SPECIAL PROJECT PROGRESS REPORT

All the following mandatory information needs to be provided. The length should *reflect the complexity and duration* of the project.

Reporting year	2021
Project Title:	Data-driven calibration of stochastic parametrization of IFS using approximate Bayesian computation
Computer Project Account:	spgbdutt
Principal Investigator(s):	Dr. Ritabrata Dutta
Affiliation:	Department of Statistics, Warwick University, UK.
Name of ECMWF scientist(s) collaborating to the project	Dr. Nils Wedi and Dr. Peter Dueben, ECMWF.
Start date of the project:	2020
Expected end date:	2022

Computer resources allocated/used for the current year and the previous one
(if applicable)
Please answer for all project resources

		Previous year		Current year	
		Allocated	Used	Allocated	Used
High Performance Computing Facility	(units)	6000000	0	6000000	35045.64
Data storage capacity	(Gbytes)	6400	0	6400	0

Summary of project objectives (10 lines max)

The aim of the project is to develop Bayesian inferential techniques suitable to be applied in the setting of ensemble NWP models. Specifically, this will allow us to perform Bayesian inference of parametrization parameters from observations. Bayesian inference, opposed to frequentist inference, provides a better way to quantify uncertainty starting from previous knowledge. The Approximate Bayesian Computation paradigm is a possible option, as it allows to perform inference tasks on highly complex models, only relying on the capacity to simulate model outputs for some parameter values. In the first phase of the project, we have tailored the aim and understood better the complexities of the issue; specifically, we will focus now on tuning spread parameters (as for instance the parameters governing the SPPT stochastic parametrization scheme, following suggestion by ECMWF scientists) and expand our toolbox beyond ABC methods (see Results section).

Summary of problems encountered (10 lines max)

Lorenzo Pacchiardi (PhD student in Department of Statistics, Oxford) collaborating in this project was supposed to attend the training week on IFS in ECMWF in March 2020, which was postponed to May 2021 due to the coronavirus pandemic. This caused some delay in understanding the details of IFS.

Additionally, Pacchiardi and Dutta diverted some of their time to work on a small research project on modelling the COVID pandemic (https://arxiv.org/abs/2006.16059); this slowed down progress for the present ECMWF project.

Summary of plans for the continuation of the project (10 lines max)

After extensive investigation of the problem backgrounds and some initial tentative trials with the Lorenz95 toy model (see the Results section below), we now have a tentative direction to tackle. Specifically, this entails developing an inferential procedure based on Scoring Rules to tune the spread parameter values, as for instance the parameters governing the SPPT stochastic parametrization scheme. We plan to test our algorithm on the "Simplified Parameterizations, privitivE-Equation DYnamics" (SPEEDY) model, which is more representative of IFS than the toy Lorenz95 model (as SPEEDY for instance includes SPPT scheme) but is still computationally tractable.

Tangentially, we will explore using a fully data-driven approach to perform probabilistic weather forecast, with probabilistic Neural Networks tuned with a Scoring Rule loss function, by making use of the WeatherBench dataset (Rasp et al., 2020).

List of publications/reports from the project with complete references

Pacchiardi, L., & Dutta, R. (2021). Generalized Bayesian likelihood-free inference using scoring rules estimators. *arXiv preprint arXiv:2104.03889*.

Pacchiardi, L., & Dutta, R. (2020). Score Matched Conditional Exponential Families for Likelihood-Free Inference. *arXiv preprint arXiv:2012.10903*.

Summary of results

If submitted **during the first project year**, please summarise the results achieved during the period from the project start to June of the current year. A few paragraphs might be sufficient. If submitted **during the second project year**, this summary should be more detailed and cover the period from the project start. The length, at most 8 pages, should reflect the complexity of the project. Alternatively, it could be replaced by a short summary plus an existing scientific report on the project attached to this document. If submitted **during the third project year**, please summarise the results achieved during the period from July of the previous year to June of the current year. A few paragraphs might be sufficient.

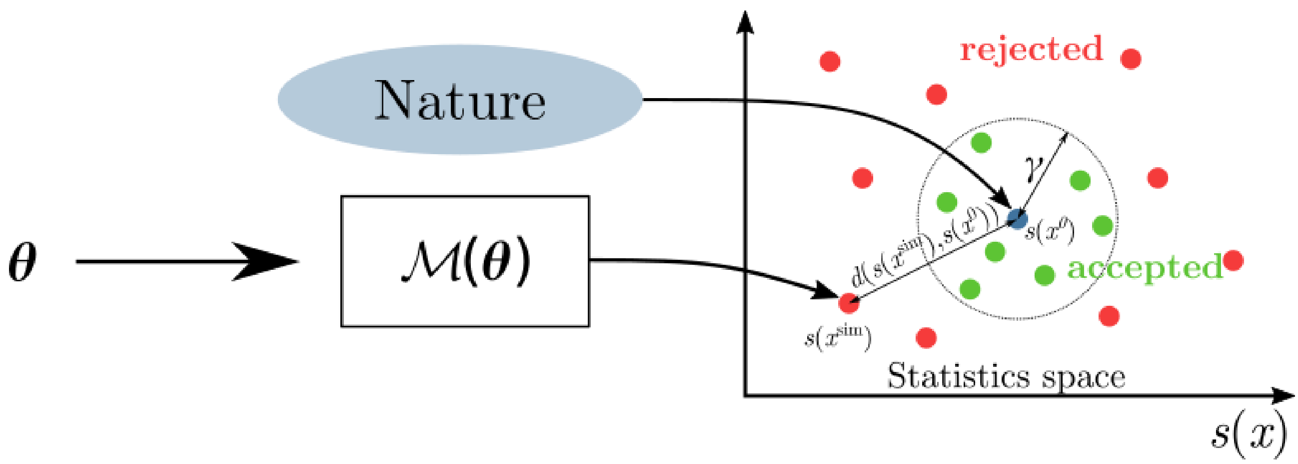
Up to this moment, we have concentrated mostly on investigating methodology and understanding how to best overcome the difficulties presented by handling IFS, which is computationally expensive and outputs large spatio-temporal data. For this reason, we do not yet have any result to present with the IFS model, but we now have a direction to explore which we believe can be applicable to IFS and yield promising results for tuning the parameters impacting the probabilistic distribution of the forecast (see the "Summary of plans" section). In the following we discuss our investigations from the start of the project.

In IFS, parametrization is used to represent sub-grid physical phenomena; these processes are represented as functions of the grid-scale fields depending on *closure parameters* with physical meaning, which are usually hand-tuned.

Additionally, IFS is an ensemble prediction model; as such, the forecast is performed multiple times by introducing small stochastic perturbations in the initial conditions and in the model parametrization schemes; the parameters guiding the amplitude of these perturbations are usually called *spread parameters,* as they directly impact the *spread* of the probability distribution of the resulting ensemble.

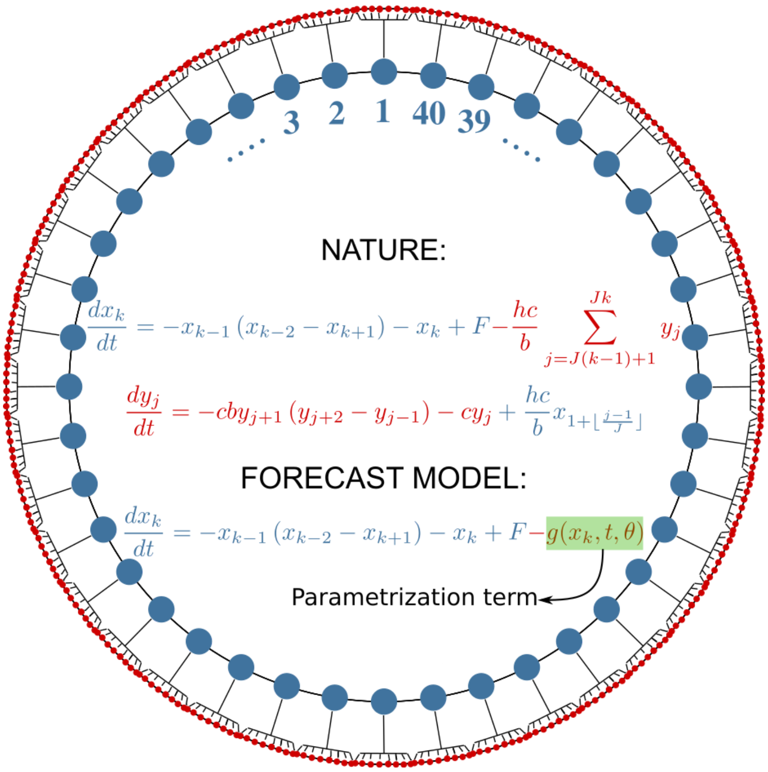
The closure and the spread parameters are conceptually different and therefore different approaches are needed to tune them. Specifically, the closure parameters affect deterministic parametrization schemes, and can therefore be tuned by considering single simulations and their difference with respect to observations (or finer scale simulations). Instead, as the *spread parameters* are only well defined in an ensemble setting, tuning those requires considering the information on the probability distribution defined by the full ensemble.

During the very first part of the project, we developed some techniques for Approximate Bayesian Computation (ABC) which could work for tuning closure parameters. Approximate Bayesian Computation is a technique for performing Bayesian parameter inference for probabilistic model for which the likelihood is not available, but simulations from the model is possible; specifically, it works by drawing a parameter value from the prior distribution, generating a simulation and thereafter accepting or rejecting the corresponding parameter value according to whether the simulation is "close enough" to an observation according to some distance function *d*; the following image shows a graphical illustration of this procedure; more efficient procedures exist, and we apply those in the following when we refer to "ABC".



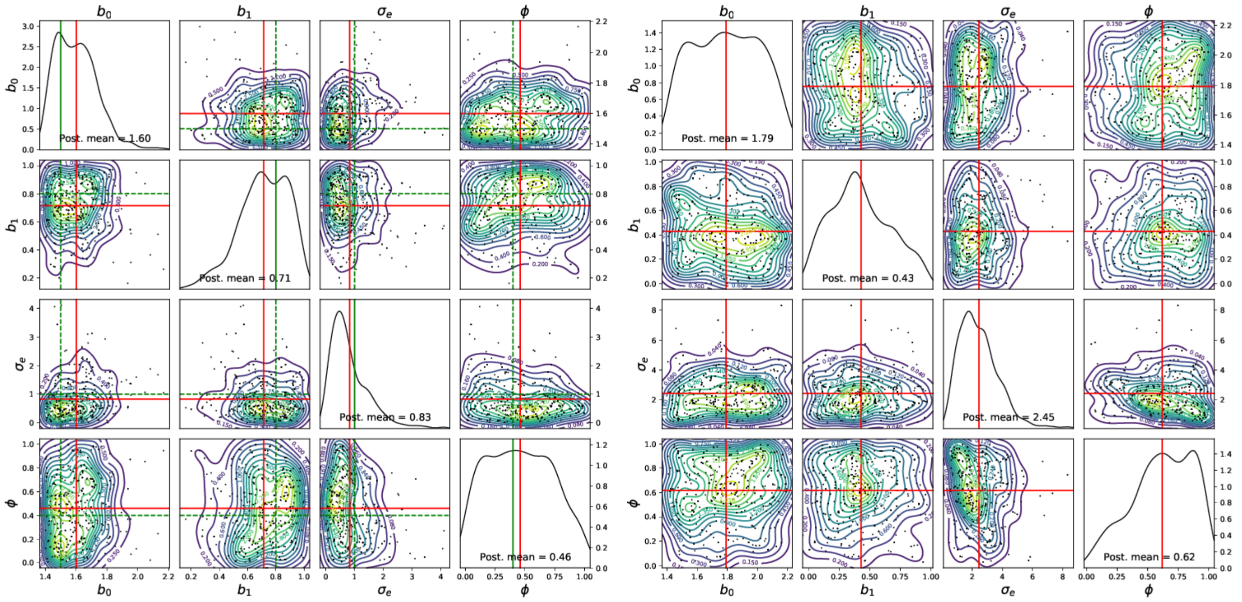
Graphical illustration of Approximate Bayesian Computation (ABC) procedure.

As a benchmark, we considered the Lorenz95 model. Specifically, the Lorenz95 model (see picture below) considers the evolution of two kinds of variables ("slow" and "fast") which are coupled by some ordinary differential equations. It is widely used as a testbench for parametrization studies. The "nature" model considers both the slow and fast variables explicitly, while in the "forecast" model the fast variables are unobserved and replaced by a parametrization term (function g) in the evolution of the slow variables, which depends on some closure parameters.



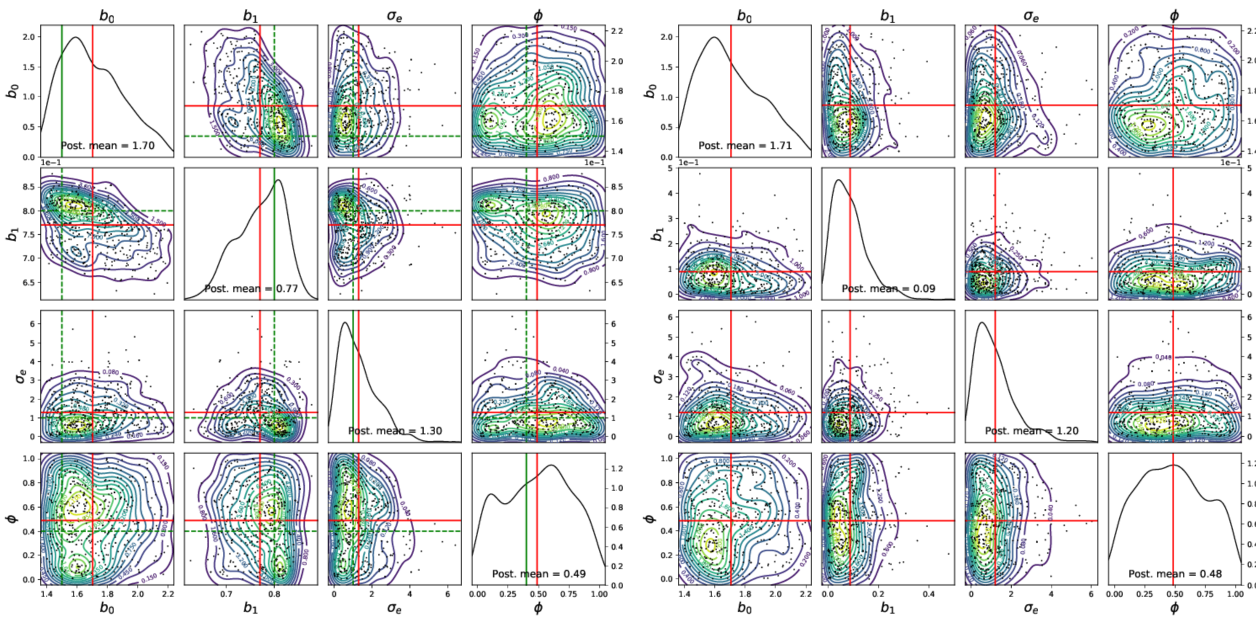
Graphical illustration of the Lorenz95 model; the blue dots represent the "slow" variable x, while the red dots represent the associated "fast" variables y.

In particular, we have considered a fixed parametrization form and have been able to prove the adequacy of ABC for retrieving closure parameter values from an observation. We have applied ABC in this setting in two different ways; in the first approach, we exploited recent techniques in ABC which employ neural networks in order to reduce the dimensionality of the data, thus drastically reducing the computational burden at the expense of a mild loss of information. Moreover, a neural network architecture specific for time-series data has been used. The results of this work have been presented in a poster at the Machine Learning for Weather and Climate Modelling in Oxford, in September 2019. In the images below, we show our results; the left hand side refers to an experiment where the observations were generated from the forecast model for a fixed parameter value (green lines), while in the experiment in the right hand side the nature model was used to generate the observation and the forecast one to perform inference. In both plots, the diagonal panels represent univariate marginal distributions over one single parameter, while the off-diagonal plots represent paired contour plots for the bivariate marginals. Red lines represent posterior mean. In the left hand side figure, the obtained posterior distribution is concentrated close to the true value of the parameter (green line) used to generate the observation.



Posterior results using the Neural Networks summary statistics ABC approach for Lorenz95, for an observation from the forecast model (left) and one from the nature model (right).

The second way in which ABC was applied was by matching the climatological distribution of the Lorenz95 parametrized model; contrarily to the above approach of distance between a set of summary statistics, here we exploit statistical divergences (specifically the Wasserstein distance) to assess the discrepancy between the (high-dimensional) probability distribution sampled by the Lorenz95 model (for different parameter values) in a long run. This builds on recent advances in ABC with statistical divergences (Bernton et al., 2019), and entails running long simulations for each parameter value in order to obtain the invariant probability measure from it, and compare with the observation via the statistical divergence. The results of this approach are presented in the images below; again, the left hand side one refers to an observation generated from the forecast model, while in the right hand side one the observation was generated by the nature model.

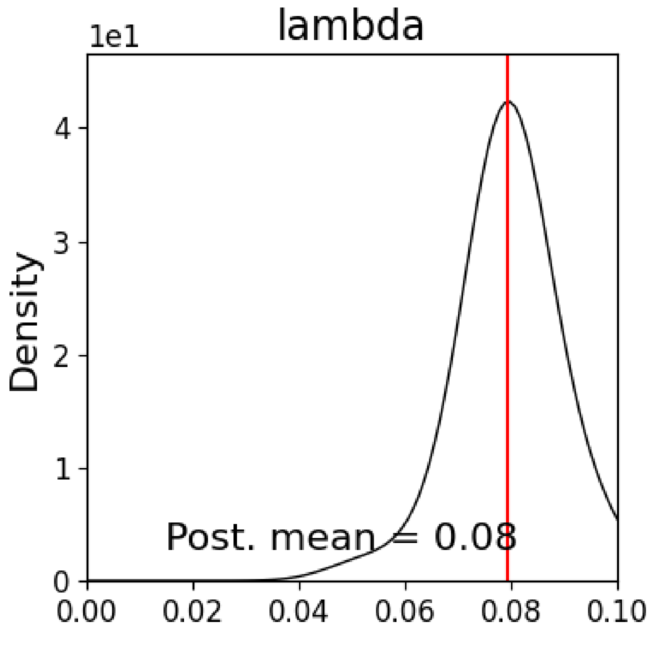


Posterior results using the Wasserstein ABC approach on the climatological distribution for Lorenz95, for an observation from the forecast model (left) and one from the nature model (right).

Although both the above approaches give satisfactory results for the closure parameters for Lorenz95 model, we remark how both could not be applied directly to tune the spread parameters of IFS, as in fact no ensemble information was considered here.

We therefore spent some time considering how tuning spread parameters could be done. To our knowledge, the only work proposing a way to automatically tune the spread parameter is Ekblom et al., 2020. In this work, the initial condition perturbation amplitude for a Lorenz95 ensemble is tuned by using a Differential Evolution approach targeting a quadratic loss function which assesses the spread-skill relationship, which is a metric comparing the ensemble spread to the error of its mean with respect to the observation. In more generality, however, ensemble methods are ultimately evaluated in terms of scoring rules, which assess the probabilistic forecast against an observation (eg. the CRPS for a univariate forecast). It could be possible therefore to exploit the scoring rules as a sort of "loss function" to directly tune the spread parameters; intuitively, tuning with respect to the criterion which is afterwards used to evaluate the model performance should return the best possible model parameters (assuming an infinite amount of computing power).

As a preliminary investigation to this, we considered using ABC by replacing the distance function with the required *scoring rule* between the probabilistic model output and the observation. This approach therefore runs an ABC algorithm on the spread parameter values, simulates an ensemble for each of them and evaluates it with respect to a fixed observation using a chosen Scoring Rule; according to the Scoring Rule value, a given weight is assigned to the parameter value. We have run some preliminary test for this approach by using an ensemble Lorenz95 model in which the closure parameters are fixed and shared between ensemble members (so no model uncertainty is added) while the initial conditions of the different ensemble members are perturbed by some noise with fixed variance lambda. Here, we use the CRPS scoring rule in an ABC setting, and tune the parameter lambda (which was given uniform prior between 0 and 0.1) using an observation generated from the full Lorenz95 model; the following picture shows how this method is capable of obtaining a posterior distribution over lambda:



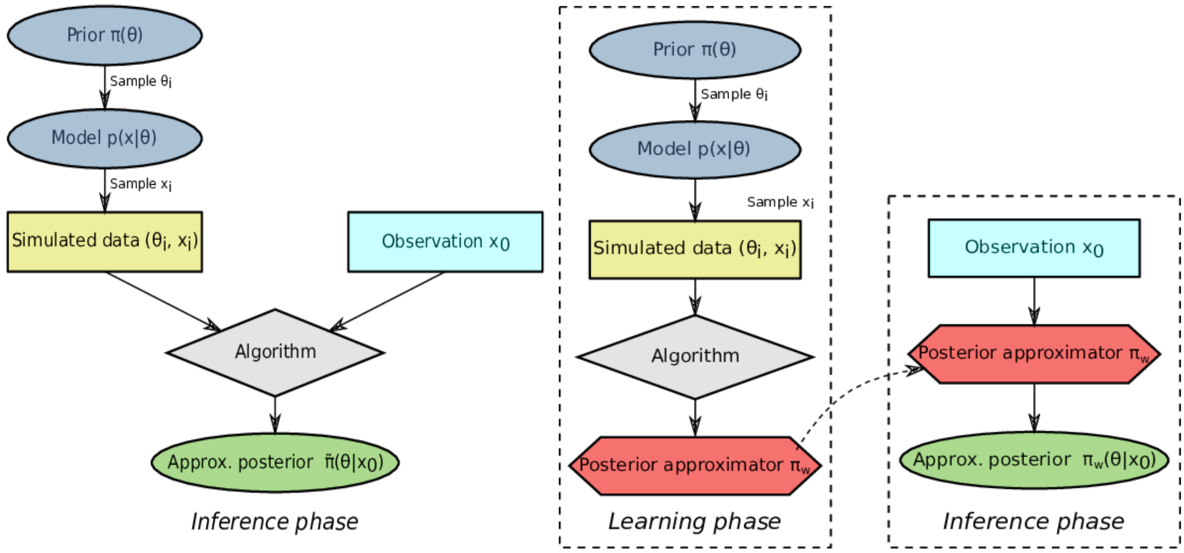
Posterior for the spread parameter in the ensemble Lorenz95 setting.

Parallely and more recently, we have investigated using the scoring rules in a Generalized Bayesian posterior fashion (see for instance Bissiri et al. 2016 for an introduction) to provide a new Likelihood-Free Inference scheme which targets an outlier-robust posterior; see Pacchiardi and Dutta, 2021, where we extensively study both theoretically and empirically this method. There, we did not consider explicitly the setting of ensemble weather models but rather the setup of stochastic simulator models (where a single simulation for a fixed parameter value gives a stochastic output), which is standard in Likelihood-Free Inference research.

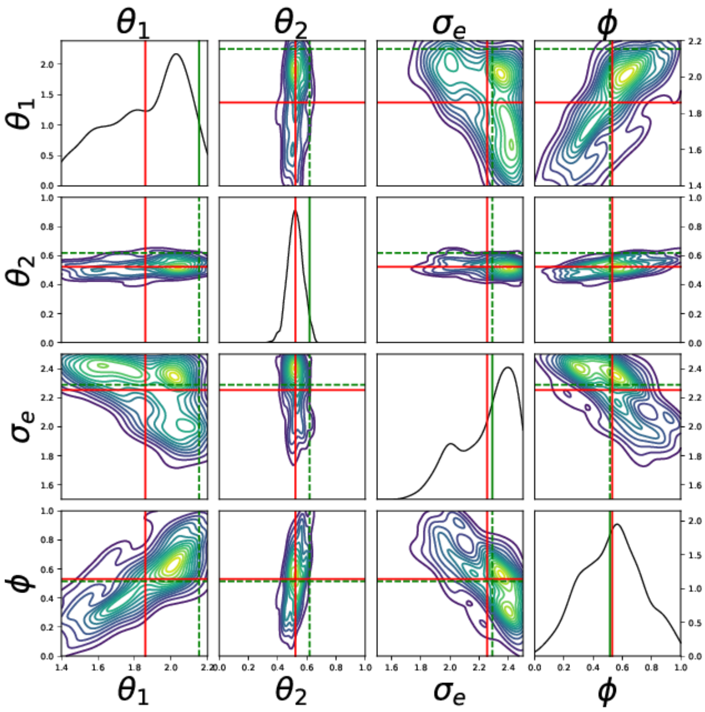
In the approach described above, we build on the recent conceptual realization in the Bayesian inference community that when the model is misspecified with respect to the data-generating process (as it is the case for NWP), generalized Bayesian posteriors work better than the standard Bayes posterior for the sake of prediction; specifically, targeting the scoring rules in some way may lead to better prediction skill, which is ultimately the aim of NWP.

However, both approaches above are computationally demanding and cannot be applied straightforwardly to the IFS; they in fact consider an ensemble run for the full length of the forecast window to evaluate the performance of a single choice of the spread parameters. In the future, we plan to develop a custom method combining these ideas in order to effectively perform tuning of spread parameters (see summary of plans). A possibility is to consider a sequential approach as in the Focused Bayesian Prediction approach (Loaiza-Maya et al., 2019), whose algorithm works by evaluating and updating the choices of parameter values after a prediction for a single time step is made, and then restarting the model and providing a new prediction for the next timestep (this is similar to what suggested in the EPPES approach in Järvinen et al, 2012, which however considered information of the different ensemble members independently and aimed to tune closure parameters). The Focused Bayesian Prediction approach was originally applied to unidimensional financial data series with likelihood models; it offers, however, an interesting starting point for further investigation.

Parallel to this main line of investigation, we have been thinking on how to reduce the computational cost of the algorithm due to the highly expensive model simulations. That can be done by training an "emulator" alongside the algorithm; this emulator (which can be for instance a Neural Network or a Gaussian Process) would learn to predict the important features of the model given the input parameters; in order to be useful, the emulator would need to state the confidence with which its predictions are correct; in this way, whenever the prediction has high confidence, full simulation from the model can be avoided; new simulations are performed if the confidence in the emulator predictions are low, which are used in both the parameter inference algorithm and in further training the emulator. An extreme version of this approach is the one in which the emulator is trained before the observations are even obtained; we call this *amortized* inference setting, as once the emulator is trained inference can be cheaply performed for infinitely many new observations. We investigated a possible approach to achieve this in a recent paper (Pacchiardi and Dutta, 2020), where we also considered the Lorenz95 model as a testbench (in a single simulation setting). This gave some interesting results; however, training the emulator beforehand presents some challenges, as a huge number of simulations need to be used, which likely makes this undoable for IFS. Some intermediate strategies as discussed above (where the emulator is trained alongside the inference scheme) could be instead applicable; we plan to investigate this, if this permits, once the custom method for spread parameters inference is proven to work.



Case-based inference (left) vs. Amortized inference (right).



Inference result for Lorenz95 with the amortized inference scheme proposed in Pacchiardi and Dutta 2020.

Additionally, the high-dimensionality of the IFS output presents some further challenges; the Scoring Rules which are used to evaluate the models have in fact estimation problems in high-dimensional data settings. We envision that some reduction of the model output to lower dimensionality will be needed in all the above cases. For instance, when evaluating a new release of IFS, a score card is used, which expresses the increase (or decrease) of the model performance in predicting some specific atmospheric phenomena with respect to the previous version. An idea which could be applicable when tuning the IFS would be therefore to consider (some of) these same measures and combine them in a single score to tune the IFS.

Finally, we mention here a slightly different line of research which we are currently planning to pursue; specifically, this would entail using a Deep Learning model to perform probabilistic weather forecast in a fully data-driven manner by exploiting the WeatherBench dataset. Previous works (with the sole exception of

Scher and Messori, 2020, to the best of our knowledge) have used that dataset to perform deterministic forecast using neural networks. In order to perform probabilistic forecasting, more advanced neural networks capable of encoding probabilistic information need to be used (for instance Bayesian Neural Networks or Normalizing Flows); additionally, using a suitable combination of scoring rules as a loss function may lead to well-calibrated and accurate probabilistic forecast.

REFERENCES:

Bernton, E., Jacob, P. E., Gerber, M., & Robert, C. P. (2019). Approximate Bayesian computation with the Wasserstein distance. arXiv preprint arXiv:1905.03747.

Bissiri, P. G., Holmes, C. C., & Walker, S. G. (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, *78*(5), 1103.

Ekblom, M., Tuppi, L., Shemyakin, V., Laine, M., Ollinaho, P., Haario, H., & Järvinen, H. (2020). Algorithmic tuning of spread–skill relationship in ensemble forecasting systems. Quarterly Journal of the Royal Meteorological Society, 146(727), 598-612.

Järvinen, H., Laine, M., Solonen, A., & Haario, H. (2012). Ensemble prediction and parameter estimation system: The concept. Quarterly journal of the royal meteorological society, 138(663), 281-288.

Loaiza-Maya, R., Martin, G. M., & Frazier, D. T. (2019). Focused Bayesian Prediction. arXiv preprint arXiv:1912.12571.

Scher, S., & Messori, G. Ensemble methods for neural network‐based weather forecasts. Journal of Advances in Modeling Earth Systems, 2020.

Pacchiardi, L., & Dutta, R. (2021). Generalized Bayesian likelihood-free inference using scoring rules estimators. *arXiv preprint arXiv:2104.03889*.

Pacchiardi, L., & Dutta, R. (2020). Score Matched Conditional Exponential Families for Likelihood-Free Inference. *arXiv preprint arXiv:2012.10903*.

Rasp, S., Dueben, P. D., Scher, S., Weyn, J. A., Mouatadid, S., & Thuerey, N. (2020). WeatherBench: A Benchmark Data Set for Data‐Driven Weather Forecasting. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2020MS002203.