

REQUEST FOR A SPECIAL PROJECT 2020–2022

MEMBER STATE: United Kingdom

Principal Investigator: Dr. Ritabrata Dutta

Affiliation: Department of Statistics, Warwick University, UK.

Address: Department of Statistics, Warwick University, Coventry, CV4 7AL.

Other researchers: Dr. Nils Wedi and Dr. Peter Dueben, ECMWF.

Project Title: Data-driven calibration of stochastic parametrization of IFS using approximate Bayesian computation

If this is a continuation of an existing project, please state the computer project account assigned previously.	Not applicable	
Starting year: (A project can have a duration of up to 3 years, agreed at the beginning of the project.)	2020	
Would you accept support for 1 year only, if necessary?	YES <input checked="" type="checkbox"/>	NO <input type="checkbox"/>

Computer resources required for 2020-2022: (To make changes to an existing project please submit an amended version of the original form.)		2020	2021	2022
High Performance Computing Facility	(SBU)	6 million	6 million	6 million
Accumulated data storage (total archive volume)	(GB)	6400	6400	6400

Continue overleaf

Principal Investigator: Dr. Ritabrata Dutta, Warwick University

Project Title: Data-driven calibration of stochastic parametrization of IFS using approximate Bayesian computation

Complex numerical models of weather/climate dynamics eg. ‘IFS’ developed by ECMWF, are used to predict the weather for days or climate for hundreds of years. These predictions are used by the general public to plan holidays, by farmers for seasonal farming activity, or to assess the effects of global warming on e.g coastal areas of Bengal.

In these simulator models of the Earth-system, the differential equations describing the dynamics of the atmosphere and the coupling to the other components (land, oceans), are integrated over a grid with some spatial resolution and over discrete timesteps. However, many physical processes operate at a smaller space (or time) scale than the considered grid, and their effect on the dynamics of the larger scale has to be taken into account. This is called *parametrization*. At each step of the integration, the effect of sub-grid processes is determined and the corresponding correction (usually referred to as *tendencies*) to the large scale atmosphere prognostic variables is computed. The most straightforward way of doing this is expressing the sub-grid processes as a deterministic function of the large scale state variables; however, this does not lead to satisfactory results, for the following reasons [6]. First of all, the presence of power-law behaviour has been observed in the atmosphere and this, together with the scale invariance of the solutions of the Navier-Stokes equations, prevents the definition of a clear-cut scale separation. In fact, scaling arguments can be used to show that, even with a very high resolution, due to the power law behaviour errors at small scale can contaminate the larger scales in finite time. Moreover, a given large-scale state does not uniquely correspond to a single sub-grid configuration. Therefore, choosing a deterministic parametrization means inserting errors in the system, the fact that these errors can grow quickly may lead to diverging evolution of the system, and no diagnosis of this is possible *a priori*, nor is possible a quantification of the confidence.

Stochastic parametrization of NWP and Calibration A setting offering a solution to this problem is the following: consider an ensemble of simulations, and randomly perturb the computation of the sub-grid processes with different random numbers in the different ensemble members. In this way, each ensemble member represents a possible realization of the sub-grid process, and the spread of the ensemble can be used to provide an estimate of the accuracy of the prediction. The operational ECMWF weather and climate simulators include the Stochastically Perturbed Parametrization Tendency (SPPT) scheme [12], that inserts stochasticity directly at the tendencies level. Specifically, after the tendencies have been computed, a multiplicative noise is added in order to represent uncertainty. This noise is correlated in space and time according to some carefully chosen scales and is the same for each prognostic variable and in each vertical column of the simulation, in order to preserve the vertical profile of the tendency. Experimentally, SPPT has been proven to improve the probabilistic forecast skills of the ensemble simulator. However, as it is a “bolt-on” to the deterministic parametrization schemes, it presents some problems: first, the correlation and intensity parameters for the perturbations are not known; secondly, it leads to some physical inconsistencies of the model, that have to be treated ad hoc. To solve the second issue, another approach the Stochastically Perturbed Parametrizations (SPP) scheme [5] is right now under development. In this scheme, uncertainty is inserted at the model level, namely by perturbing the uncertain physical parameters that define the sub-grid processes. In this way, physical consistency is preserved. Up to now, this scheme has not reached the same capabilities as SPPT, but research is still going on. However, note that this does not solve the problem of fixing the perturbation parameters, as the random perturbation still follows a space-time pattern with some intensity and correlation scale.

These parameters of SPP and SPPT, described above are presently hand-tuned, which may cause unreliable weather forecasting, more evidently in the zones with fewer observations eg. in the trop-

ics. In this project, we will specifically consider the challenging case of *weather prediction in the tropics*, trying to improve it by rigorously learning these parameters in a data-driven manner using approximate Bayesian computation [3]. For this the real weather observation data will be provided by *ECMWF*, while for the tropics we will use a dataset collected by ‘*Laboratoire de météorologie dynamique, France*’ (Concordiasi).

Approximate Bayesian computation to estimate parameters of SPP/SPPT: Bayesian inference schemes for NWP crucially depends on our ability to draw samples from the posterior distribution of parameters of the mechanistic models (θ) given the observed data x^0 , denoted as $p(\theta|x^0)$, which is not known due to the unavailability of the likelihood function of θ for the models. The fundamental rejection Approximate Bayesian computation (ABC) offers a way to overcome this problem, by sampling from an approximate posterior distribution. To sample from this, we first sample θ^* from the prior distribution $\pi(\theta)$ and then simulate x^{sim} from the NWP model $\mathcal{M}(\theta^*)$ using θ^* and finally accepting or rejecting θ^* depending on the probability $\mathbb{K}_\gamma(d(x^{\text{sim}}, x^0))$, where $\mathbb{K}_\gamma(d(x^{\text{sim}}, x^0))$ is a probability density function with a large concentration of mass near x^0 , γ is a threshold and $d(x^{\text{sim}}, x^0)$ is a chosen metric on the dataspace. Finally, in all of the ABC algorithms sampling from the approximate posterior distribution, we decrease $\gamma \rightarrow 0$ at each generations of the chosen sequential algorithm, to improve the approximation and hence being able to approximately sample from the true posterior distribution. Although the choice of $d(x^{\text{sim}}, x^0)$ for ABC is a difficult task and becomes crucial for the present problem as the dataset is of spatio-temporal nature containing several weather variables.

Distance learning for ABC: As for NWP model, we are only interested in inference of the parameters, we may want to learn the distance on data space by inducing the geometry in parameter space on data space. To do so, we first simulate n data sets $\{x_1, \dots, x_n\}$ from n parameters $\{\theta_1, \dots, \theta_n\}$, respectively, which are randomly sampled from prior distribution. We can capture the geometry in the parameter space, through two sets of pairwise similarity and dissimilarity constraints, $\mathbb{S} = \{(x_i, x_j) | x_i \text{ and } x_j \text{ are similar}\}$ and $\mathbb{D} = \{(x_i, x_j) | x_i \text{ and } x_j \text{ are dissimilar}\}$, where x_i and x_j are considered similar if $d_E(\theta_i, \theta_j) < \varepsilon$ and dissimilar otherwise for some $\varepsilon > 0$. A *Mahalanobis distance* between x_1 and x_2 , $d_M(x_1, x_2) = (x_1 - x_2)^T M (x_1 - x_2)$, where M is a $d \times d$ positive semi-definite matrix, under the above similarity and dissimilarity constraints, can be learned by finding M satisfying:

$$\min_M \left[\sum_{i,j=1}^n (x_i^T M x_i - x_i^T M x_j) K_{ij} \right], \text{ s.t. } M \geq 0 \text{ and } K_{i,j}^x = \begin{cases} +1, & \text{if } (x_i, x_i) \in \mathbb{S} \\ -1, & \text{if } (x_i, x_i) \in \mathbb{D}. \end{cases}$$

Existing metric learning methods can be divided into two categories: linear metric learning methods to learn Mahalanobis distance as above and nonlinear methods applying kernel tricks [11] or deep neural networks to model higher-order correlations. In this project, our focus will be on deep metric learning (DML) which combines the ability of deep learning to learn nonlinear feature representation and discrimination power of metric [9, 10] and is the state-of-the-art. We plan to use DML algorithms using convolutional neural networks (CNNs) [13], multilayer perceptron or a deep adversarial framework [1]. We will highlight here that we are proposing DML to learn the distances for ABC rather than purely using deep learning for prediction. *This enables us to bring together the best of two worlds: deep data-driven learning of nonlinear maps and scientific models based on physical principles, providing a novel framework to combine machine learning and natural sciences.*

We should note that the similarity and dissimilarity illustrated in the simple example above is defined using contrastive loss between pairs of data points. In this project, we will also explore losses using more than two data points for DML, eg. triplet loss [11, 7], N-pairs loss [8], facility location framework [4], angular loss [10] etc, and develop new losses capturing the geometry better for the improvement of the the inferential performance of ABC.

ABCpy: HPC-driven Python package for ABC For this project we will use a Python package called ABCpy developed by the PI with the help of Swiss National Supercomputing Centre (CSCS). This Python package needs MPI for parallelization/nested-parallelization. Further information about this software can be found in <https://github.com/eth-cscs/abcpy>. ABCpy implements the most advanced ABC algorithms and optimally utilizes high performance computing (HPC) facilities. The ABC algorithms in ABCpy are sequential algorithms, for which the jobs at each of the sequence is parallelized. The computational cost of ABC algorithms depend on three multiplication factors, (a) simulation time of simulator model, (b) iteration steps, (c) sample points. At each iteration step, we need to simulate from the ‘simulator model’ for ‘sample point’ many times. Now, if the simulator model takes more than 30 milliseconds for one simulation, then the algorithm scales almost linearly upto ‘sample points’ many core, if each simulation is done on a core (This happens as we need to simulate sample points many simulations and we do them at each core). In figure [1], I consider two models with 12 seconds and 2 minute simulation time correspondingly, iteration steps = 2 and sample point =5000. This is a very simplified example used to illustrate that we get a linear speedup for expensive simulator model and a reasonable speedup for the cheaper model. Further speedup studies with the first model (simulation time 12 secs), can be found in [2].

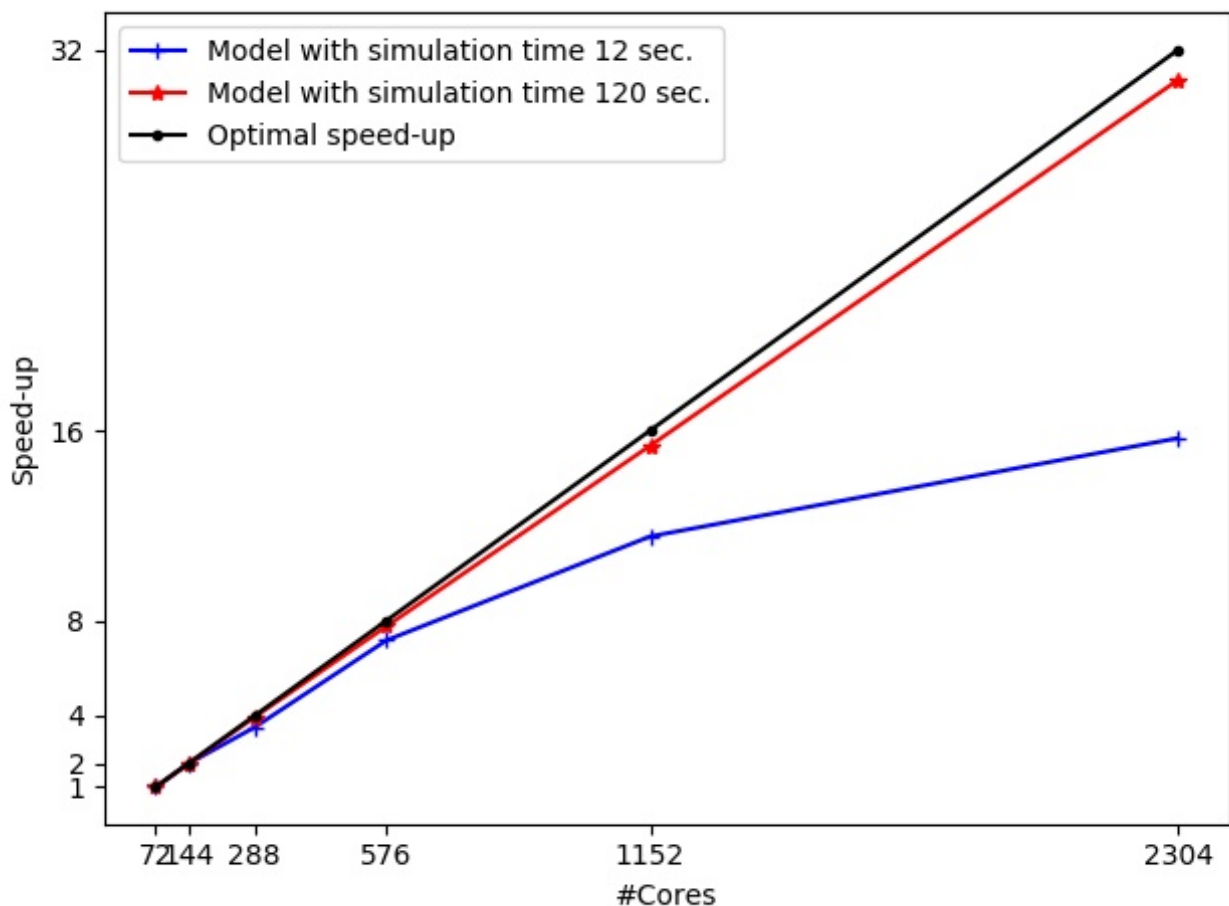


Figure 1: Speed-up of ABCpy

Requested resources In this project proposal we consider NWP model implemented in openIFS for which the simulation time is 30 minutes and more, the iteration steps would be 10-20 and sample points should be between 10000-50000. Hence we expect to have linear speedups in this case upto number of cores equal to the sample points. We can also further parallelize our simulation model

openIFS using 24 cores of a node. Then we will get a linear speed upto sample point many nodes, using nested parallelization scheme of ABCpy. Based on these parameters, we expect to run yearly 5 big jobs (50 nodes for 24 hours) and 50 small jobs (10 nodes for 5 hours) totalling to 6 million SBUs.

References

- [1] Yueqi Duan, Wenzhao Zheng, Xudong Lin, Jiwen Lu, and Jie Zhou. Deep adversarial metric learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2780–2789, 2018.
- [2] Ritabrata Dutta, Marcel Schoengens, Avinash Ummadisingu, Nicole Widmer, Jukka-Pekka Onnela, and Antonietta Mira. Abcpy: a high-performance computing perspective to approximate bayesian computation. *arXiv preprint arXiv:1711.04694*, 2017.
- [3] Jarno Lintusaari, Michael U Gutmann, Ritabrata Dutta, Samuel Kaski, and Jukka Corander. Fundamentals and recent developments in approximate Bayesian computation. *Systematic biology*, 66(1):e66–e82, 2017.
- [4] Hyun Oh Song, Stefanie Jegelka, Vivek Rathod, and Kevin Murphy. Deep metric learning via facility location. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5382–5390, 2017.
- [5] Pirkka Ollinaho, Sarah-Jane Lock, Martin Leutbecher, Peter Bechtold, Anton Beljaars, Alessio Bozzo, Richard M Forbes, Thomas Haiden, Robin J Hogan, and Irina Sandu. Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ecmwf ensemble. *Quarterly Journal of the Royal Meteorological Society*, 143(702):408–422, 2017.
- [6] TN Palmer. Towards the probabilistic earth-system simulator: a vision for the future of climate and weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 138(665):841–861, 2012.
- [7] Matthew Schultz and Thorsten Joachims. Learning a distance metric from relative comparisons. In *Advances in neural information processing systems*, pages 41–48, 2004.
- [8] Kihyuk Sohn. Improved deep metric learning with multi-class n-pair loss objective. In *Advances in Neural Information Processing Systems*, pages 1857–1865, 2016.
- [9] Evgeniya Ustinova and Victor Lempitsky. Learning deep embeddings with histogram loss. In *Advances in Neural Information Processing Systems*, pages 4170–4178, 2016.
- [10] Jian Wang, Feng Zhou, Shilei Wen, Xiao Liu, and Yuanqing Lin. Deep metric learning with angular loss. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2593–2601, 2017.
- [11] Kilian Q Weinberger and Gerald Tesauro. Metric learning for kernel regression. In *Artificial Intelligence and Statistics*, pages 612–619, 2007.
- [12] Antje Weisheimer, Susanna Corti, Tim Palmer, and Frederic Vitart. Addressing model error through atmospheric stochastic physical parametrizations: impact on the coupled ecmwf seasonal forecasting system. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2018):20130290, 2014.
- [13] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Deep metric learning for person re-identification. In *2014 22nd International Conference on Pattern Recognition*, pages 34–39. IEEE, 2014.