The use of machine-learning to improve localization of background covariances in Variational Data Assimilation.

Supervisor: Dr Alison Fowler

How can we use machine learning (ML) to reduce sampling errors in ensemble-based data assimilation (DA)?"

Data assimilation is the process of combining imperfect models with observations. Many operational weather centres use ensembles to quantify imperfections in model forecasts to help determine how the forecasts should be corrected with observations. Ensemble information is extremely valuable but does introduce statistical noise into the DA process. A method called localization helps solve this problem, but it can be expensive in practice. It is proposed that ML can improve efficiency and help to solve this problem.

As an example application, we propose using ML to mimic advection processes to describe how localization should change over time. This is important when the localization length scale is shorter than the propagation of error structures through the DA time window. Just as ML can emulate forecast models, it should be able to take account of differences in advection velocities over time and space.

We plan to use the state-of-the-art Joint Centre Satellite Data Assimilation (JCSDA) technical infrastructure and the incredibly flexible background error covariance framework with the System Agnostic Background Error Representation (SABER) repository. This is fully in line with the Met Office's system development framework.

A simple model will be used for the initial investigation and optimal localization theories will be used to train the ML schemes. Thereafter extensions to the project could involve using it with a scale-dependent localization scheme (Caron, 2023) and moving to more realistic models of the atmosphere.

Making better use of near-surface observations in coupled atmosphere-ocean prediction

Supervisor: Professor Amos Lawless

In the last few years operational weather forecasting centres, such as the Met Office, have started to use coupled atmosphere-ocean models to produce their regular weather forecasts. Using a coupled model allows a better representation of the influence of the ocean on the atmosphere, which is important for predicting high-impact weather events such as storms and tropical cyclones. The models are initialised using the latest observations of the atmosphere and ocean with a mathematical technique known as data assimilation, which combines the observations with the model, taking into account their respective uncertainties. Some observations that are close to the surface (such as satellite measurements of sea-surface temperature) give useful information about both the atmosphere and ocean state, but current data assimilation methods are not adequate for exploiting this information fully. In this project we will address this challenge in two stages. First, we will improve the mathematical mapping between the model state and the observations in the data assimilation process, known as the observation operator. We will design an improved operator that incorporates more completely the physical relationship between the measurements and the coupled atmosphere-ocean state. We will then investigate how we can best generate ensembles of coupled model forecasts to obtain information about the

uncertainty in our predictions by a better representation of the uncertainty in near-surface observations. New methods will be developed mathematically and tested using idealised models. The most promising ideas will then be tested in a research mode of the Met Office's forecasting system.

Where do we need the high resolution in the ocean?

Supervisor: Professor Bryan Lawrence

Most of the European community uses coupled global ocean-atmosphere models (GCMs) which include an ocean component with a globally uniform mesh, even though we know that we need finer meshes in some regions than others. We do this because the impact of using variable resolution on climate is hard to understand, and both running and analysing variable resolution models is not easy. The overall purpose of this project is to address some of those concerns by experimenting with models which focus higher resolution in selected regions, to address the question "Can we use variable resolution in the ocean component of very high resolution coupled model systems in such a way that the simulations can be made enough cheaper to afford more of them, and hence better sample climate variability?"

The exact models to be used are not yet known but are likely to involve ones where either the ocean resolution is natively variable or there are one or more embedded high-resolution regions. Many different directions are possible ranging between very scientific and quite computational, addressing questions such as: "How much does atmospheric variability depend on ocean variability in regions with large-scale eddies or low eddy activity?"; "Can we resolve those regions at lower resolution than active regions with small-scale eddies and still get realistic and useful climate simulations?"; "Can this approach lead to simulations that can be made cheap enough to generate usefully larger ensembles?"; "How can we best analyse models to answer these sorts of questions?"

Cloud feedbacks on tropical climate at subseasonal scales

Supervisor: Professor Chris Holloway

In the tropics, clouds and the large-scale circulation are intimately coupled. The large-scale circulation controls the location of clouds, which in turn feedback on the circulation though latent and radiative heating and transport of air parcels. Consequently, clouds play a crucial role in numerous tropical features, including the Intertropical Convergence Zone, the Madden-Julian Oscillation, and the El Nino Southern Oscillation. Cloud-circulation coupling also contributes to uncertainty in cloud responses to climate warming.

Despite its importance, the coupling between clouds and circulation remains relatively poorly understood. This is at least in part due to the fact that representing this coupling in numerical models requires a computationally expensive combination of high resolution to resolve the convective and cloud scales and a large domain to capture circulation features.

This project will exploit kilometre-scale simulations to investigate the role of clouds in shaping tropical climate dynamics on subseasonal timescales. Clouds and circulation features will be evaluated against observations to understand how well these are captured by the model. Cloud locking experiments will be used to identify the role of clouds in tropical

dynamics in the model. Based on this work and expertise from AFESP modelling centre partners, further experiments will study how perturbing specific model parameters change tropical cloud properties and consequently circulation with the aim of identifying physical processes where there is potential for improvement.

Downstream impacts of embedded convection in warm conveyor belts

Supervisor: Dr Oscar Martinez-Alvarado

Recent research has shown that embedded convection within warm conveyor belts (WCBs) can influence the occurrence of heavy precipitation and the development of heatwaves. The misrepresentation of WCB embedded convection can importantly impact the development of forecast errors, affecting weather prediction skill. However, further research is needed to understand case-to-case variability.

Separate research on convection-permitting climate simulations has shown more intense precipitation and damaging winds than their lower-resolution counterparts, likely produced by processes, such as WCB embedded convection, that are not well represented at low resolutions. These results highlight the need to understand the effects of these processes on downstream HIW under different model resolutions.

This project will investigate the following research questions:

- What processes determine the effects of WCB embedded convection on downstream impact?
- How do differences in process representation in numerical models across different resolutions lead to divergent evolutions and forecast error?

Machine learning driven balance relationships for next generation data assimilation systems

Supervisor: Dr Ross Bannister

This proposal is about km-scale data assimilation (DA). Effective DA improves predictability by extending the time range of useful forecasts, and eliminating the presence of spin-down effects (transient error growth rates at the initialisation time). Doing this well – using imperfect observations – requires knowledge of how each model grid point and variable should be properly coupled to other (neighbouring) grid points and variables (the 'forecast error statistics'). In this respect, current DA systems were designed with large-scale systems in mind, and so a different approach is needed for high-resolution models. The goal of this project is to study machine learning (ML) techniques in the context of this problem. This is a new frontier in DA science.

There are exciting research questions relevant to km grid length models (regional and global models). How to estimate the 'true' forecast error statistics? How do they change in time? Which ML methods can be applied to this problem to efficiently reproduce the true statistics? How can these methods be trained? How do the outcomes of the ML compare with traditional methods? What are the implications for forecasting?

The key innovation will be in the combination of ML with DA in order to determine and efficiently use flow-dependent forecast error statistics. Such knowledge is useful for the current application of DA to state-of-the-art high-resolution models, but may also help guide purely ensemble-based DA methods for the purpose of reducing sampling noise. The

method could also be applied to the representation of flow-dependent atmosphere-ocean coupled covariances.

How skilful are AI-based forecasts of monsoon weather?

Supervisor: Dr Reinhard Schiemann

Artificial-intelligence-based forecasts may revolutionise weather prediction, yet limited understanding of how well, and when and why, these new models work creates unease when applying them operationally. There is a pressing need to learn how to use these models safely and evaluate their usefulness to ensure they meet the needs of forecasters and contribute to public safety (Ebert-Uphoff and Hilburn, 2023).

Our overarching question is: how confident can we be in AI-based forecasts of monsoon weather at lead times between a few days and four weeks ahead? The project will consider:
- How does the skill of AI models in predicting monsoon weather patterns compare to that of traditional NWP models as a function of lead time?
- How does skill depend on large-scale extra/tropical forcing, e.g., phases of the BSISO and Silk Road Pattern, stage of the monsoon front progression?
- How can we best create AI ensemble forecasts, what is the probabilistic skill, and can it be improved with post-processing?
- How well are extreme precipitation events and their precursors (TCWV, circulation) forecast?

We will run novel AI models (PanguWeather (Huawei), GraphCast (Google DeepMind), FourCastNet (Nvidia), potentially ECMWF's AIFS) to produce 30-day hindcasts for monsoon seasons on the UoR or JASMIN cluster. One week of global 0.25° forecasts takes about one minute so we will be able to create large ensembles by perturbing initial conditions. We will target seasons after the AI models' training period (1979-2020) and use conventional methods and explainable AI to identify how and where wind and humidity biases develop.

S2S prediction of marine heat waves and associated compound events

Supervisor: Professor Ted Shepherd

Marine heat waves (MHWs) are an extreme event type of growing scientific interest, because they have important effects on ecosystems and fisheries (Rodrigues et al. 2019). The frequency and intensity of MHWs are expected to increase with global warming, as is evident in recent observed trends. From an S2S perspective the trend is a complicating factor: using a fixed SST threshold means the same statistical event reflects a different configuration of causal factors and is thus a different physical event, whilst using a moving threshold means that the impacts will be different. **The main research question in this project is how to meaningfully evaluate and communicate MHW predictions on the S2S timescale in a non-stationary climate.** This question lies at the intersection of statistical and physical reasoning. Our working hypothesis is that this can be done by treating MHWs as short-term (probabilistic) events riding on top of both long-term (trend) and medium-term (low-frequency variability) components, both of which can be regarded as known in the S2S context. This combination of causal factors is important because compound aspects of the event, e.g. drought over adjacent land areas associated with anti-cyclonic atmospheric blocking conditions, will have a different relationship with the MHW on

the different timescales. The spatial structure of the MHW within the ocean would be an important reflection of these differences.

Understanding the potential for month-ahead prediction for monsoon systems

Supervisor: Professor Andy Turner

Active and break events in monsoon rainfall, lasting a week or so, represent the extremes of intraseasonal variability, with profound implications for water supply and agriculture. The Boreal summer intraseasonal oscillation, which varies over 30-60 days, provides the known predictability at the S2S range, but its prediction skill in models is limited. This project will seek to answer the question: Can monsoon intraseasonal prediction be extended to the one-month lead time? The approach will be to first reduce the dimensionality of reanalysis data, e.g., using empirical orthogonal functions, and then use novel methods such as decision trees and causal networks to identify drivers of subseasonal monsoon variability, to seek previously unknown sources of predictability in the tropics and extratropics. The project will also assess the stratosphere, which has known interactions with ENSO and the MJO but whose influence on the monsoon is poorly explored. We will use conditional approaches to understand whether the links between subseasonal drivers and monsoon rains, and thus subseasonal prediction skill, are dependent upon the state of slowly varying seasonal drivers such as ENSO, the IOD or QBO. Modelling work at climate and convection-permitting scales will be used to understand to what extent subseasonal drivers can be represented, whether the kilometre scale is necessary to simulate such behaviour, and to perform case-study experiments to define the mechanisms involved. We will leverage the new NERC MiLCMOP project (starting 2023) in which we test the role of the extratropics in monsoon onset using modelling and causal analysis techniques.

Breaking the barriers of predictability in hydrological forecasting

Supervisor: Professor Hannah Cloke OBE

Global forecasts of upcoming river flows and water resources can now be made from 1 week to a few months ahead, but the skill of these forecasts varies widely in space and time. Anticipatory humanitarian action has the potential to provide targeted, timely support before a disaster strikes. Humanitarian organisations operating in parts of the world with extremely vulnerable communities, such as those displaced by conflict in South Sudan, have reached out for improved information to support anticipatory action to prepare for floods and droughts. But it is in just such places that hydrological forecasting models have poor skill. The aim of this PhD is to investigate the current barriers to improving hydrological prediction at the Earth System scale, and consider how these interact to currently limit the lead time of skilful subseasonal hydrological forecasts. The project will explore how ongoing developments in predicting hydrological flows in Earth System could provide more accurate and earlier forecasts of upcoming floods and droughts. These developments range from better representation of soil and snowmelt processes, groundwater dynamics, inclusion of dams, reservoir representation and upstream water management in the river channel network, improvements in the representation and postprocessing of precipitation, and new observations such as data from SWOT.

Discovering the Mechanisms Behind "Forecast Busts"

Supervisor: Professor Robert Plant

Numerical weather prediction (NWP) occasionally presents "forecast busts", in which the forecast skill at five- to six-day lead time drops to almost zero across the world's leading NWP centres. Such failures can be linked via Rossby wave dynamics to an initial poor representation of Mesoscale Convective Systems (MCS) upstream, which is in turn related to systematic difficulties in simulating moist convection. Our main research question is why errors in the representation of an MCS are normally benign for predictability but occasionally catastrophic? Answering this will provide insight into how multi-scale error growth in NWP systems depends on flow regimes.

The project will involve making detailed investigations of simulated cases and comparing situations that do and do not lead to busts. A key element will be understanding the development of errors in the forecast model: their propagation downstream and growth upscale from order 1km to the synoptic. This will involve analyses and interpretations in the frameworks of potential vorticity using process-oriented diagnostics to unravel the contributions of different physical mechanisms in the model. We will include simulations with different representations of convection, expected to lead to some significant changes at the MCS stage which may then alter the coupling to larger-scale weather patterns. Will new treatments and/or higher resolution solve the forecast bust problem, and why, or might they even make it worse?

Developing Artificial Intelligence Approaches to Enhance Representations of Turbulence in Atmospheric Models

Supervisor: Dr Todd Jones

Turbulence is one of the many processes represented in simulations of the Earth System. Better understanding of the effects of turbulence on dynamics and clouds in weather/climate models would allow improving its parameterisation in these models. Explainable Artificial Intelligence (XAI) is concerned with allowing human users to better understand the role various variables play in complex models, while Machine Learning (ML) techniques such as parameter tuning allow tuning model parameters for better performance in downstream tasks. This project will investigate and contrast three approaches to the parameterisation of turbulence: traditional parameterisations, traditional parameterisations tuned using ML, and emulations of turbulence developed using XAI techniques. The goal of this project is to investigate whether AI/ML techniques can be used to improve the accuracy and computational efficiency of these parameterizations, either by improving scientific outcomes at the same computing cost or making simulations cheaper. The main outcomes of the project will be determining whether AI/ML techniques can improve the fidelity of the Met Office NERC convection (MONC) model simulations and making them more computationally efficient. We are very sure that the "right way" to use AI/ML in any given turbulence situation is not well known, but improving turbulence in MONC could lead directly to improvements in future weather and climate models.

Improving river streamflow forecasts using deep learning techniques

Supervisor: Dr Kieran Hunt

Artificial neural networks - particularly those designed to process sequential data, known as long short-term memory networks (LSTMs) - can improve forecasts in areas where data sparsity or incomplete process knowledge challenges conventional dynamical models. This project will explore how LSTMs can fill such gaps, mitigate against biases in input data, and produce accurate river streamflow forecasts. To achieve this, the project will vastly expand on a prototype LSTM designed to ingest weather forecasts and produce operational streamflow forecasts over the US.

Research questions include:

- What is the optimal setup for the LSTM? How are gridded variables best processed - by statistical preprocessing, or by additional components in the LSTM (e.g. convolutional layers)?
- How can km-scale hydrological models and science be improved to develop a hydrological Digital Twin?
- How well does the LSTM perform against benchmark dynamical models in both data-rich and data-poor regions? How can performance in data-poor regions be improved?
- Can the LSTM effectively forecast extreme flooding events that are beyond the distribution of the training data?
- What is the effect of climate change on streamflow and flooding risk in vulnerable catchments?

Improving the Efficiency of Weather and Climate Prediction by developing Mathematical Methods to take Long Time Steps

Supervisor: Professor Hilary Weller

State of the art weather and climate prediction models are efficient and even accurate when large time steps are taken due to the use of semi-Lagrangian transport schemes. However semi-Lagrangian is not conservative – the total amount of the transported quantities can change due purely to numerical errors. Next generation models, such as ECMWF's FVM (finite volume model) are designed for higher resolution and designed to run on massively parallel computer architectures and they do not use semi-Lagrangian transport so as to avoid conservation errors. However this means that they face time step restrictions that can be severe, for example in the presence of strong updrafts associated with severe weather. The supervisors have started to develop implicit time stepping schemes that are stable and accurate for much longer time steps and retain exact conservation. These schemes need further development and testing before they can be used operationally. We need to answer questions such as:

1. Can larger time steps using implicit transport lead to reduced computational cost in comparison to (traditional) explicit transport with a smaller time step?

2. Would it be beneficial to use implicit transport only in the vertical direction, where it is most needed?

3. If implicit transport is used only locally, where it is most needed, how does this affect load balancing between multiple computing processors?

4. With the use of implicit time stepping, can we increase the vertical resolution without reducing the time step and still improve accuracy?

Multi-Scale Parameterization for the Convection Grey Zone

Supervisor: Professor Bob Plant

The project will assess the prospects for a ***multi-scale parameterization*** suitable for use in the convective grey zone: what benefits could that provide, and what properties would the scheme need to have to realize these benefits? The concept is feasible only because of the functionality of LFRic, which can solve the model dynamics and (parts of) the physics on different grids. Thus, we will pioneer use of LFRic within academia and tackle scientific problems that were not previously accessible.

It is characteristic of a grey zone that many studies show clear benefits at o(1km) scale when switching off the convective parameterization, while many others point to serious drawbacks. Our underpinning assumption is that parameterizations *do* provide valuable information if applied on a scale where their assumptions are valid. We envisage running the parameterization at o(10km), with information derived from it leading to additional tendencies imposed on the dynamics grid of o(1km). We call this a multi-scale parameterization because radiation, microphysics etc continue to operate at o(1km).

We will use the new CoMorph parameterization in online diagnostic mode within sub-km simulations. We will compare its predictions of convective properties with those produced by the resolved convection and assess dependencies on the space-time scales of the inputs provided to CoMorph. A machine-learning element might be useful to capture the relationships. From these findings, concepts for multi-scale parameterizations will be devised and tested in o(1km) simulations, establishing what structural characteristics are necessary for a useful multi-scale representation that better matches sub-km results.

Unleashing the potential of high-resolution sub-seasonal tropical cyclone predictions

Supervisor: Associate Professor Linda Hirons

Flooding and strong winds associated with tropical cyclones in the South-West Indian Ocean (SWIO) severely impact vulnerable Southern African communities every year. Access to reliable, actionable early warnings on sub-seasonal timescales are essential for supporting anticipatory actions and building resilience. This ambitious project seeks to address a critical question towards this: **Do sub-seasonal predictions of tropical cyclones improve when harnessing regional kilometre-scale model configurations?**

High-resolution modelling has significantly advanced our ability to forecast the intensity and track of tropical cyclones on daily timescales. However, whilst sub-seasonal forecasts can identify regions at risk of cyclones, they struggle to accurately predict observed intensification and longevity of cyclones. Addressing this gap, the project will first quantify forecast skill of existing sub-seasonal predictions for high-impact SWIO cyclones. Building on this, it will then perform regional kilometre-scale simulations driven with state-of-the-art sub-seasonal forecasts. The project will evaluate how and why this modelling approach improves predictions of key cyclone characteristics which remain challenging to forecast: genesis, track, intensification and associated rainfall. In collaboration with regional institutions - Meteo-Madagascar, East African Climate Prediction and Applications Centre (ICPAC) - we will contribute to international efforts to improve sub-seasonal tropical cyclone advisories. The student will engage with relevant international University of Reading projects (REPRESA and ACACIA), and support early warnings issued to vulnerable communities, aligning with global initiatives such as the WMOs Sub-seasonal to seasonal Applications for Agriculture and Environment (SAGE). This project aims to advance both scientific understanding and useful application of sub-seasonal predictions to mitigate the impacts of cyclones.

The Roles of Tropical Waves in the Sub-seasonal Prediction of Tropical Cyclones

Supervisor: Associate Professor Xiangbo Feng

Tropical cyclones (TCs) are among the most hazardous weather systems worldwide. They pose a major threat to life, properties and ecosystems in coastal regions. As climate change amplifies exposure and vulnerability to natural hazards, gaining a better understanding of TC activity has never been more critical. However, the sub-seasonal timescale is still considered as a "predictability desert" for extreme weather due to limited understanding of the sources of S2S variability. The current prediction models have very limited or no skill in predicting the deviation of TC activity (i.e., anomaly) from seasonal climatology beyond a week.

A recent study (Feng et al., 2023) has shown that pre-existing tropical waves, including equatorial waves (EQWs), can explain the sub-seasonal variability of TCs from a global perspective. However, the nature of the wave-TC interaction and how the wave-associated dynamical and thermodynamical processes influence TC genesis and intensification is largely unknown. Furthermore, it is critically important to diagnose and understand how these interactions are represented in global forecast models, to better understand the large-scale drivers and potentially improve the sub-seasonal forecasts in practice.

The main research questions are: (1) what dynamical and convective processes are responsible for the modulation of TC activity (e.g. TC seeds, generation, and intensification) by tropical waves? (2) How well are these processes represented in the k-scale climate models and operational forecasts (e.g., Met Office and ECMWF systems)? (3) How sensitive is the TC-wave interaction to large-scale drivers (e.g., regionality, seasonality, El Niño Southern Oscillation)?

Prediction of high impact extratropical cyclone clustering events

Supervisor: Professor Helen Dacre

Clusters of extratropical cyclones, which follow one another in close succession, can result in rain falling on already saturated ground leading to costly flooding events. For example, in the UK, in December 2015 with storms Desmond, Eva and Frank, and more recently in February 2022 with storms Dudley, Eunice and Franklin. These cyclones cannot be treated as isolated meteorological events, because they have the potential to interact with each other – either directly or indirectly through feedbacks onto the background flow.

The PhD project aims firstly to assess the predictability of extratropical cyclone clusters by comparing global ECMWF ensemble forecasts with reanalysis data. The project will then focus on their interaction and feedback mechanisms, which are often underestimated in models. The recent introduction of a 100-member extended range ensemble at ECMWF provides an opportunity to test the hypothesis that model inadequacies in representing these feedbacks weaken responses to local and remote processes, ultimately limiting predictability. Using data from the ensemble will enable discernment of differences between well-performing and underperforming members. Additionally, the research will evaluate model errors in a cyclone-centred framework, offering an innovative approach to identifying poorly represented physical processes in the forecast model. There is also potential to evaluate forecasts against observations taken by aircraft during the January 2026 NAWDIC field

campaign which will focus on mid-latitude atmospheric dynamics and their relation to high impact weather in the North Atlantic region.

Predicting harmful algal toxins in the ocean through satellite data assimilation into a novel biogeochemical model

Supervisor: Professor Shovonlal Roy

This project will develop a novel modelling framework to address the challenges of understanding and predicting harmful algal blooms (HABs) and the toxins they release in the marine ecosystems. HABs disrupt vital ecosystem services, leading to substantial revenue losses for fisheries and aquaculture productions, and pose serious human health risks through the consumption of contaminated seafood. This development is needed to better understand the unpredictable behaviour of HAB forming algal species in response to environmental changes and to simulate algal toxins at regional and global scales. The project will leverage two parallel research strands currently being developed by the supervisors from UoR and MO: (1) a novel biogeochemical model explicitly representing toxin-producing HAB species in the ocean and capable of simulating their harmful dynamics, building on the framework of an intermediate complexity ocean biogeochemical model; and (2) advanced spatio-temporal maps of biological carbon held in marine phytoplankton derived from Earth Observation of ocean colour. By integrating these strands, the student will assimilate advanced EO products into the new biogeochemical model to predict the dynamics of HABs and associated algal toxins in specific regions of the global ocean. The assimilated model output will map the high-risk regions for HABs and algal toxin concentrations, which can help management of the fisheries and aquaculture production in the affected regions.

Understanding (and predicting) tropical cyclone genesis and evolution in the Indian Ocean at coupled km-scale. Project UPTICK

Supervisor: Professor Pier Luigi Vidale

Future projected changes in tropical cyclones (TCs) are uncertain due to our limited understanding of the processes that drive TC precursors, genesis and intensification and how these processes interact with the large-scale environment. Existing literature suggests that, at least in some regions, TC precursor systems evolve from organised convective systems, which are typically unresolved or poorly represented in climate model simulations. TC intensification rates are also strongly influenced by model resolution, but as demonstrated by hurricane Otis, limitations in current modelling predictions can lead to poor forecasts and devastating impacts.

In this project we will use the Met Office's new regional, coupled km-scale modelling capability to answer the research question: What are the processes that drive TC genesis and intensification in the Indian Ocean, how might these change in future and what are the implications for TCs landfalling Africa? Despite their devasting impacts, little is known about how TCs will change in the region. The CP4A (Convection-Permitting 4km-model for Africa) domain will be extended across the Indian Ocean basin to encompass TC genesis regions. With km-scale resolution and atmosphere-ocean coupling, the model should be capable of representing the key processes for TC evolution and hence enable new insights and understanding. Simulations using present-day and future climate conditions will be used to test how these processes and interactions may change in future. These new simulations will complement atmosphere-only CP4A simulations to identify the impacts of coupling on TC

evolution, including potential re-intensification in the Mozambique channel, evidenced recently by Cyclone Freddy.

Large ensembles of machine learning forecasts for advanced nonlinear filters in atmospheric data assimilation

Supervisor: Dr Eviatar Bach

Recently, machine learning (ML) weather forecasting models have shown deterministic forecast skill approaching that of physics-based models, at a small fraction of the computational cost. This provides the opportunity to create very large ensembles of ML forecasts, with the potential to improve data assimilation (DA), the process of optimally combining forecasts and observations to estimate the state of the atmosphere.

Many atmospheric processes are known to be highly nonlinear and non-Gaussian, especially at convective scales. However, DA algorithms currently used in operational centres are only optimal in the linear, Gaussian setting. Advanced nonlinear DA algorithms such as particle filters can estimate the true state of the system without such restrictions, and there is thus great interest in their application to the atmosphere. However, their use has so far been considered unfeasible due to the large ensembles required.

The aim of this project is to test the use of ML ensembles for enabling advanced nonlinear DA methods, following previous work that has shown that a large ML ensemble can augment a small ensemble of physics-based forecasts. This will involve:

- Assessing the probabilistic skill of ML ensembles and investigating how this skill can be improved during training or with transfer learning.
- Developing a particle filter algorithm which incorporates both physics-based and ML forecasts, following previous work on the multi-model ensemble Kalman filter.
- Experiments with this particle filter algorithm and large ML ensembles on models of increasing complexity, ultimately testing in a close-to-operational setting at the ECMWF.