University of Reading

PhD Project Proposal 1

Title: *Inference of nonlinear dynamics through data assimilation and machine learning*

Start Date: October 2022.

Supervisors: J. Bröcker (Maths/Stats), A. Carrassi (Met), V. Ojha (CS), T. Kuna (L'Aquila).

The project: The problem of inferring dynamical laws from noisy and incomplete observations has long been recognised as very important but notoriously difficult. The "classical" Bayesian or maximum likelihood approaches to this problem assume a parametrised form for the to-be-inferred dynamical laws and then estimate the distribution of those parameters, conditionally on the history of observations. These approaches however involve solving the nonlinear filtering problem, which is known to have an infinite complexity (except for a few special cases such as linear dynamics or finite state Markov chains), and this difficulty is inherited to the problem of inferring dynamical laws.

A large amount of research has been dedicated to finding approximate and efficient methodologies that work in realistic scenarios. Data assimilation (DA) is the most notable example, providing various approaches to approximate the nonlinear filter. These approaches in turn allow to apply (approximately) Bayesian or maximum likelihood estimators, and such methods have proved successful (at least empirically) in estimating the parameters of the assumed underlying dynamical laws and recently in fully retrieving those laws (given a prescribed model class) from data [1].

At the same time, machine learning (ML) is currently underpinning pivotal changes to inference problems in applications. Leveraging novel and efficient big data algorithms, ML is showing immense capabilities in retrieving dynamical laws from data, without the need to explicitly assume a parametric model class.

Combining ML and DA has recently attracted the attention of researchers, and novel powerful DA-ML hybrid approaches have been developed under an approximate Bayesian perspective [1]. While showing enormous practical potential, DA-ML methods have not yet been subject of a throughout mathematical examination, particularly with regards to convergence and consistency properties, i.e., if and in what sense they approximate the true dynamical laws as data and sample size are increased. Since this understanding is pivotal in guiding the design and improvement of methods, our aim in this project is to advance our theoretical understanding of these novel DA-ML methods.

To achieve this, we shall leverage a recent landmark theoretical result by Douc et al [4], proving consistency of the maximum likelihood estimator for a large class of hidden Markov dynamics. Nevertheless, challenges remain in extending or modifying these results to the DA-ML estimation methodology outlined above.

The project is structured along the following successive tasks:

A. Based on the analysis in [4], DA-ML algorithms (e.g., as in [2,3]) will be analysed theoretically for consistency. This will require extending the methodology of [4] as we will deal *merely* with *approximations to* the maximum likelihood estimator.

B. The analysis from (A) will inform the development of improved numerical algorithms.

C. The results from (A,B) will be studied numerically using finite dimensional chaotic dissipative ODEs (Lorenz'96) but also PDEs (e.g., 2D—Navier—Stokes).

D. Based on progress and student inclination, the analysis will be extended to investigate asymptotic distributions of parameter estimates.

Student profile: Bachelor 2:1 or Master in Mathematics, Statistics or Computer Science.

References:

1. Bocquet, M., Brajard, J., Carrassi, A. and Bertino, L. (2020) Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximisation. *Foundations of Data Science*, 2 (1)

2. Brajard, J., Carrassi, A. , Bocquet, M. and Bertino, L. (2021) Combining data assimilation and machine learning to infer unresolved scale parametrisation. Philos. Trans. R. Soc. London, Ser. A, 379 (2194).

3. Bröcker, J. et al, Probabilistic evaluation of time series models: A comparison of several approaches, *Chaos* 19, 043130 (2009), doi 10.1063/1.3271343.

4. R Douc *et al*., Consistency of the maximum likelihood estimator for general hidden Markov models, *Ann. Statist.* 39:1, (2011), pp 474–513, doi 10.1214/10-AOS834.