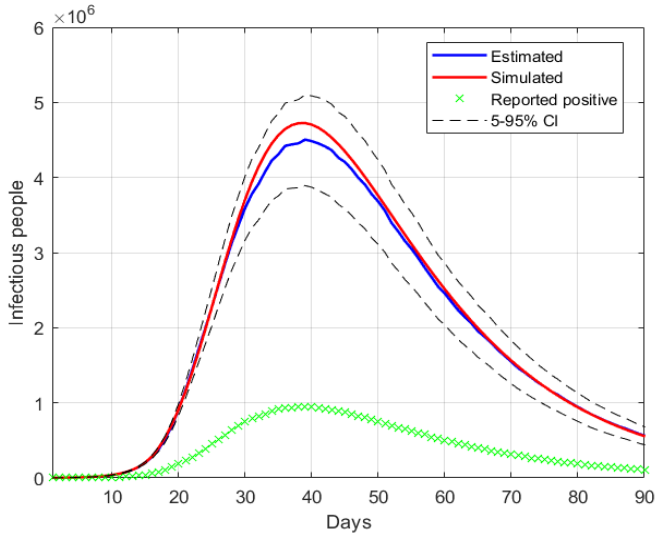


Fig. 4. Estimation of the infectious people in the base case

B. *Error assessment*

Once the proposed estimation technique is validated in the base case, where only Gaussian noise is considered, the effect of different error sources is studied in the following scenarios.

- Step in $t(n)$

An abrupt change is simulated in the testing ratio from $t(n)$ =0.2 to 0.3 at day 25, representing an increase in the availability of the tests (this has been observed in practice in several countries). Fig. 5 shows the estimation of $r(n)$, along with the simulated value and the measurements of the geometric ratios r_P, r_D and r_R. Note that the step in $t(n)$ is observed as an impulse in the ratio r_P, which is quite effectively filtered out by the proposed KF implementation.

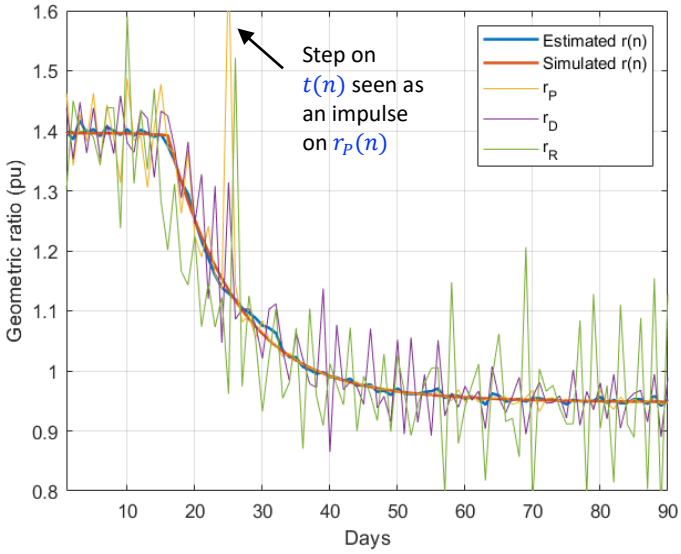


Fig. 5. Estimation of r(n) with a step on the testing ratio

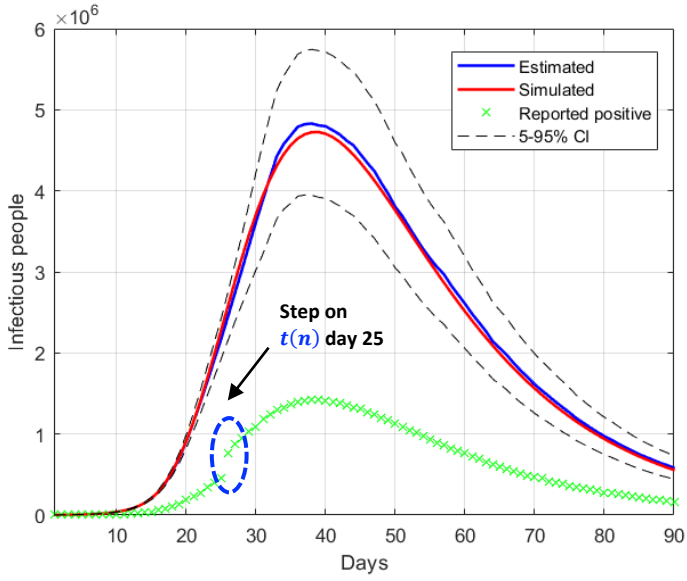


Fig. 6. Estimation of the infectious people with a step on the testing ratio

The estimation of $I(n)$ is represented in Fig. 6, where the actual simulated value is again very close to the estimated one, giving evidence of the good performance of the method in the presence of a step in the testing ratio.

- Deviations in $t(0)$

Finally, the last scenario considered in this section shows how errors in the initial guess of the testing ratio, $t(0)$, with respect to the assumed value $t(0) = 0.2$, affect the results. Given that the errors in this factor only affect the estimation of the infectious people, $I(n)$, the representation of the estimated $r(n)$ is not repeated. Fig. 7 shows the estimation of $I(n)$ for $t(0) = 0.15$ and $t(0) = 0.25$ ($\pm 25\%$ error).

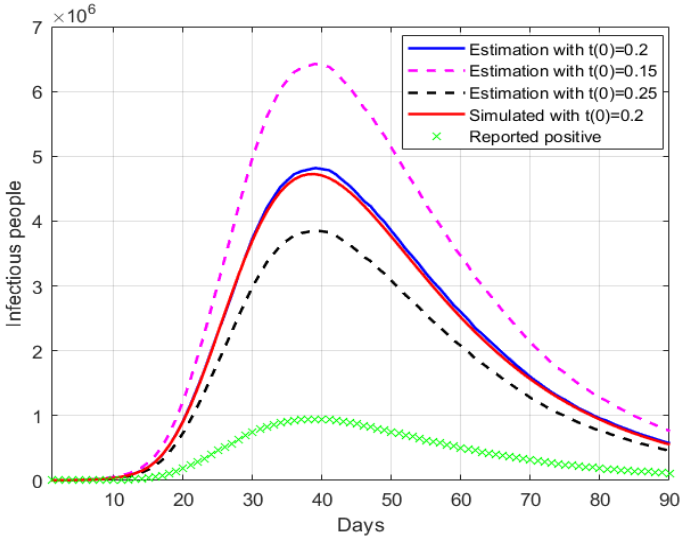


Fig. 7. Estimation of the infectious people for a range of t(0) values.

The results in Fig. 7 clearly show the importance of having an accurate guess of the number of infectious at the onset of the outbreak, as this initial error propagates proportionally up to the peak. Note, however, that any epidemiological model, such as SIR, faces the same challenge.

V. Case Studies

In this section, the proposed KF-based estimation technique is applied to the real data reported by different countries. For convenience, the results have been divided in two subsections: 1) the time period when massive lockdowns occurred in most countries, denoted in the media as the "first wave" of the pandemic [22], and 2) the subsequent transient period, once the lockdowns are relaxed, usually through several de-escalation phases, towards the so-called "new normality".

1. Lockdown period (first wave)

A total of four countries have been considered in this period: China, South Korea, Spain and the UK. At the early stage of the pandemic, the information provided by these countries was sufficient to allow the application of the proposed methodology.

Figs. 8-15 represent the estimated sequence of the geometric ratio, $r(n)$, and the number of infected people, $I(n)$, for the four countries. The KF implementation is tuned as described in Section III for the covariance matrices Q and R, and the initial values of the vector $x(0)$ and the covariance matrix $P(0)$. In order to estimate the number of infectious people, $I(n)$, according to (13), an initial value for the parameter $t(0)$ is needed. In absence of a better clue, $t(0) = 0.2$ is considered in all cases, expect for Spain (see the discussion of this particular case below). The points for which $r(n) = 1$ (peak of the epidemic) are highlighted with a dot. The following remarks can be made from those results:

- A different evolution of $r(n)$ can be observed for the Asian countries (China and South Korea), where the effects of Covid-19 started earlier. Once the geometric ratio $r(n) < 1$, the trend for South Korea is to remain roughly constant throughout the considered period, whereas for China a certain rebound can be noticed after March 10.

- The estimation results obtained for Spain show an asymptotic trend towards $r(n) = 0.95$. A slight increase is observed in $r(n)$ between April 10 and 15, probably influenced by a sudden increase in the number of tests.

  Regarding the parameter $t(0)$, in the Spanish case we have taken into account the results of a massive seroprevalence test performed by the government in the first half of May [23], from which it was concluded that the total number of infected people was around 5.2% of the population (approximately 2.3 million people). In view of this valuable information, the initial value $t(0)$ has been adjusted so that the cumulative number of infected people matches the result of the survey on the date it was released (May 13), leading to $t(0)$=0.12. This provides the estimation of $I(n)$ shown in Fig. 13, where a maximum value of the active infectious people of around 1.3 million can be noticed by mid-April. Fig. 16 represents the estimation of the cumulative infectious people for the Spanish territory, where the total number of infected people matches the results of the survey.

- With the available information in mid-May, some countries had already left behind the peak of the epidemic (i.e., $r(n) < 1$). For those cases (China, South-Korea and Spain), a rather accurate early forecasting of the epidemic evolution can be made, around 10 to 14 days before the peak, by fitting a decreasing exponential to a window of past estimated data. This prediction is shown with green dotted lines in Figs. 8-13.

- Regarding the UK, where the peak of the number of infectious people had not been reached in the period considered (i.e., $r(n) > 1$), an exponential fitting (made between around mid-April and early May) and the corresponding extrapolation is considered for this case. According to such fitting, the peak should take place in the second half of May.

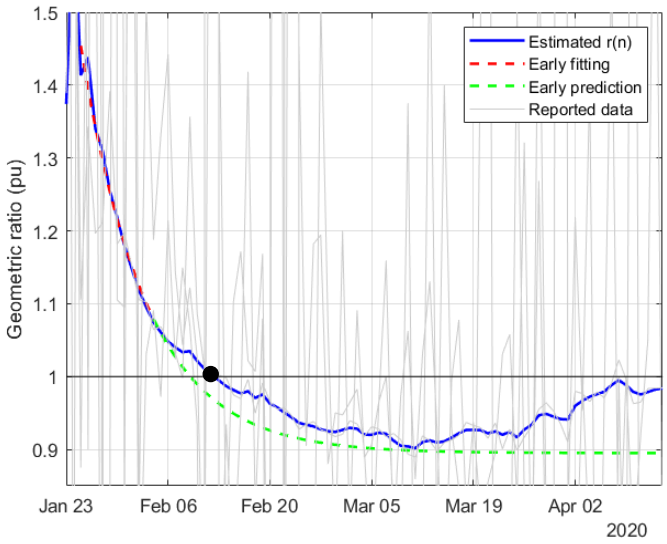


Fig. 8. Estimation of r(n) in China in the first period considered

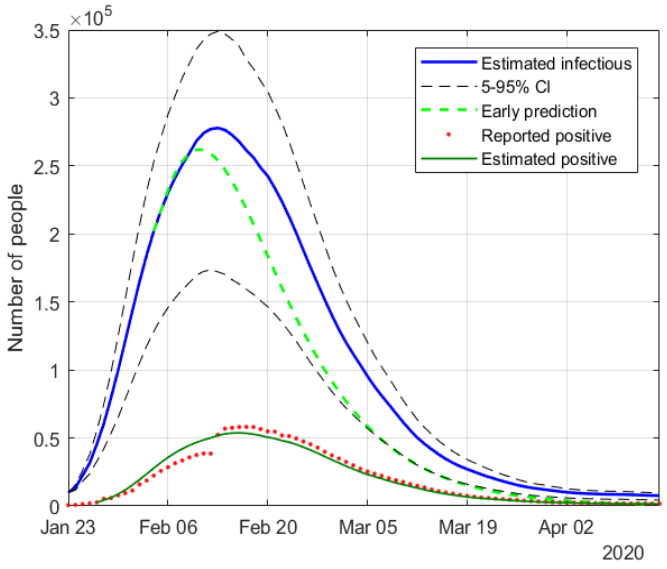


Fig. 9. Estimation of infectious people in China in the first period considered

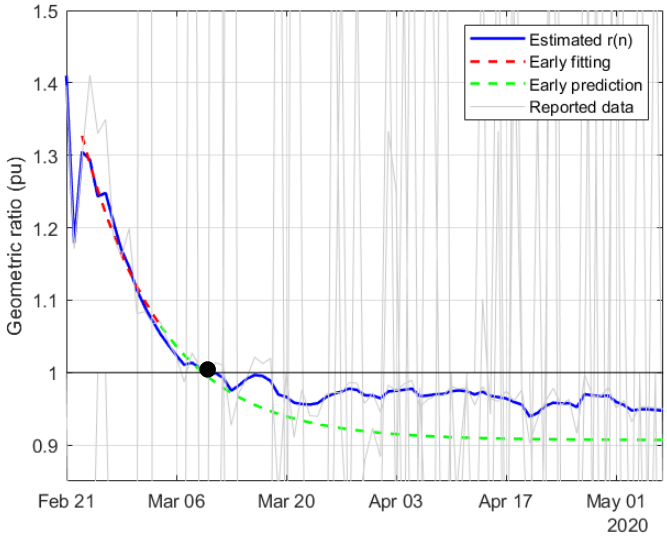


Fig. 10. Estimation of r(n) in South-Korea in the first period considered

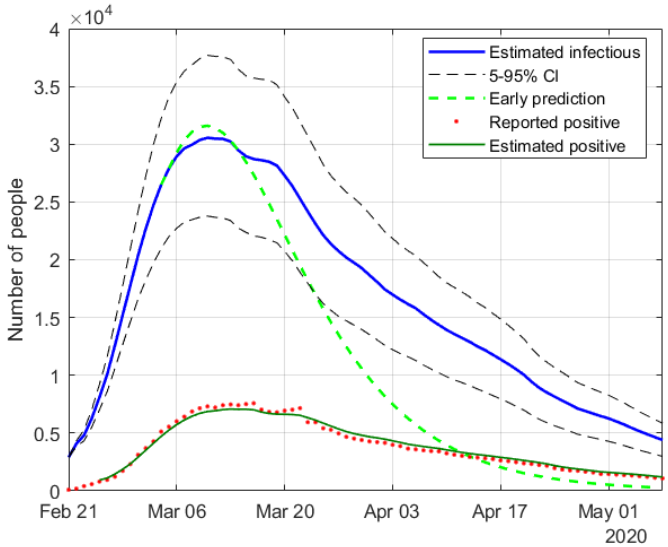


Fig. 11. Estimation of infectious people in South-Korea in the first period

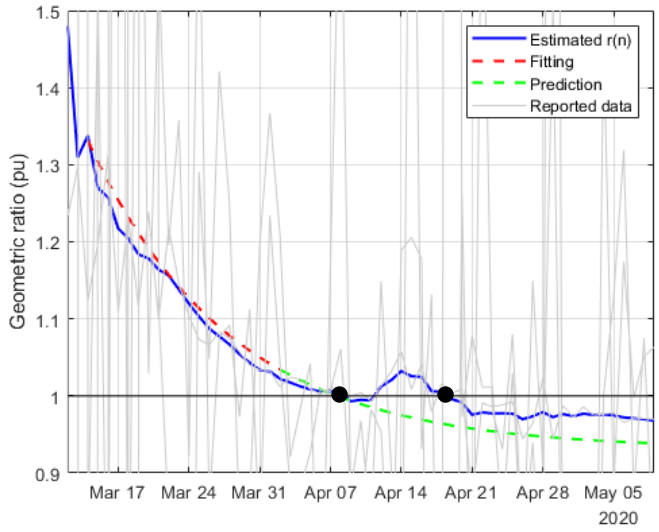


Fig. 12. Estimation of r(n) in Spain in the first period considered

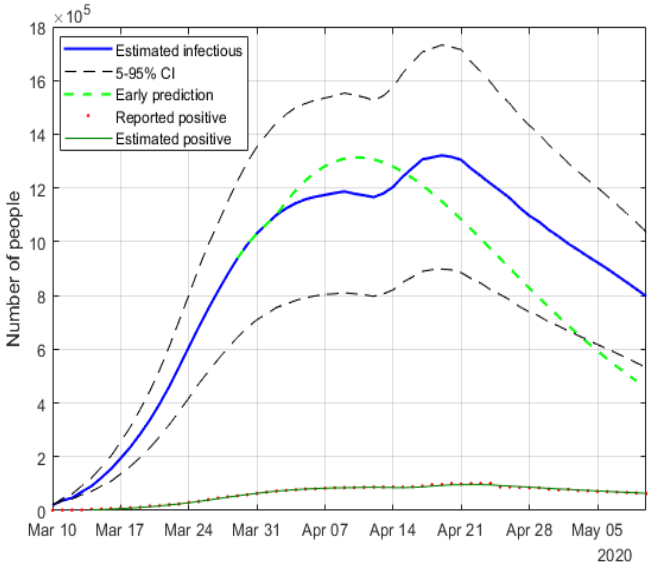


Fig. 13. Estimation of infectious people in Spain in the first period considered

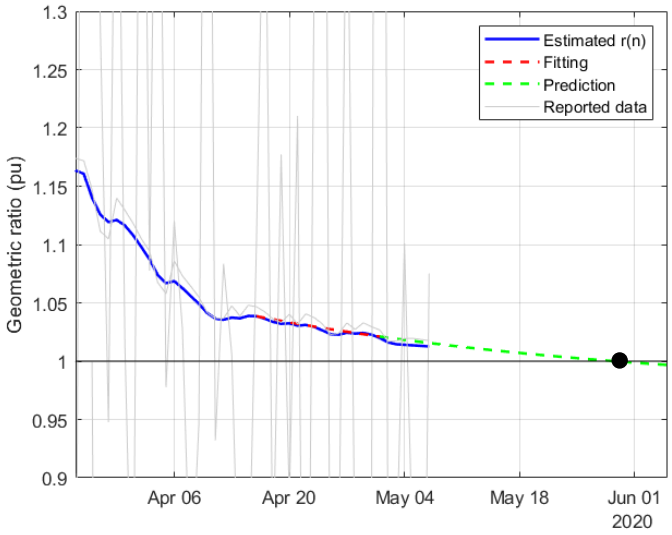


Fig. 14. Estimation of r(n) in UK in the first period considered

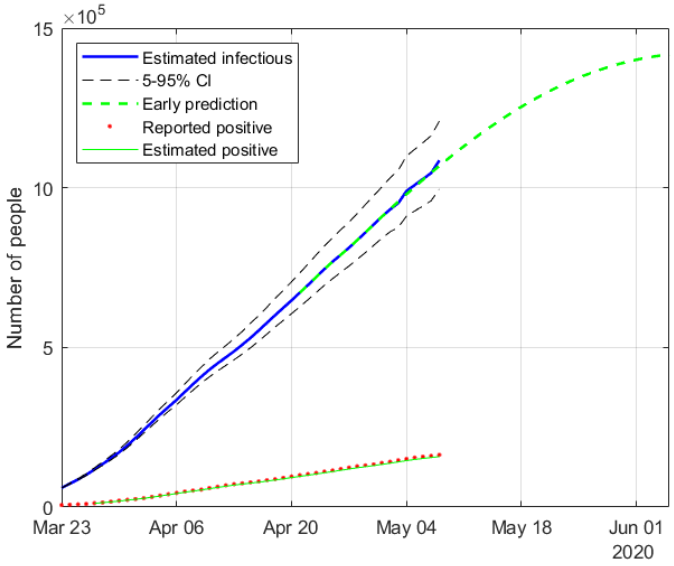


Fig. 15. Estimation of infectious people in UK in the first period considered

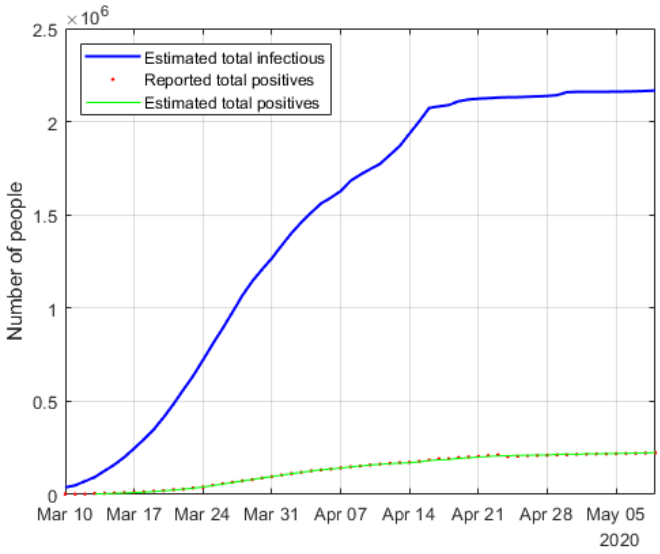


Fig 16. Estimation of cumulative infectious people in Spain

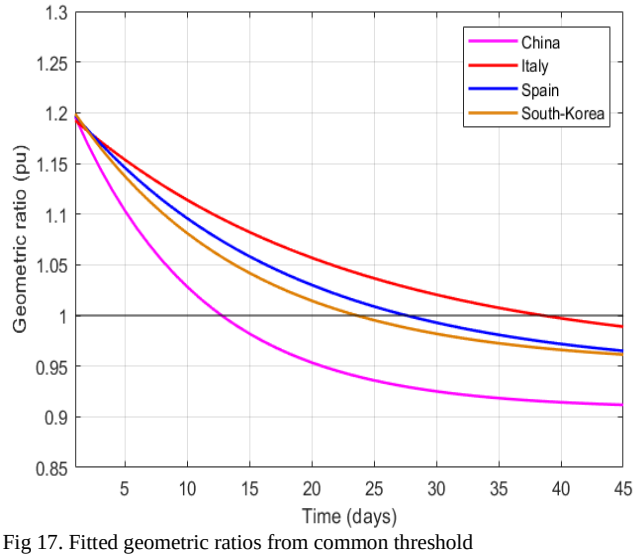


Fig 17. Fitted geometric ratios from common threshold

Finally, Fig. 17 shows the evolution of the above-mentioned fitted exponential curves for different countries (Italy has been included in the representation in order to establish a more complete comparison), all of them represented from a common threshold $r(n) = 1.2$, so that the corresponding time constants can be easily compared. In light of this representation, it can be noticed that the reduction of the geometric ratio is faster in China (just 13 days from $r(n) = 1.2$ to $r(n) = 1$), possibly as a consequence of a more severe lockdown, followed by Spain and South Korea (between 25 and 27 days to reach $r(n) = 1$), showing similar trends, and finally Italy (40 days to reach $r(n) = 1$).

2. Post-lockdown period (second wave)

As the pandemic evolves, it becomes more difficult to properly report on a regular basis all the information involved in the estimation of active positives. Many countries (notably Spain) stopped reporting the number of recovered people, probably owing to the remarkable increase in the number of asymptomatic positive cases, which never entered a hospital and hence never counted as recovered or dead. For this reason, it is not possible to accurately update the estimations of the geometric ratios of active positives, $r(n)$, for some of the countries considered in the early stages. Instead, Figs. 18-25 represent the estimated geometric ratio, $r(n)$, and the number of active infectious people, $I(n)$, for four countries (USA, Italy, India and Brazil), all of them specially affected by the pandemic and still reporting the information required by the proposed estimation technique.

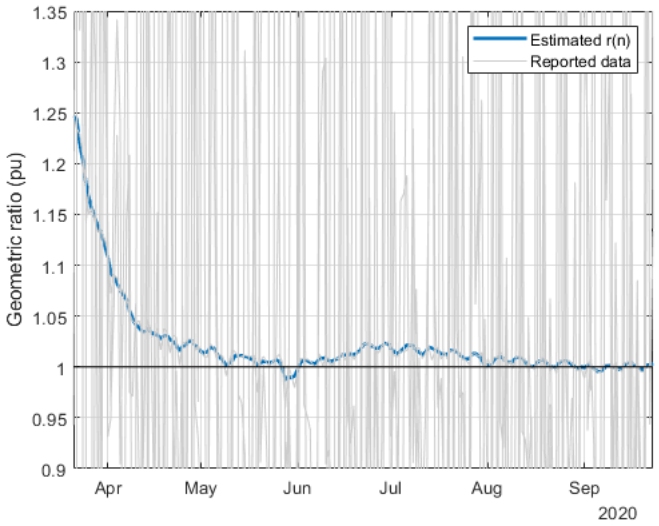


Fig. 18. Estimation of r(n) in USA in the second period considered

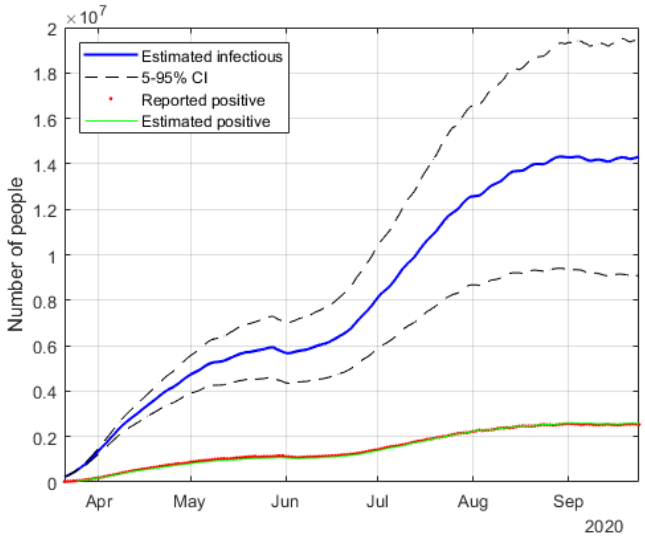


Fig. 19. Estimation of infectious people in USA in the second period considered

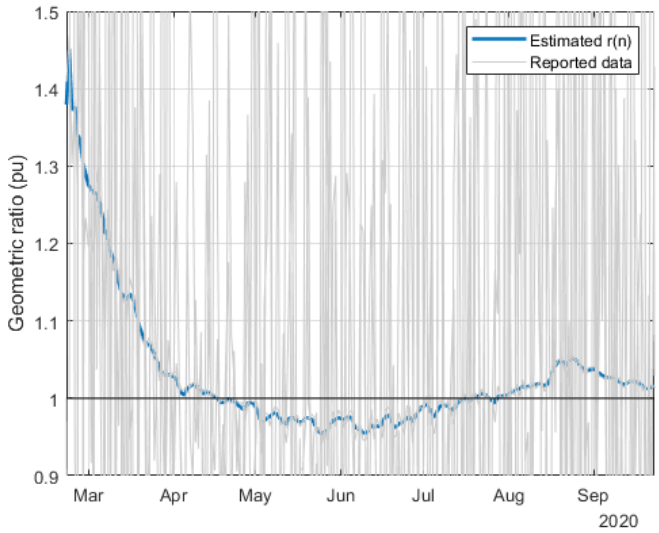


Fig. 20. Estimation of r(n) in Italy in the second period considered

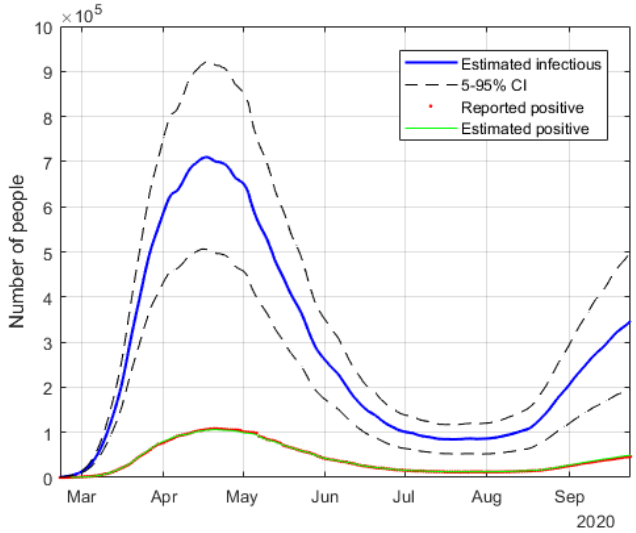


Fig. 21. Estimation of infectious people in Italy in the second period considered

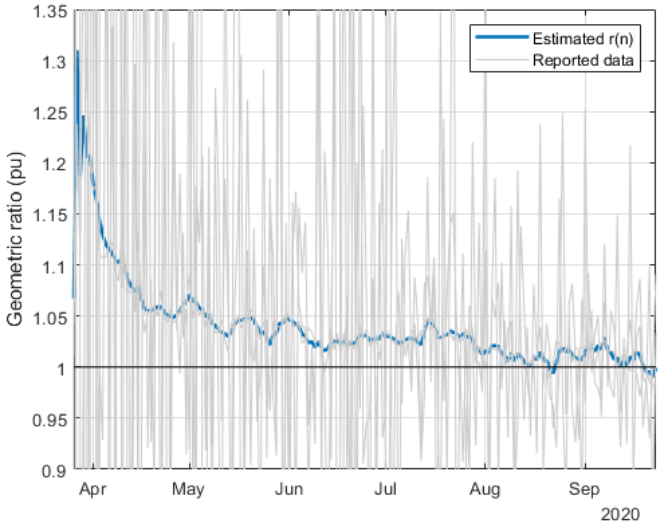


Fig. 22. Estimation of r(n) in India in the second period considered

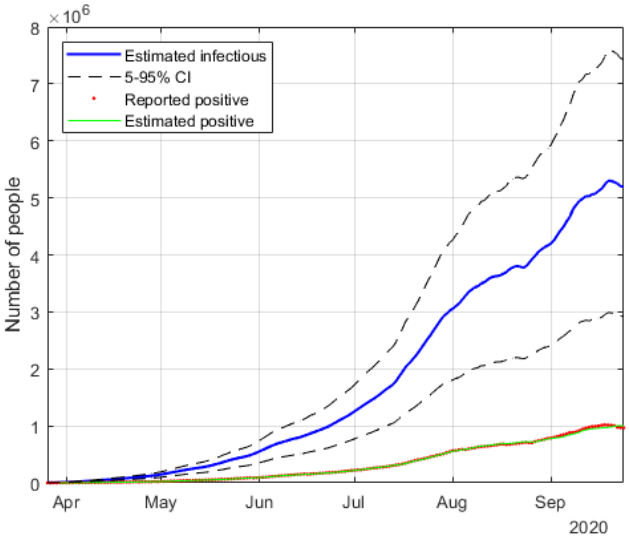


Fig. 23. Estimation of infectious people in India in the second period considered

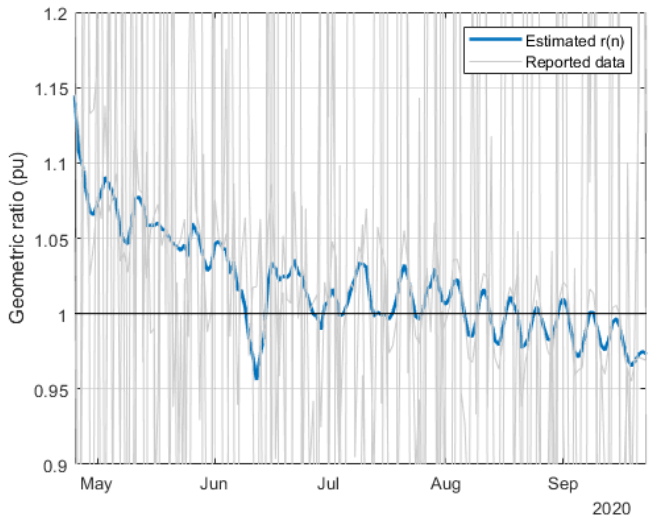


Fig. 24. Estimation of r(n) in Brazil in the second period considered

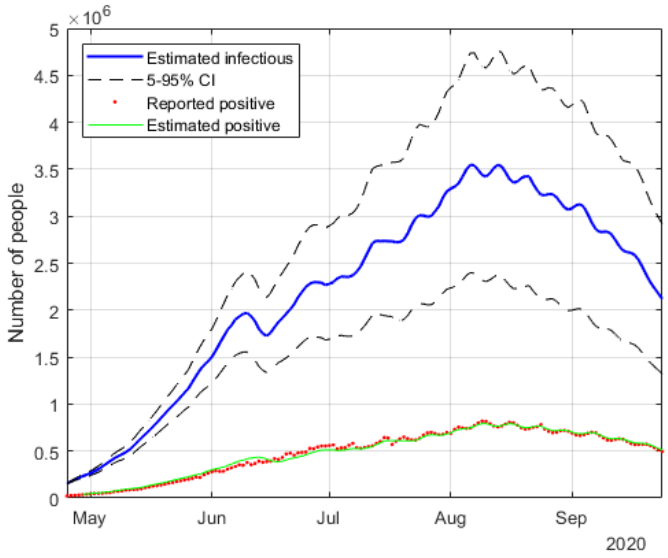


Fig. 25. Estimation of infectious people in Brazil in the second period considered

Similar assumptions as in the previous section are made regarding the KF tuning. The following remarks can be made from the results presented in Figs. 18-25:

- The number of infectious people in the USA briefly reached a peak by the end of May ($r(n) \approx 1$). However, Fig. 18 also shows that, afterwards, $r(n)$ has remained somewhat around or above one, which means that the outbreak is not still under full control, and that additional actions should be taken in this country in order to substantially reduce the number of infectious people.

- In Italy, the effect of relaxing the social distance measures can be easily noticed, with an increase of the ratio $r(n)$ from June. However, a new reduction is apparent in mid-August, probably due to a reinforcement in the severity of the mobility constraints, or simply owing to the holiday season.

- The curve of the geometric ratio in India is similar to that of the USA, with an asymptotic yet much slower trend to a value smaller than one. At the time of writing, it is observed that $r(n) > 1$, which means that the pandemic is still uncontrolled in this country.

- As far as Brazil is concerned, the most noticeable difference when compared to other countries lies in the almost periodic oscillations of $r(n)$ around the exponentially decreasing trend, the period being of about a week, which is probably due to the poor quality of the reported data.

VI. Conclusion

This work has addressed the problem of monitoring and tracking the evolution of a viral epidemic, such as Covid-19, through the application of signal processing techniques to the time series of data reported by governments and health agencies. Three main contributions can be pointed out: 1) the exclusive use of time-varying geometric ratios of daily data to track the disease, rather than the customary virus reproductive number (R_0); 2) the development of a simple algebraic model relating the geometric ratio of infectious people, $r(n)$, with those of positives, reported and dead; 3) the application of a nonlinear KF, along with a smoothing technique, to estimate the evolution of $r(n)$. By properly fitting the estimated values of $r(n)$ to a decreasing exponential, an accurate prediction of the epidemic peak can be made, as early as two weeks before the peak actually takes place.

The proposed methodology has been satisfactorily tested on a simulated case, in the presence of Gaussian noise and other sources of uncertainty, the main one being the number of infectious people at the onset of the outbreak.

The estimation technique has also been applied to a pool of countries, and the results obtained are divided in two periods:

- A first period, when most of the countries imposed a lockdown. Four territories are reported in this scenario, namely: China, South Korea, Spain, and the UK. The evolution of $r(n)$ reflects in all cases the severity of the lockdown, allowing the first peak of the epidemic to be forecasted well in advance. In some cases, a slightly increasing trend is apparent in the evolution of this ratio once the lockdown is removed, suggesting that additional mobility restrictions might be necessary.

- For the countries that have continued reporting the required information, the estimation of $r(n)$ is extended up to the moment of writing this manuscript, reflecting the panoply of post-lockdown measures taken by most of them. In this case, four countries are reported: the USA, Italy, India and Brazil. The results show how the geometric ratio $r(n)$ keeps rather close to 1, or slightly above, which explains the onset of the second wave we are currently facing in many places.

In light of the presented results, it can be concluded that the proposed methodology can effectively characterize, by means of the ratio $r(n)$, the evolution of the virus spread, when adequate information of active positives, recovered and

deceased people is available. This information on the state and dynamics of the epidemic can be used by the governing authorities in order to take the corresponding actions:

- An increasing trend of the geometric ratio represents a virus spread which might turn out of control, especially when $r(n) > 1$, leading to more restrictive policies.

- On the contrary, values of $r(n) < 1$ with decreasing trend indicate a situation where the severity of the social distancing measures can be alleviated.

As shown in the simulated scenario, the proposed methodology is not only suitable for the Covid-19, but also for other pandemics that can be characterized using the SIRD model, and for which the required information is available. Future work is aimed to the application of KF-based estimators to new models that can arise with less informative scenarios.

REFERENCES

[1] Rohman, Ahsan, and Zaber (2020, May 19). Lockdown vs. Social Distancing: Need for Effective Communication, The Jakarta Post. https://www.thejakartapost.com/academia/2020/05/19/lockdown-vs-social-distancing-need-for-effective-communication.html

[2] T. Pueyo, "Coronavirus: Why you must act now". https://medium.com/@tomaspueyo/coronavirus-act-today-or-people-will-die-f4d3d9cd99ca

[3] Imperial College COVID-19 response team, "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand". https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf?referringSource=articleShare

[4] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc Royal Soc Math Phys Eng Sci. 1927;115(772):700–21.

[5] Giordano, G., Blanchini, F., Bruno, R. et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med (2020). https://doi.org/10.1038/s41591-020-0883-7

[6] Huang, Weipeng & Provan, Gregory. (2016). An Improved State Filter Algorithm for SIR Epidemic Forecasting. 10.3233/978-1-61499-672-9-524.

[7] Computational biology and complex systems (BIOCOMSC), UPC, "Analysis and prediction of COVID-19 for different regions and countries". https://biocomsc.upc.edu/en/covid-19/daily-report

[8] Madden LV. Quantification of disease progression. Protection Ecology 1980; 2: 159-176.

[9] D. Simon, "Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches". ISBN: 13978-0-471-70858-2

[10] B.Cazelles, N.P.Chau. Using the Kalman filter and dynamic models to assess the changing HIV/AIDS epidemic. Mathematical Biosciences. Vol. 140, Issue 2, March 1997, pp. 131-154

[11] C. Rondon-Moreno, F. Arroyo Marioli, F. Bullano, "Dynamics of Transmission and Control of COVID-19: A Real-time Estimation using the Kalman Filter doi: https://doi.org/10.1101/2020.04.19.20071886

[12] Singh, Dr & Kumar, Suraj & Dixit, Prachi & Bajpai, Manish. (2020). Kalman Filter Based Short Term Prediction Model for COVID-19 Spread. 10.1101/2020.05.30.20117416.

[13] Muhammad Aslam, Using the kalman filter with Arima for the COVID-19 pandemic dataset of Pakistan, Data in Brief, Volume 31, 2020, 105854, https://doi.org/10.1016/j.dib.2020.105854.

[14] Lin, Feng & Muthuraman, Kumar & Lawley, Mark. (2010). An optimal control theory approach to non-pharmaceutical interventions. BMC infectious diseases. 10. 32. 10.1186/1471-2334-10-32.

[15] Osemwinyen, Amenaghawon & Diakhaby, Aboubakary. (2015). Mathematical Modelling of the Transmission Dynamics of Ebola Virus. Applied and Computational Mathematics (New York). 4. 313-320. 10.11648/j.acm.20150404.19.

[16] Chatterjee, S., Sarkar, A., Chatterjee, S. et al. Studying the progress of COVID-19 outbreak in India using SIRD model. Indian J Phys (2020). https://doi.org/10.1007/s12648-020-01766-8

[17] https://www.worldometers.info/coronavirus/country/spain/

[18] https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series

[19] J. Li, D. Blakeley, R. J. Smith, "The Failure of R0". Computational and Mathematical Methods in Medicine Volume 2011, doi:10.1155/2011/527610

[20] Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. "High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2". Emerg Infect Dis. Jul. 2020. https://doi.org/10.3201/eid2607.200282

[21] Evensen, G., "Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics". Journal of Geophysical Research. 99 (C5): 143–162, May 1994.

[22] Lockerd Maragakis, Lisa, "First and second waves of Coronavirus", https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/first-and-second-waves-of-coronavirus

[23] https://www.newtral.es/estudio-de-seroprevalencia-un-5-de-espana-con-indicios-de-haber-pasado-covid-19/20200513/

REFERENCES

[1] «Universidad Politécnica de Valencia,» [En línea]. Disponible: https://www.upv.es/materiales/Fcm/Fcm08/fcm8_3.html. [Último acceso: Enero 2019].

[2] Pezeshki H, Wolfs P. "Consumer phase identification in a three phase unbalanced LV distribution network". In: 2012 3rd IEEE PES innovative smart grid technologies Europe (ISGT Europe), Berlin; 2012. P. 1–7.

[3] Pezeshki H, Wolfs P. "Correlation based method for phase identification in a three phase LV distribution network". In: 2012 22nd Australasian Universities Power Engineering Conference (AUPEC), Bali; 2012. P. 1–7.

[4] Tang X, Milanovic JV. "Phase Identification of LV Distribution Network with Smart Meter Data," In: 2018 IEEE Power & energy society general meeting (PESGM), Portland, OR; 2018. P. 1–5.

[5] Satya Jayadev P, Rajeswaran A, Bhatt NP, Pasumarthy R. "A novel approach for phase identification in smart grids using Graph Theory and Principal Component Analysis". In: 2016 American Control Conference (ACC), Boston, MA; 2016. P. 5026–31.

[6] Sontag, Eduardo (1998), "Mathematical Control Theory: Deterministic Finite Dimensional Systems". Second Edition, Springer, ISBN 0-387-98489-5

[7] Zhenyu Huang, Kevin Schneider and Jarek Nieplocha,"Feasibility studies of applying Kalman Filter techniques to power system dynamic state estimation", in: International Power Engineering Conference (IPEC), Dec. 2007.

[8] H. Khazraj, F. Faria da Silva and C. L. Bak, "A performance comparison between extended Kalman Filter and unscented Kalman Filter in power system dynamic state estimation," 2016 51st International Universities Power Engineering Conference (UPEC), Coimbra, 2016, pp. 1-6, doi: 10.1109/UPEC.2016.8114125.

[9] M. S. Sachdev, H. C. Wood and N. G. Johnson, "Kalman Filtering Applied to Power System Measurements for Relaying," in IEEE Power Engineering Review, vol. PER-5, no. 12, pp. 52-53, Dec. 1985, doi: 10.1109/MPER.1985.5528637.

[10] Paul Zarchan; Howard Musoff (2000). "Fundamentals of Kalman Filtering: A Practical Approach." American Institute of Aeronautics and Astronautics, Incorporated. ISBN 978-1-56347-455-2.

[11] E. Ghahremani, I. Kamwa, "Dynamic state estimation in power system by applying the extended Kalman filter with unknown inputs to phasor measurements," in: IEEE Trans. Power Syst. 26 (November (4)) (2011).

[12] Qi Jilong, Tian Yantao, Gong Yimin and Zhucheng, "A sensorless initial rotor position estimation scheme and an Extended Kalman Filter observer for the direct torque controlled Permanent Magnet Synchronous Motor Drive", in: International Conference on Electrical Machines and Systems, Oct. 2008.

[13] Xiaoliang Jiang, Pindong Sun and Z.Q. Zhu, "Modeling and simulation of parameter identification for PMSM based on EKF", in: International Conference on Computer, Mechatronics, Control and Electronic Engineering, Aug. 2010.

[14] Xi Xiao and Changming Chen, "Reduction of Torque Ripple Due to Demagnetization in PMSM Using Current Compensation", in: IEEE Transactions on Applied Superconductivity, Vol. 20 , Issue: 3 , June 2010.

[15] Z. Ning, M. Da, H. Zhenyu, G. Welch, "Dynamic state estimation of a synchronous machine using PMU data: a comparative study," IEEE Power & Energy Society General Meeting, January, 2015.

[16] S. Wang, W. Gao, A.P. Sakis Meliopoulos, "An alternative method for power system dynamic state estimation based on unscented transform," IEEE Trans. Power Syst. 27 (May (2)) (2012).

[17] J. Qi, K. Sun, J. Wang, H. Liu, Dynamic state estimation for multi-machine power system by unscented Kalman filter with enhanced numerical stability, IEEE Trans. Smart Grid (March) (2016).

[18] F.M. Taimah, A.N. Merzah, Power system dynamic state estimation based on Kalman filter, Int. J. Comput. Appl. (0975-8887) 154 (Nov, 11) (2016).

[19] J. Zhao, L. Mili, "Robust unscented Kalman filter for power system dynamic state estimation with unknown noise statistics," IEEE Trans. Smart Grid (October) (2017).

[20] A. Rouhani, A. Abur, "Constrained iterated unscented Kalman filter for dynamic state and parameter estimation," IEEE Trans. Power Syst. 33 (May (99)) (2018) 1, https://doi.org/10.1109/TPWRS.2017.2764005.

[21] M. Huang, W. Li, W. Yan, "Estimating parameters of synchronous generators using square-root unscented Kalman filter," Electr. Power Syst. Res. 80 (September (9)) (2010) 1137–1144.

[22] Y. Wehbe, L. Fan, "UKF based estimation of synchronous generator electromechanical parameters from phasor measurements," North American Power Symposium (NAPS), September, 2012.

[23] J.C.N. Pantoja, A. Olarte, H. Daz, "Simultaneous estimation of exciter governor and synchronous generator parameters using phasor measurements," Electric Power Quality and Supply Reliability Conference (PQ), June, 2014, pp. 43–49.

[24] S. Dutta Chowdhury, N. Senroy, "PMU data based online parameter estimation of synchronous generator," International Conference on Power Systems (ICPS), March, 2016.

[25] Divya G Pillai, A. Vivek and V. Srikanth, "Non-linear state estimation of PMSM using derivative-free and square-root Cubature Kalman Filter", in: International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), July 2017.

[26] Ienkaran Arasaratnam, "Cubature Kalman Filtering: Theory and Applications,",Ph.D. Thesis, School of Graduate Studies of McMaster University, Hamilton, Ontario 2009.

[27] Yan Sun and Qijun Chen, "Joint estimation of states and parameters of vehicle model using cubature Kalman filter," 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, 2016, pp. 977-982, Oct. 2016.

[28] Gopinath G. R. and S. P. Das, "Speed and position sensorless control of Interior Permanent Magnet Synchronous Motor using Square-root Cubature Kalman filter with joint parameter estimation," 2016 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), Trivandrum, 2016, pp. 1-5, Dec. 2016.

[29] «Wikipedia,» [En línea, Último acceso: Enero 2019]. Available: https://es.wikipedia.org/wiki/Filtro_de_Kalman.

[30] Bucy, R.S. and Kalman, R.E., "New results in linear filterig and prediction theory", in: ASME. J. Basic Eng., 83(1), pp. 95-108, 1961

[31] S Simon D., "Optimal State Estimation: Kalman H Infinity and Nonlinear Approaches". ISBN: 13 978-0-471-70858-2.

[32] H. Pesonen and R. Pich´e, "Cubature-based Kalman filters for positioning", in: Positioning Navigation and Communication (WPNC), 2010 7th Workshop on , vol., no., pp.45,49, 11-12, March 2010.

[33] Evensen, G. (1994). "Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics". Journal of Geophysical Research. 99 (C5): 143–162

[34] Afrasiabi S, Afrasiabi M, Rastegar M, Mohammadi M, Parang B, Ferdowsi F. Ensemble Kalman Filter based Dynamic State Estimation of PMSG-based Wind Turbine. 2019 IEEE Texas Power and Energy Conference (TPEC), College Station, TX, USA. 2019. p. 1–4.

[35] H. Ghassempour, Z. Miao, L. Fan, W. Jiang, D. Manjure, Identification of synchronous generator model with frequency control using unscented Kalman filter, Electr. Power Syst. Res. 126 (2015) 45–55.

[36] J.J. Hyun, L. Hyung-Chul, Analysis of scaling parameters of the batch unscented transformation for precision orbit determination using satellite laser rangingdata, J.Astron. Space Sci. 28 (2011) 183–192, https://doi.org/10.5140/JASS.2011.28.3.183.

[37] S. Julier, The scaled unscented transformation, American Control Conference, 2002. Proceedings of the 2002, IEEE Press, 2002, pp. 4555–4559.

[38] Milano F., Power System modelling and scripting. ISBN: 978-3-642-13668-9.

[39] Arpan Koirala, Pablo Arboleya, Bassam Mohamed, Suárez-Ramón Lucía. Non-Synthetic European Low Voltage Test System. Int J Electr Power Energy Sys 2019; 118:10.1016.

[40] M.A. González-Cagigal, J.A. Rosendo-Macías, A. Gómez-Expósito, "Parameter estimation of fully regulated synchronous generators using Unscented Kalman Filters", in: Electric Power Systems Research, Volume 168, 2019, Pages 210-217, November, 2018.

[41] M.A. González-Cagigal, J.A. Rosendo-Macías, A. Gómez-Expósito, "Parameter Estimation of Wind Turbines with PMSM using Cubature Kalman Filters," in: IEEE Transactions on Power Systems, Volume .35, 2020, pp. 1796-1804, May 2020.

[42] M.A. González-Cagigal, J.A. Rosendo-Macías, A. Gómez-Expósito, "Application of nonlinear Kalman filters to the identification of customer phase connection in distribution grids," in: International Journal of Electrical Power and Energy Systems, Volume 125, 2021.

[43] J.A. Rosendo-Macías; M.A. González-Cagigal, A. Gómez-Expósito, "PMU-Based Estimation of Renewable Power Plants Parameters." In: IEEE Powertech Conference. Milán (Italia). 2019

[44] A. Gómez-Expósito, J.A. Rosendo-Macías and M.A. González-Cagigal, "Modelo y análisis de la evolución de una pandemia vírica mediante filtros de Kalman: el caso del Covid-19 en España", In: https://idus.us.es/handle/11441/94508

[45] A. Gómez-Expósito, J.A. Rosendo-Macías and M.A. González-Cagigal, "Monitoring and tracking the evolution of a viral epidemic through nonlinear Kalman filtering: Application to the Covid-19 case", In: https://www.medrxiv.org/content/10.1101/2020.05.11.20098087v1

[46] Rohman, Ahsan, and Zaber (2020, May 19). Lockdown vs. Social Distancing: Need for Effective Communication, The Jakarta Post. https://www.thejakartapost.com/academia/2020/05/19/lockdown-vs-social-distancing-need-for-effective-communication.html

[47] T. Pueyo, "Coronavirus: Why you must act now".

https://medium.com/@tomaspueyo/coronavirus-act-today-or-people-will-die-f4d3d9cd99ca

[48] Imperial College COVID-19 response team, "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand".

https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf?referringSource=articleShare

[49] Computational biology and complex systems (BIOCOMSC), UPC, "Analysis and prediction of COVID-19 for different regions and countries". https://biocomsc.upc.edu/en/covid-19/daily-report

[50] Madden LV. Quantification of disease progression. Protection Ecology 1980; 2: 159-176.

[51] B.Cazelles, N.P.Chau. Using the Kalman filter and dynamic models to assess the changing HIV/AIDS epidemic. Mathematical Biosciences. Vol. 140, Issue 2, March 1997, pp. 131-154

[52] C. Rondon-Moreno, F. Arroyo Marioli, F. Bullano, "Dynamics of Transmission and Control of COVID-19: A Real-time Estimation using the Kalman Filter. doi: https://doi.org/10.1101/2020.04.19.20071886

[53] Singh, Dr & Kumar, Suraj & Dixit, Prachi & Bajpai, Manish. (2020). Kalman Filter Based Short Term Prediction Model for COVID-19 Spread. 10.1101/2020.05.30.20117416.

[54] Muhammad Aslam, Using the kalman filter with Arima for the COVID-19 pandemic dataset of Pakistan, Data in Brief, Volume 31, 2020, 105854, https://doi.org/10.1016/j.dib.2020.105854.

[55] Lin, Feng & Muthuraman, Kumar & Lawley, Mark. (2010). An optimal control theory approach to non-pharmaceutical interventions. BMC infectious diseases. 10. 32. 10.1186/1471-2334-10-32.

[56] Osemwinyen, Amenaghawon & Diakhaby, Aboubakary. (2015). Mathematical Modelling of the Transmission Dynamics of Ebola Virus. Applied and Computational Mathematics (New York). 4. 313-320. 10.11648/j.acm.20150404.19.

[57] Chatterjee, S., Sarkar, A., Chatterjee, S. et al. Studying the progress of COVID-19 outbreak in India using SIRD model. Indian J Phys (2020). https://doi.org/10.1007/s12648-020-01766-8

[58] Lockerd Maragakis, Lisa, "First and second waves of Coronavirus", https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/first-and-second-waves-of-coronavirus

[59] https://www.newtral.es/estudio-de-seroprevalencia-un-5-de-espana-con-indicios-de-haber-pasado-covid-19/20200513/

Glossary

EPS	Electric Power System
PMU	Phasor Measurement Unit
DSE	Dynamic State Estimator
KF	Kalman Filter
UKF	Unscented Kalman Filter
EKF	Extended Kalman Filter
CKF	Cubature Kalman Filter
EnKF	Ensemble Kalman Filter
VSC	Voltage Source Converter
CPI	Customer-Phase Identification
PCA	Principal Component Analysis
LASSO	Least Absolute Shrinkage and Selection Operator
Covid	Coronavirus disease
SIR	Susceptible, Infected, Recovered
SIRD	Susceptible, Infected, Recovered, Deceased