CCFE
CULHAM CENTRE FOR FUSION ENERGY

UK Atomic Energy Authority

UKAEA-CCFE-CP(19)09

Wayne Arter, Ana Osojnik, Coralia Cartis, Godwin Madho, Steven Tobias

# Data assimilation approach to analysing systems of ordinary differential equations

This document is intended for publication in the open literature. It is made available on the understanding that it may not be further circulated and extracts or references may not be published prior to publication of the original when applicable, or without the consent of the UKAEA Publications Officer, Culham Science Centre, Building K1/0/83, Abingdon, Oxfordshire, OX14 3DB, UK.

Enquiries about copyright and reproduction should in the first instance be addressed to the UKAEA Publications Officer, Culham Science Centre, Building K1/0/83 Abingdon, Oxfordshire, OX14 3DB, UK. The United Kingdom Atomic Energy Authority is the copyright holder.

The contents of this document and all other UKAEA Preprints, Reports and Conference Papers are available to view online free at https://scientific-publications.ukaea.uk/

# Data assimilation approach to analysing systems of ordinary differential equations

Wayne Arter, Ana Osojnik, Coralia Cartis, Godwin Madho, Steven Tobias

This paper has been submitted to
IEEE International Symposium on Circuits and Systems (ISCAS), Florence, Italy, 27-30 May 2018

# Data assimilation approach to analysing systems of ordinary differential equations

Wayne Arter
CCFE, Culham Science Centre,
Abingdon, Oxon OX14 3DB, UK
Email: wayne.arter@ukaea.uk

Ana Osojnik
and Coralia Cartis
Mathematical Institute,
University of Oxford,
Oxford, OX2 6GG, UK

Godwin Madho
and Steven Tobias
Dept. of Applied Mathematics,
University of Leeds, Leeds, LS2 9JT, UK

***Abstract*—The problem of parameter fitting nonlinear oscillator models to noisy time series is addressed using a combination of Ensemble Kalman Filter and Optimisation techniques. Preliminary results for acceptable sampling rates and noise levels are presented. Application to the understanding and control of tokamak nuclear reactor operation is discussed.**

## I. Introduction

ITER, the world's device for exploring the physics and technology of magnetically confined nuclear fusion as a route to power generation, is currently under construction near Cadarache in S. France. For critical operational reasons, ITER needs to be able to model discharge evolution in real time as accurately as possible, the model in effect forming part of a closed loop control system. For example, even at current power levels in the similar but smaller EU funded device JET, limiting the temperature of the first wall is critical to the extent that magnetic-flux sweeping to spread the divertor heat load is becoming the norm. However, there is not at present a satisfactory model capable of modelling temperature distributions on plasma facing components (PFCs) which can operate at sufficient speed.

Even for operation of existing devices, a model will only be satisfactory if it has the capability to predict behaviour in new regions of operating space, hence to meet ITER needs it is critical for extrapolation that there be a physical basis to a model, which obviously has to incorporate time dependence. Under these conditions, the developing field of Uncertainty Quantification (UQ) indicates the favoured approach is data assimilation (DA)[1] using a multi-parameter surrogate model containing simplified physics. Amongst other benefits DA will help avoid problems of over-fitting noisy data, while being capable of incorporating both point sampled data from Langmuir probes (LP), 2-D video images of the PFCs, and line-of-sight data as obtained by bolometry, see [2]. The current work represents the first part of a programme designed both to develop suitable surrogate models and maximise the speed and robustness of DA fitting algorithms.

The surrogates chosen for initial investigation are in mathematical terms, nonlinear low order ordinary differential equations. As explained in Section II-A, not only might these systems reproduce the magnetic field behaviour of the device to sufficient accuracy, but they also represent the resonant response of a simple but nonlinear electrical circuit. Of course, as such they might be analysed by a variety of techniques already developed for such systems, eg. the matlab$^{TM}$ system identification tool-box, but since it is conceivable that partial differential equation models could be used in the course of ITER operation in future decades, DA is preferred as more amenable to extension to very computationally expensive models.

Following description of the model and representative data quality in Section II-A, the Ensemble Kalman Filter (EnKF) approach to DA is outlined in Section III-A. Difficulties caused by sampling effects and noise lead to the need for a complementary optimisation step to help identify likely parameters. This is discussed in Section IV, with preliminary results in Section V and finally Section VI is brief discussion of conclusions and possible future developments.

## II. Model and Data

### A. Models

In the plasma, there is a range of microscopic and macroscopic instabilities, such as sawtooth and ELMs, which served to motivate this work. The sawtooth instability is a relaxation oscillation in the centre of the plasma at large electric currents, mainly observed through oscillations in electron temperature and density, followed by subsequent movement of particles and energy as a heat pulse from the centre of the plasma to the boundary. Edge-localised modes occur during sufficient increase of input power, when the edge of the plasma, characterised by large differences in electron density and temperature, undergoes short heat and particle eruptions. In addition to instabilities being prone to nonlinear interactions, a wide range of spatial and time scales also make simulations of large scale behaviour of the tokamak plasma at high temperatures difficult and computationally demanding. However, simplifications of tokamak geometries under symmetry considerations enables the study of sawteeth and ELMs via simple ordinary differential equation (ODE) models that reproduce their behaviour as outlined [3]. In particular, the source states the equations follow on from equivariant bifurcation theory, where low order Taylor expansions of equations are used in regions of parameter space where qualitative behavioural change is shown. The simplest coupled equations, called Axisymmetric

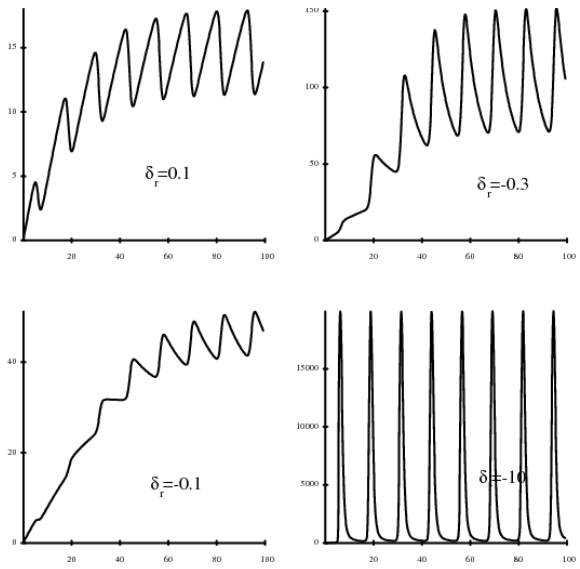


Fig. 1. Plots of $b(t)$ for ANAC model as parameter $\delta_r$ is varied, showing potential to represent both sawteeth and spiking oscillations.

and Non-Axisymmetric Coupled model (ANAC), observed to qualitatively fit the experimental data, are the following

$$\ddot{a} = \gamma a + 2\mu a^3 \quad (1)$$
$$\dot{b} = \alpha - \beta b^2 - (1+\delta_r b)a^2 \quad (2)$$

where dot notation represents derivatives with respect to time t. The final goal is to perform data assimilation with this model. However, due to complexity of the coupled system, simpler models are first looked at for purposes of testing the data assimilation algorithm and the optimisation scheme, as well as to gradually build up complexity and understanding of the problem. Models considered along with brief outlines of their dynamical properties are presented below.

The simplified non-coupled version of the system is given by Equation (1), which represents a non-linear oscillator provided $\gamma < 0$ when solutions correspond to motion in a potential well of the form $4V(a) = -2\gamma a^2 - \mu a^4$. The system state vector is defined as $\mathbf{x} = (a, \dot{a})$ and the model parameters as $\lambda = (\gamma, \mu)$.

The Lorenz system is a set of three ordinary differential equations which were first derived by E. N. Lorenz[4]. The system Equation (3) is known for its non-periodic/chaotic behaviour. In all our studies $\sigma = 10$, $r = 28$ and $b = 8/3$.

$$\begin{aligned} \dot{x} &= \sigma(y - x), \\ \dot{y} &= rx - y - xz, \\ \dot{z} &= xy - bz \end{aligned} \quad (3)$$

### B. Data

All testing described herein is done using synthetic data. The aim is to make those as realistic as possible in order for the results to be meaningful for real experimental data of plasma magnetic field measurements, which are taken during occurence of ELMs or sawteeth in the tokamak. The assumptions made are the following:

- The sampling rate of the data (denoted as $\nu$), the number of observations per period of oscillation, is at least 6. We are able to determine the period oscillation, i.e. the typical time scale, via the Fast Fourier Transform.
- Noise is present in data. One source of noise are errors due to truncation of recorded data points, in addition to other sources of noise difficult to identify.
- Measurements are taken directly of system variables, therefore, the linear observation model is simply the identity matrix, $h = I$, the unit matrix.

Synthetic data are generated by perturbing the solutions obtained by numerical integration at each timestep, assumed fixed. The perturbations are simply numbers from a uniform or Gaussian distribution with parameter $\sigma$. It is further assumed that the timestep of observations is a positive integer multiple of the numerical integration timestep.

## III. Data Assimilation

Many simulations are initial-value problems where the determination of the initial condition is very important. Observations can be used to obtain the initial condition but are usually non-uniformly spread in space and time. Hence an estimate has to be provided for the initial condition on all grid points. After running the simulation for a few time steps the observation and the result of the model are combined to obtain an analysis state. This state is then used as initial condition for the next set of forecast. The method of combining observations and the numerical model is called data assimilation (DA). The analysis cycle described above is run at intermittent time intervals to obtain new initial conditions for our forecast. A general outline of the process is shown in Figure 2. The calling of the analysis cycle is defined by the user or can be initiated by an increase in the model error. Generally there is a forecast, $x^f$, an observation, $y^o$ and a first guess, $H(x^f)$, where $H$ is the observation map operator. Given the innovation or observational increment being the difference between observed and model first guess, $y^o H(x^b)$, the analysis can be obtained by adding this innovation to the model forecast with weights $W$ which are determined by statistical error covariance.

$$x^a = x^b + W\left(y^o - H(x^b)\right) \quad (4)$$

Many assimilation schemes are based on Equation (4) and differ only by how they combine the observation and background to produce the analysis. A specific scheme, namely EnKF, is now described.

### A. EnKF

A very simple derivation for the EnKF approach to Data Assimilation, suitable for a first introduction begins with Bayes' theorem in probability, which states that

$$P(x|d) \propto P(x)P(d|x) \quad (5)$$

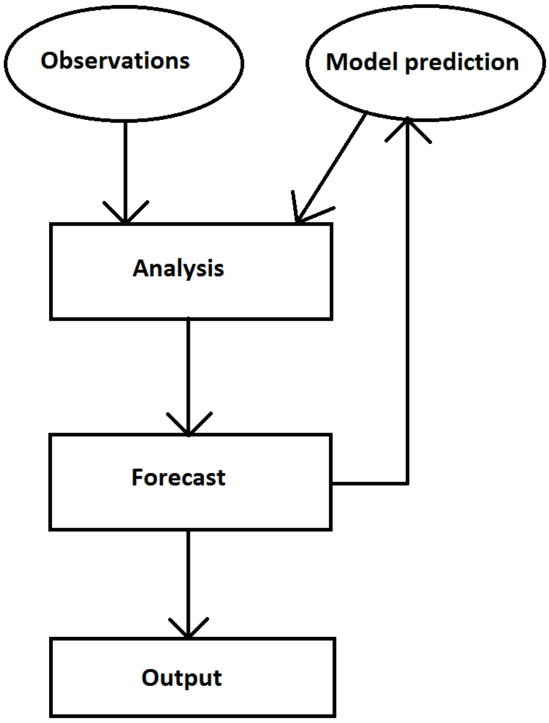


Fig. 2. General Schematic for Data Assimilation.

when Gaussian distributions for $P(x)$, $P(d|x)$ imply

$$P(x|d) \propto \exp\{-\frac{1}{2}T(x)\} \quad \text{where} \tag{6}$$

$$T(x) = (x-x^f)(C_{ff})^{-1}(x-x^f) + \tag{7}$$

$$+ \ (x-d)(C_{\epsilon\epsilon})^{-1}(x-d) \tag{8}$$

Minimising $T$ gives the analysis $x = x^a$,

$$\delta T(x) = 0 = 2(x-x^f)(C_{ff})^{-1} + 2(x-d)(C_{\epsilon\epsilon})^{-1} \tag{9}$$

as

$$x^a = \frac{x^f(C_{ff})^{-1} + d(C_{\epsilon\epsilon})^{-1}}{(C_{ff})^{-1} + (C_{\epsilon\epsilon})^{-1}} \tag{10}$$

$$= \frac{x^f C_{\epsilon\epsilon} + dC_{ff}}{C_{\epsilon\epsilon} + C_{ff}} \tag{11}$$

$$\overline{(p^a)^2} = C_{ff}\left(1 - \frac{C_{ff}}{C_{ff}+C_{\epsilon\epsilon}}\right) = C_{aa} \leq C_{ff} \tag{12}$$

The above generalises to give the vector formulation of EnKF

$$\mathbf{x}^a - \mathbf{x}^f = C_{ff}(C_{ff}+C_{\epsilon\epsilon})^{-1}(\mathbf{d} + \delta\mathbf{d} - H\mathbf{x}^f) \tag{13}$$

where $H\mathbf{x}^f$ is what is observed, and $\delta\mathbf{d}$ makes explicit the addition of a perturbation to the observed data.

Combined state and parameter estimation with EnKF

For many dynamical real-world problems a common problem are poorly known model parameters. Evensen [5] introduces a scheme for estimating poorly known model parameters in addition to system state during data assimilation with an Ensemble Kalman Filter. The main difference with the classic system state estimation algorithm is that poorly known model parameters $\lambda$ are added to the state vector $\mathbf{x}$. Therefore, at every time step the new state vector is $\mathbf{z}_k = (\mathbf{x}_k, \lambda_k x)$ the ability to change in time and should eventually converge to the true values, i.e. $lim_{k\to\infty}\lambda_k = \lambda^t$ , under the assumption that model parameters are constants. This was implemented by revising the dynamical model to account for this by adding $d(\lambda_i)dt = 0$ for $i = 1, \ldots n_\lambda$ to the original system.

The reasons for choosing EnKF may be enumerated as

1) Extensibility - from 1 to $10^8$ variables.
2) Simultaneous parameter and state estimation.
3) Wide range of practical applications - oil reservoir modelling, weather forecasting, image processing, MHD.
4) Convergence studies
   a) Successful empirical parameter fitting for Lorenz chaos [6].
   b) Recent positive analytic results [1]

### B. EnKF from matlab to python

EnKF-matlab is described in [7], and may be freely downloaded from the website http://enkf.nersc.no/, This version is used as a stepping stone to understand how to program EnKF and its working. In the matlab code there is a file which holds 10 000 sample points each with its own x,y and z values. At the start of the simulation this sample file is shuffled, the first point now is taken to be the initial condition for the true state and the points following in the sample are taken to be the initial conditions for the ensemble state. The size of the ensemble is chosen by the user and can be changed through the parameter file. Both the true state and the ensemble are propagated in series using the same time step which is specified in a custom file. The model steps before assimilation is specified by the user in the parameter file. Depending on the settings for variables such as inflation and localization in the parameter file different assimilation algorithms are followed. The default version of the matlab code uses[7] to work out the corrections, but for comparison purposes the version of EnKF described above was used. with localization and inflation turned off.

The matlab code for this algorithm was straightforwardly converted to Python. As seems to be standard [6], the Lorenz model was used to test the new code. A typical verification test is shown in Figure 3.

## IV. Optimisation and CMA-ES

Three optimisation algorithms have been tested, two local solvers for nonlinear least squares problems, DFOGN [8] and DFO-LS [9] and a stochastic search method called Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) [10]. The variety of types of solvers was chosen to see variations in performance, and their appropriateness for this problem.

Local solvers perform the search iteratively by zooming in on error function values in the close neighbourhood of the current best value held; the new best candidates are taken deterministically to be the best neighbouring points. Solvers DFOGN and DFO-LS in particular choose the next

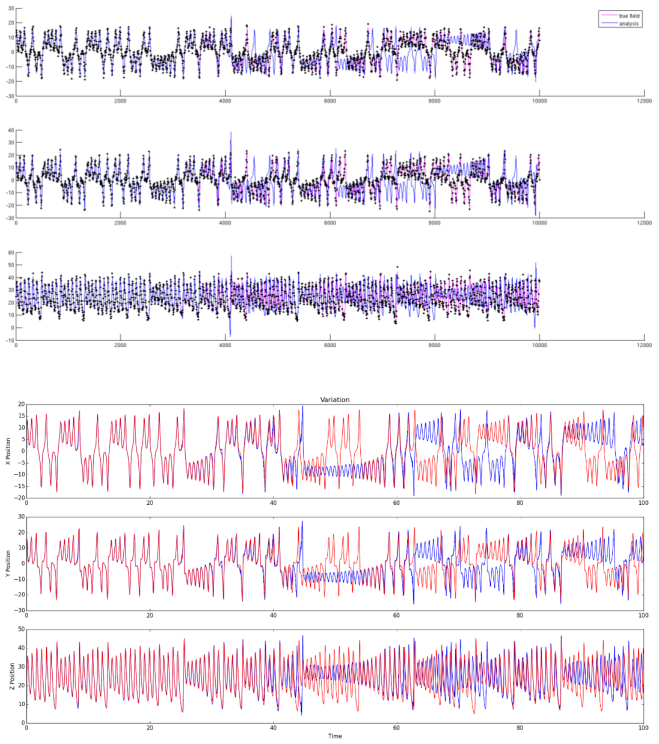


Fig. 3. Results for Lorenz system, verifying EnKF-python (bottom) against EnKF-matlab (top). The blue line is the ensemble average and the red line is the true state.

iterate based on the approximation of the error function in the neighbouring region via a linear model, as this is cheaper and easily applicable to black-box or noisy functions.

On the other hand, stochastic search implies that there is an element of randomness in searching for new best candidates that minimise the error function. CMA-ES specifically searches the neighbourhood of some starting guess by moving and reshaping a normal distribution based on sorting between samples from the distribution.

All solvers were initially run for 100 different starting guesses of parameter and initial state values, over which optimisation was attempted. In reality it might occur that observed data can provide a very good guess of the initial system state. To imitate and test such situations, the true initial state was perturbed with small noise level of 0.05. Optimisation was for such cases performed over parameters only with the same 100 starting candidates for parameters. Robustness of solvers for these problems to noisy or sparse data sets was tested by varying noise levels and sampling rates of synthetic observations.

## V. Results

All solvers were initially run for 100 different starting guesses of parameter and initial state values, over which optimisation was attempted. In reality it might occur that observed data can provide a very good guess of the initial system state. To imitate and test such situations, the true

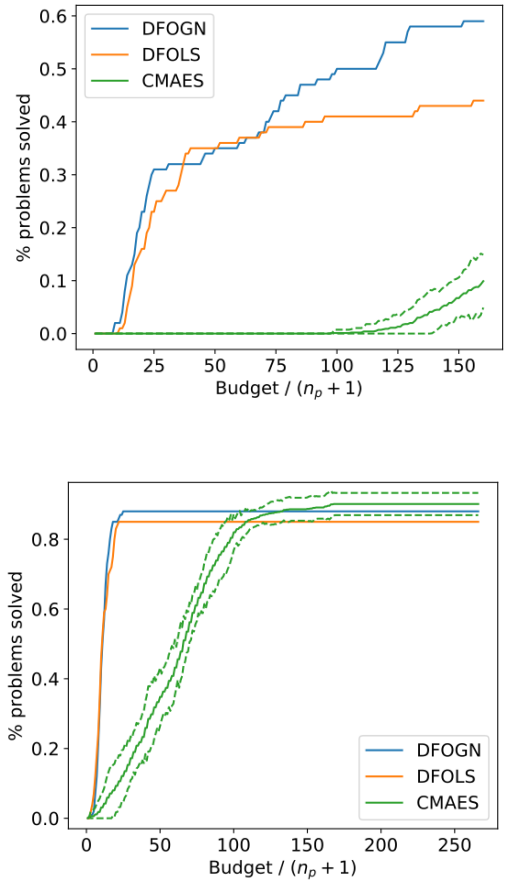


Fig. 4. Performance profiles for combined parameter and initial state estimation (top) and estimation of just parameters (bottom) with $\nu = 12$ and Gaussian noise $\sigma = 0.5$.

initial state was perturbed with small noise level of 0.05. Optimisation was for such cases performed over parameters only with the same 100 starting candidates for parameters. Robustness of solvers for these problems to noisy or sparse data sets was tested by varying noise levels and sampling rates of synthetic observations.

Comparison of derivative-free solvers DFOGN, DFO-LS and CMA-ES showed that for combined parameter and initial state estimation, with and without noise, DFOGN was the best-performing solver overall. The increase in the number of solved problems to high accuracy was the steepest and the final proportion reached was the highest at about 60%. It was followed closely by DFO-LS, whereas CMA-ES consistently performed considerably worse on average (see Figure 4 (top)). Stagnation of local solvers at local function minima was observed. Local minimum of a function is a point which, compared to points in its close vicinity, gives the smallest value of the function. Increased sampling rate seemed to have decreased the chances of local solvers getting stuck at these points.

Sole parameter estimation, as in Figure 4 (bottom), showed

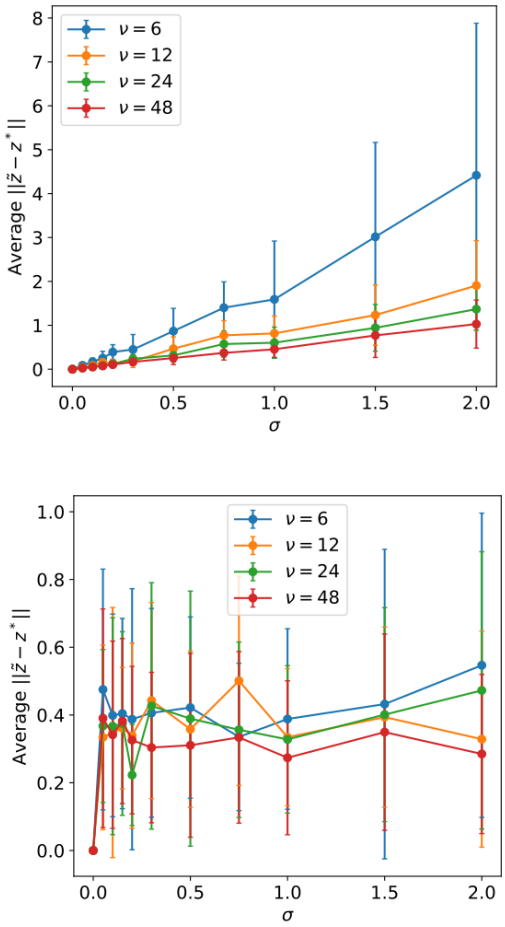


Fig. 5. Plots of average error of estimates for combined parameters and initial state case (top) and just parameters (bottom). $\nu$ is the sampling rate and $\sigma$ is the Gaussian noise level.

very encouraging results for all solvers, as within 500 function evaluations the solvers found the minimum for about 90% of starting guesses. This was the case for the error function measuring error from observations with as well as without noise. DFOGN and DFO-LS were consistently the fastest solvers over CMA-ES here as well.

When varying noise levels and sampling rates (denoted as $\nu$) with DFOGN solver, it was found that in general larger sampling rates seem to mitigate large noise levels. There is a clear trade-off visible between the two when estimating parameters and the initial state in Figure 5 (top). With just parameters, this trend is less strong, and one can notice in Figure 5 (bottom) that the average error over the runs for different starting guesses stays constant. This is an encouraging result pointing to errors in parameter estimation occurring mostly due to small noise in the known initial state.

The new approach developed for initialisation of the ensemble Kalman filter involves using observations over the first period for parameter and initial state estimation. The optimised values of parameters are used to initialise the model

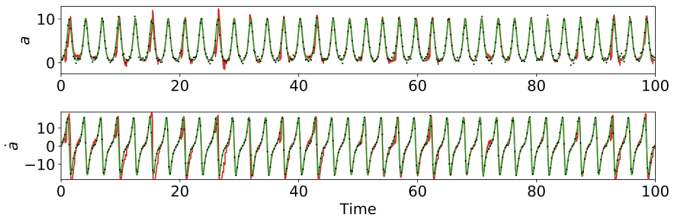


Fig. 6. Ensemble average (red) with observations (black), reference solution (green) and standard deviation (blue) for sampling rate 12 and Gaussian noise level $\sigma = 0.5$

to be assimilated, whereas optimised initial state is used for initialisation of the initial ensemble of state vectors. Results for one run including error estimates can be found in Figure 6, where performance plots and average root mean square errors are similar to the case of initialisation with true values, as desired.

## VI. Conclusion

Software has been developed for parameter identification of noisy time-dependent systems. Its robustness has been confirmed using synthetic data representative of simple nonlinear oscillators with additive noise. It is expected to be most useful in situations where it is desired to produce surrogates of varying degrees of sophistication, the more detailed ones including spatial as well as temporal variation. Even the simple surrogates may however be used to validate physical assumptions about tokamak dynamics, as well as employment in system control.

## Acknowledgment

This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement number 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission To obtain further information on the data and models underlying this paper please contact PublicationsManager@ccfe.ac.uk. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

## References

[1] K. Law, A. Stuart, and K. Zygalakis, *Data assimilation: a mathematical introduction*. Springer, 2015, vol. 62.

[2] T. Craciunescu, M. Murari, E. Peluso, M. Gelfusa, and J. Contributors", "A New Approach to Bolometric Tomography in Tokamaks," in *This proceedings*, 2018.

[3] A. Murari, W. Arter, D. Mazon, M. Gelfusa, V. Ferat, and J.-E. Contributors, "Symmetry based analysis of macroscopic instabilities in Tokamak plasmas," JET-EFDA, Tech. Rep. PR(11)07, 2011, http://www.iop.org/Jet/fulltext/EFDP11007.pdf.

[4] E. Lorenz, "Deterministic Nonperiodic Flow," *Journal of the Atmospheric Sciences*, vol. 20, no. 2, pp. 130–141, 1963.

[5] G. Evensen, *Data assimilation: the ensemble Kalman filter*. Springer, 2009.

[6] A. Majda and J. Harlim, *Filtering complex turbulent systems*. Cambridge University Press, 2012.
[7] P. Sakov, G. Evensen, and L. Bertino, "Asynchronous data assimilation with the EnKF," *Tellus A*, vol. 62, no. 1, pp. 24–29, 2010.
[8] L. Roberts and C. Cartis, "A Derivative-Free Gauss-Newton Method," *To appear*, 2017.
[9] L. Roberts, C. Cartis, J. Fiala, and B. Marteau, "A Derivative-Free Solver for Least-Squares Problems," *To appear*, 2017.
[10] Hansen, "CMA-ES, Covariance Matrix Adaptation Evolution Strategy for non-linear numerical optimization in Python. Version 1.1.7. 2008-2014." https://pypi.python.org/pypi/cma, Tech. Rep., Visited on 07/07/2017.