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,
Chris Jones
and Steven Tobias
Dept. of Applied Mathematics,
University of Leeds, Leeds, LS2 9JT, UK

Abstract—The problem of parameter fitting for nonlinear oscillator models to noisy time series is addressed using a combination of Ensemble Kalman Filter and optimisation techniques. Encouraging 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 the 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 indicates that 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 overfitting noisy data, while being capable of incorporating both point sampled data from Langmuir probes, 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 Toolbox. 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. MODELS AND DATA

A. Models

In the tokamak plasma, there is a range of microscopic and macroscopic instabilities, such as sawtooth and edge-localised modes (ELMs), which serve 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 in [3]. There, the equations follow on from equivariant bifurcation theory, where low-order Taylor expansions of equations are used in regions of the parameter space where qualitative behavioural change is shown. The simplest coupled equations,
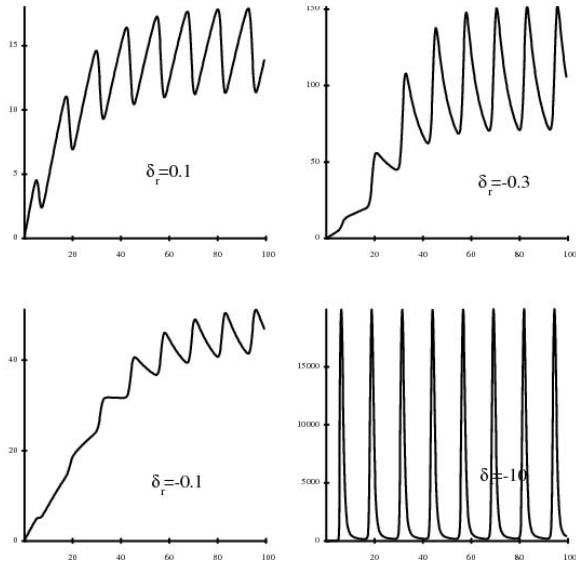
Fig. 1. Plots of $b(t)$ for ANAC model as parameter δ_r is varied, showing potential to represent both sawteeth and spiking oscillations.

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

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

where the 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 considered for purposes of testing the data assimilation algorithm and the parameter estimation optimisation scheme, as well as to gradually build up complexity and understanding of the problem. The models we consider are:

- The simplified noncoupled version of the system is given by Equation (1), which represents a nonlinear oscillator provided $\gamma < 0$. Solutions in this case correspond to motion in a potential well of the form $4V(a) = -2\gamma a^2 - \mu a^4$. We define the system state vector as $\mathbf{x} = (a, \dot{a})$ and the vector of model parameters as $\boldsymbol{\lambda} = (\gamma, \mu)$. In all our studies $\boldsymbol{\lambda} = (10, -0.1)$.
- The Lorenz system [4] is

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

The system (3) is known for its non-periodic/chaotic behaviour. In all our studies $\sigma = 10$, $r = 28$ and $b = 8/3$.

B. Data

All testing described herein is done using synthetic data. The aim is to make the latter 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 occurrence of ELMs or sawteeth in the tokamak. We make the following assumptions:

- The data sampling rate, ν, or the number of observations per period of oscillation, is at least 6. From the data we are able to determine the period oscillation (i.e. the typical time scale) via the Fast Fourier Transform.
- Data contains noise from various sources; these can be errors due to truncation of recorded data points or from other sources which are difficult to identify.
- We measure system variables directly, so the observation model is assumed to be the identity matrix, $H = I$ (see Equation (4) below).

Synthetic data is generated by perturbing the solutions obtained by numerical integration at each time step, which is assumed to be fixed. The perturbations are randomly generated from the uniform or Gaussian distribution with variance σ^2. Observations are subsampled from synthetic data so that the observation time step is a positive integer multiple of the numerical integration time step.

III. DATA ASSIMILATION

Model simulations are often initial value problems where determination of the initial condition is very important. Observations can be used to obtain the initial condition but are usually non-uniformly distributed in space and time. Hence an estimate has to be provided for the initial condition on all grid points. After running the simulation over a few time steps, the observation and the model forecast are combined to obtain an analysis state. This state is then used as initial condition for the next forecast. The method of combining observations and the 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. 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 forecasted observation, $H(x^{\mathrm{f}})$, where H is the observation map operator. Given the innovation or observational increment, which is the difference between the observation and its model forecast, namely, $y^{\mathrm{o}} - H(x^{\mathrm{f}})$, the analysis can be obtained by adding the innovation to the model forecast using an optimal weighting W, which depends on statistical forecast and observation error covariance; that is

$$x^{\mathrm{a}} = x^{\mathrm{f}} + W\left(y^{\mathrm{o}} - H(x^{\mathrm{f}})\right). \qquad (4)$$

Many assimilation schemes are based on Equation (4) and differ only by how they combine the observation and the forecast to produce the analysis. A common statistical DA method, called Ensemble Kalman Filter (EnKF), is now described.

A. EnKF

Derivation of a Kalman Filter approach to DA relies on the Bayes' theorem in probability, which states that

$$P(x|y^{\mathrm{o}}) \propto P(x)P(y^{\mathrm{o}}|x). \qquad (5)$$

In EnKF, $P(x)$ and $P(y^{\mathrm{o}}|x)$ are assumed Gaussian. For a 1-D example with the identity observation map $H = I$, this implies

$$P(x|y^{\mathrm{o}}) \quad \propto \quad \exp\{-\tfrac{1}{2}T(x)\}, \quad \text{where} \tag{6}$$

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

$$(x - y^{\mathrm{o}})(C_{\mathrm{ee}})^{-1}(x - y^{\mathrm{o}}), \tag{8}$$

where C_{ff} and C_{ee} are the forecast and observation error covariances, respectively. Minimising $T(x)$ gives the analysis $x = x^{\mathrm{a}}$, since when

$$\frac{\mathrm{d}}{\mathrm{d}x}T(x) = 2(x - x^{\mathrm{f}})(C_{\mathrm{ff}})^{-1} + 2(x - y^{\mathrm{o}})(C_{\mathrm{ee}})^{-1} = 0 \tag{9}$$

we have that

$$x = \frac{x^{\mathrm{f}}(C_{\mathrm{ff}})^{-1} + y^{\mathrm{o}}(C_{\mathrm{ee}})^{-1}}{(C_{\mathrm{ff}})^{-1} + (C_{\mathrm{ee}})^{-1}} = \frac{x^{\mathrm{f}}C_{\mathrm{ee}} + y^{\mathrm{o}}C_{\mathrm{ff}}}{C_{\mathrm{ee}} + C_{\mathrm{ff}}} =: x^{\mathrm{a}}. \tag{10}$$

and there also follow error estimates for x^{f}. In a general vector formulation of EnKF with a linear observation operator, the above generalises to

$$\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{f}} = C_{\mathrm{ff}}H^{T}(HC_{\mathrm{ff}}H^{T} + C_{\mathrm{ee}})^{-1}(\mathbf{y}^{\mathrm{o}} + \delta\mathbf{y} - H\mathbf{x}^{\mathrm{f}}), \tag{11}$$

where $\delta\mathbf{y}$ is a Gaussian perturbation of the observation, and covariance matrices are computed as Monte Carlo estimates from the ensemble runs.

The reasons for choosing EnKF may be enumerated as

1) Allowing nonlinear model and observation operators.
2) Extensibility - from 1 to 10^{8} variables.
3) Simultaneous parameter and state estimation (see below).
4) Wide range of practical applications - oil reservoir modelling, weather forecasting, image processing, MHD.
5) Convergence studies: successful empirical parameter fitting for Lorenz chaos [5]; recent analytical convergence results [1].

B. Combined state and parameter estimation with EnKF

For many dynamical real-world problems a common problem are poorly known model parameters. Evensen [6] introduced a scheme for estimating poorly known model parameters in addition to system state during DA with the EnKF. The main difference with the classic system state estimation algorithm is that poorly known model parameters $\boldsymbol{\lambda}$ are added to the state vector $\mathbf{x}$. Therefore, at every time step the new state vector is $\mathbf{z} = (\mathbf{x}, \boldsymbol{\lambda})$, that changes in time. $\boldsymbol{\lambda}$ should eventually converge to the true values $\boldsymbol{\lambda}^{\mathrm{t}}$, under the assumption that model parameters are constants. To account for this, we revised the dynamical model by adding $\mathrm{d}\boldsymbol{\lambda}/\mathrm{d}t = 0$ to the original system of ODEs.

C. EnKF implementation: from Matlab to Python

EnKF-Matlab is described in [7], and may be freely downloaded from the website http://enkf.nersc.no. This version was used as a stepping stone, to understand how to program EnKF
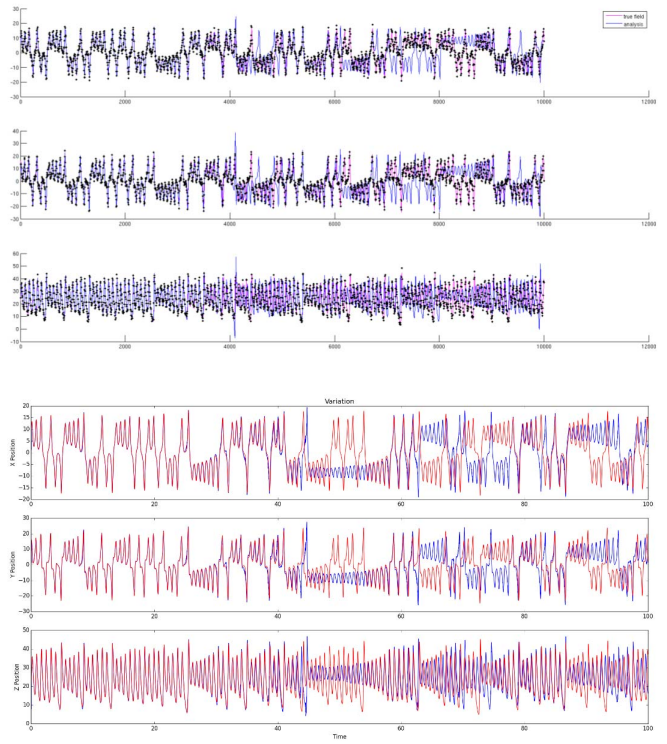

Fig. 2. Results for Lorenz system, verifying EnKF-python (bottom) against EnKF-matlab (top). The blue line is the ensemble average and the red line is the true state.

and its working [1]. The MATLAB code was straightforwardly converted to Python. As seems to be standard [5], the Lorenz model (3) was used to test the new code. A typical verification test is shown in Figure 2. The capability of EnKF-Python was then extended to have joint parameter and state estimation. We found that the performance of the latter was highly sensitive to initial guesses of parameter values. This further motivated us to consider automatic ways of tuning initial guesses of parameters and state for improved EnKF performance.

IV. OPTIMISATION ALGORITHMS AND TESTS

In addition to poorly known model parameters, noise in the initial state is also often encountered. For successful initialisation of EnKF, it would be beneficial if parameters were tuned beforehand, and a good guess of the initial state was available. We set up an optimisation scheme for estimation of the model parameters and the initial state using observations over the time window of the first period of oscillation, to be

[1]The MATLAB code contains a file with $10\,000$ sample points (x, y, z). At the start of the simulation, a random point from this set is taken to be the initial condition for the true state, whereas the leftover points are taken to be initial conditions for ensemble runs. The size of the ensemble is chosen by the user in the parameter file. Both the true state and ensemble states are propagated in series using the same time step, specified in the parameter file along with the model step. Settings such as inflation and localization in the parameter file determine the assimilation algorithm to be followed. The default version of the MATLAB code uses [7] to work out analysis corrections. For comparison purposes, we use this version, with localization and inflation turned off.

used to aid initialisation of EnKF. We define the optimisation problem of minimising a function quantifying the error between obervations and model predictions as

$$\underset{\mathbf{x}_0, \boldsymbol{\lambda}}{\text{minimise}} \quad f(\mathbf{x}_0, \boldsymbol{\lambda}) = \sum_{k=1}^{N} \|\mathbf{y}_k - H(\mathbf{x}_k^{\text{f}})\|^2, \qquad (12)$$

where dependency of f on $\mathbf{x}_0$ and $\boldsymbol{\lambda}$ is through the model forecast $\mathbf{x}_k^{\text{f}}$, and where N is the number of observations available.

Two optimisation algorithms are tested, a local, derivative-free, solver for nonlinear least squares problems such as (12), named DFO-GN [8], and a stochastic search method called Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) [9]. These two types of solvers have state of the art performance in their respective algorithm classes, and are chosen to see variations in performance, and their appropriateness for parameter and initial state estimation in nonlinear models such as ANAC[2].

Local solvers perform the search for a (local) minimizer point iteratively, by zooming in on error function values in a neighbourhood of the current best point held; new best-point candidates are taken deterministically to be the best neighbouring points within a trust region. Model-based derivative-free solver DFO-GN, in particular, chooses the next iterate based on the approximation of the error function in the neighbouring region via a linear model, which is relatively inexpensive and easily applicable to black-box or noisy functions.

On the other hand, stochastic search methods use an element of randomness in searching for new best candidates that minimise the error function. CMA-ES searches the neighbourhood of the current best minimum point by sorting between samples from a normal distribution centred at this point. At the next iteration, the distribution is moved and appropriately reshaped in the direction of best samples.

V. RESULTS

To define our problem set, we pick 100 different starting guesses of parameter and initial state values[3], and assess performance of solvers on this set. A key measure of the effectiveness of an optimisation algorithm is how the proportion of problems solved (to global optimality, which is the true state value) varies with the number of function evaluations (budget) scaled with the dimension of the problem.

In real problems, the observed data may provide a very good guess of the initial system state. To imitate and test such situations, we perturb the true initial state with a small error of up to 0.05, and fix it at this value. Optimisation is then carried out with the same set of 100 starting guesses for parameters. Robustness of solvers for these problems to noisy or sparse data sets is tested by varying the noise levels and sampling rates of the synthetic observations.

---

[2]DFO-GN scales well with increasing number of state variables/parameters, and can successfully solve, from (noisy) error evaluations alone, data fitting problems of $O(100)$ dimensions [8].

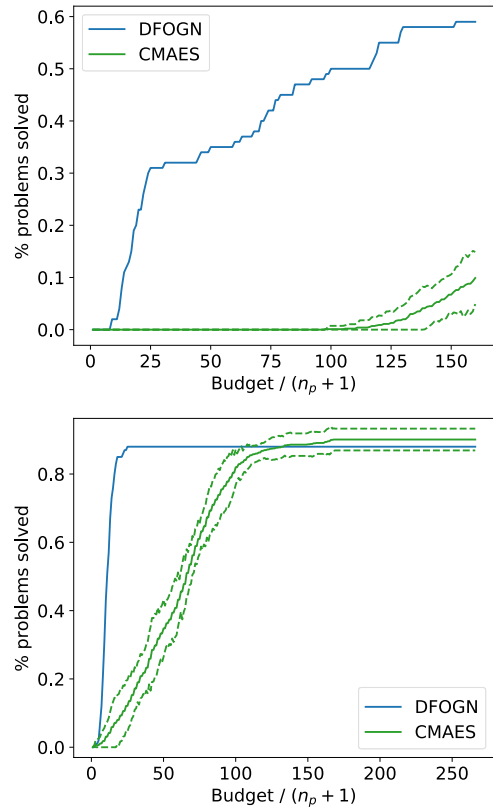[3]These points are chosen on a latin hypercube in intervals around true values.

Fig. 3. Data performance profiles for combined parameter and initial state estimation (top) and estimation of just parameters (bottom) using synthetic data with observation sampling rate $\nu = 12$ and Gaussian noise $\sigma = 0.5$.

Comparison of data performance profiles for derivative-free solvers DFO-GN and CMA-ES shows that for combined parameter and initial state estimation with and without noise, DFO-GN is the best-performing solver overall. This is since the increase in the number of problems solved to accuracy of 10^{-5} is steeper than for CMA-ES, and the final proportion reached is greater, at about 60%; see Figure 3 (top). CMA-ES consistently performs considerably worse on average (see Figure 3 (top)), with smaller number of problems solved to high accuracy and slower progress. Stagnation of the local solver at local function minima, frequently encountered with this type of problems and solvers, is observed. Our results indicate that increased sampling rates decrease the number of distinct local (but not global) points this solver gets stuck at. The profiles in Figure 3 do not seem to have reached a plateau yet, therefore, a valid speculation is that an increase in solvers' performance could be achieved with larger budgets.

For parameter estimation (only), we see in Figure 3 (bottom) encouraging results using both solvers, which find the (global) minimum for about 90% of starting guesses within a reasonable budget. DFO-GN again consistently makes faster progress than CMA-ES, using observations with noise and without, although we also observe that CMA-ES can solve comparable proportions of problems, given sufficiently large budgets.

By varying noise levels and sampling rates with DFO-GN, we find that in general larger sampling rates seem to mitigate large noise levels. A clear trade-off is visible between the two for parameter and initial state estimation, as in Figure 4 (top). When estimating only parameters, this trend is less strong, and we see in Figure 4 (bottom) that the average error over the runs for different starting guesses remains constant. This is promising as it suggests that the main source of error in parameter estimation is the small noise in the known initial state.

The new approach developed for initialisation of EnKF involves using observations over the first period for parameter and initial state estimation. The optimised values of parameters are used to initialise the model to be assimilated, whereas the optimised initial state is used for initialisation of the initial ensemble of state vectors. Results for one run can be found in Figure 5, where performance is 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
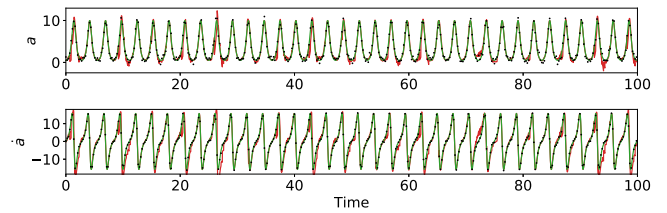
Fig. 5. Ensemble average (red) with observations (black) and the reference solution (green) for an EnKF run with observation sampling rate $\nu = 12$ and Gaussian noise level $\sigma = 0.5$. EnKF is initialised with parameters and the initial state found with DFO-GN. The resulting average RMSE for $(a, \dot{a})$ is $(0.811, 2.949)$, which is comparable to $(1.15, 2.99)$, the average RMSE obtained using true initialization values for state and parameters.

assumptions about tokamak dynamics, as well as employment in system control.

Complete details of the results presented here can be found in [10]; available upon request. The DFO-GN solver (Python) is available on GitHub[4], and is released under the open-source GNU General Public License.

ACKNOWLEDGMENT

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] A. Majda and J. Harlim, *Filtering complex turbulent systems*. Cambridge University Press, 2012.
[6] G. Evensen, *Data assimilation: the ensemble Kalman filter*. Springer, 2009.
[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] C. Cartis and L. Roberts, "A Derivative-Free Gauss-Newton Method," *Technical Report, University of Oxford, arXiv:1710.11005*, 2017.
[9] N. 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.
[10] A. Osojnik, "Techniques for initialising simple data assimilation calculations for plasma models," Technical Report, InFoMM CDT, University of Oxford, Tech. Rep., 2017.
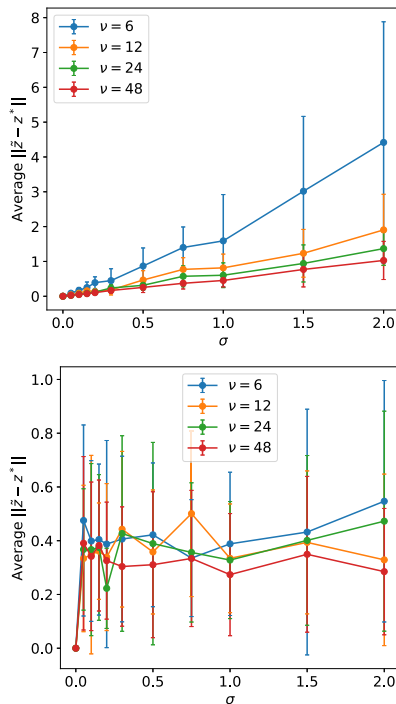
Fig. 4. Plots of average error of estimates for combined parameters and initial state case (top) and just parameters (bottom). ν is the sampling rate and σ is the Gaussian noise level.