SELECTION OF DYNAMICAL MODEL USING ANALOG DATA ASSIMILATION

Juan Ruiz[1], Pierre Ailliot[2,5], Thi Tuyet Trang Chau[3], Valérie Monbet[4], Pierre Tandeo[5]

***Abstract*—Data assimilation is a relevant framework to merge a dynamical model with noisy observations. When various models are in competition, the question is to find the model that best matches the observations. This matching can be measured by using the model evidence, defined by the likelihood of the observations given the model. This study explores the performance of model attribution based on model evidence computed using data-driven data assimilation, where dynamical models are emulated using machine learning methods.**

I. MOTIVATION

Data Assimilation (DA) algorithms have been developed to estimate initial conditions for the numerical simulation of complex dynamical systems. More recently, DA methods have been extended as a tool to evaluate different competing numerical models. In this line, [1] proposed a metric called the Contextual Model Evidence (CME), which is based on the likelihood of a set of observations conditioned on a particular numerical model. This likelihood is local, because it is computed on short time intervals, using the forecasts of the model. However, when dealing with high-dimensional systems, running various DA systems, each one with a different model, becomes prohibitive. In this work we propose a data-driven approach, based on the Analog DA (AnDA, [2]) that allows efficient computation of CME based on existing model simulations.

II. METHOD

AnDA uses the analog forecasting method. It is based on a catalog of numerical simulations or historical observations. The objective is to find (in the catalog) analog situations of the current state of the system, to extract the corresponding successors of these analogs, and to adjust a regression between analogs and successors to produce probabilistic forecasts. A key advantage of AnDA is that it can be applied independently to a subset of variables (e.g., a local domain or some integral quantity) without the need to consider the full state space. In this study, the CME method is tested in the AnDA framework.

III. EVALUATION

Preliminary results indicate that CME, using AnDA, is as good as the model-driven method in identifying which are the models that better fit the observations. The method has been tested on the Lorenz-96 system, with a set of noisy observations generated with the forcing term $F = 8$, and using partial observations, covering 12.5% of the domain. Several Lorenz-96 models, with various parameters, from $F = 6$ to $F = 10$ have been compared with the true model (i.e., using $F = 8$).

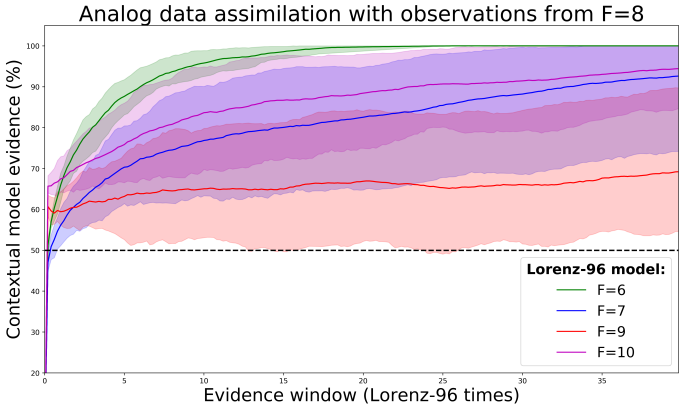

Fig. 1. Percentage of model evidence as a function of the evidence window. Observations are coming from the Lorenz-96 model, with forcing term $F = 8$, and using an additive Gaussian noise with unit variance. The observations are compared to other Lorenz-96 models, using $F = 6$ to $F = 10$. The procedure has been repeated on 10 independent observation sets to get an estimate of the confidence interval.

Results of the CME computed using the AnDA method are given in Fig. 1. The figure shows that the probability to detect the true model from the noisy observations is high, especially for very separated models, and that the evidence to distinguish among different models can be found using observations over a relatively short period of time. Moreover, this detection probability increases with the size of the evidence window on which the computation of the likelihood is done.

The proposed methodology has also been tested on an intermediate complexity climate model. Firsts results show that models with various parameterizations can be attributed.

Corresponding author: J. Ruiz, jruiz@cima.fcen.uba.ar[1] [1]Atmospheric and Oceanography Research Center (UBA-CONICET) [2]University of Brest [3]Laboratoire des Sciences du Climat et de l'Environnement (LSCE) [4]University of Rennes I [5]IMT-Atlantique

REFERENCES

[1] A. Carrassi, M. Bocquet, A. Hannart, and M. Ghil, "Estimating model evidence using data assimilation," *Quarterly Journal of the Royal Meteorological Society*, vol. 143, no. 703, pp. 866–880, 2017.

[2] P. Tandeo, P. Ailliot, J. Ruiz, A. Hannart, B. Chapron, A. Cuzol, V. Monbet, R. Easton, and R. Fablet, "Combining analog method and ensemble data assimilation: application to the lorenz-63 chaotic system," in *Machine learning and data mining approaches to climate science*, pp. 3–12, Springer, 2015.