ISDA
Bologna 2023

16th – 20th
October 2023

Bologna
Palazzo dei Congressi
Piazza della Costituzione, 4

INVITED SPEAKERS

Peter Bauer
ECMWF

Jana de Wiljes
University of Potsdam, Germany

Alban Farchi
CEREA, Ecole des Ponts
and EDF R&D, France

Erik Van Vleck
University of Kansas, USA

Xuguang Wang
University of Oklahoma, USA

Milija Zupanski
Cooperative Institute for
Research in the Atmosphere, USA

LOCAL ORGANISING COMMITTEE

Alberto Carrassi
University of Bologna, Italy

Pier Paolo Alberoni
ARPAE, Italy

Carlo Cacciamani
Italia Meteo

Chiara Marsigli
ARPAE, Italy

9th International Symposium on Data Assimilation

ABSTRACT BOOK

INDEX

John Walter Acevedo Valencia,
Deutscher Wetterdienst – Germany

With Walter Acevedo Valencia, Frederik Kurzrock, Sven Ulbrich, Jasmin Vural, Maria Reinhardt, Olivier Liandrat and Roland Potthast

Assimilation of Camera Cloud Motion Vectors (CAM-CMVs) into a Regional NWP system

Nowadays, the skilful forecast of lower tropospheric winds largely depends on the assimilation of high-quality wind observations coming from different instruments such as radiosondes, radar wind profilers and Doppler wind lidars. These observational sources present nonetheless strong limitations, such as the sparse temporal coverage of radiosondes, coming from the small number of launches per day, and the low geographical coverage of the latter two sources, due to the relatively high costs of these devices. To tackle these issues, the German weather service has recently tested for the first time the assimilation of CAM-CMVs into a regional numerical weather prediction system. These ground-based wind retrievals were derived from the combination of an infrared all-sky camera and a ceilometer located at the Lindenberg Meteorological Observatory — Richard-Assmann-Observatory (MOL-RAO) in Germany. For the retrieval, consecutive brightness temperature images were geometry-corrected and fed into an Optical Flow algorithm, generating a set of CMVs. These CMVs were rescaled according to the cloud base height (CBH) measured by the ceilometer and spatially averaged to produce CAM-CMVs. Afterwards, these observations were assimilated into the KENDA (Kilometer Scale Ensemble Data Assimilation) system, which uses the limited-area version of the nonhydrostatic ICON (ICON-LAM) model and the Local Ensemble Transform Kalman Filter (LETKF). In order to assess the potential impact of assimilating these remotely sensed wind observations in a wind-data poor area, we define an experimental setup with a 100x100 km wide box centered on MOL-RAO, where all upper air wind observations were deactivated except our CAM-CMVs. The experiments show that, provided an exhaustive quality control, CAM-CMVs can constitute an innovative complementary observational source of lower tropospheric winds and their assimilation can considerably improve the forecast of the wind field as well as temperature and humidity for lead times up to 24 hours.

Mario Adani,
ENEA – Italy

With Francesco Uboldi

Data assimilation experiments over Europe with the Chemical Transport Model FARM

An Optimal Interpolation scheme has been implemented for the Chemical Transport Model FARM, core of the MINNI modelling system, operational within the regional Copernicus Atmospheric Monitoring Service. The scheme assimilates regulated pollutant concentration observations from ground stations. Two assimilation experiments are validated with independent observations, and compared with CAMS Validated ReAnalysis ensemble for year 2018 over the European domain. Results show that both assimilation experiments are consistent with the VRA ensemble for all pollutants. When the Spatial Consistency Test is used to selectively prevent the assimilation of “bad” information, results are greatly improved, comparing well with the ensemble median. The operational products of the assimilation scheme are continuously evaluated in the framework of CAMS.

Kelbch Alexander,
Deutscher Wetterdienst – Germany

Ensemble-based regional reanalysis system for Central Europe: ICON model setup, test experiments and outlook

The development of regional reanalyses aims at the provision of high-resolution data sets that are suitable for climate applications and climate services. As the desired high-resolution information can barely be provided by observation data, a growing interest in high-quality regional reanalyses is recognisable. Particular demand arises from the renewable energy sector. Further quality gains are expected by using an ensemble approach as the desired uncertainty can be provided. Our project aims to develop and evaluate an operational ensemble-based regional reanalysis system incorporating the current NWP model of DWD (ICON).
For our new reanalysis system ERA5 observation data will be assimilated by using DWD's operational LETKF implementation KENDA. We present the Basic Cycling Environment (BACY) – our testing environment at DWD. BACY consists of shell scripts and is characterized as modular, robust and highly versatile. With BACY, test experiments have been performed to 1) assess the specifications for our future reanalysis system and 2) evaluate the performance of the observation data assimilation. We also present the resulting set of "frozen" BACY specifications such as model resolution, number of ensemble members, etc. as well as results of the assimilation experiments.

Brian Ancell,
Texas Tech University – USA

Ensemble Sensitivity-Based Subsetting: Progress Toward Operational Use

Ensemble sensitivity is a statistical tool applied within an ensemble that reveals the atmospheric flow features related to a chosen forecast response later in the forecast window. Using these relationships, a subset of members with the smallest errors in sensitive regions can be chosen that could improve probabilistic forecasts of the response relative to the full ensemble. Similar to ensemble data assimilation, this process incorporates observational information to beneficially update forecast distributions.
Ensemble sensitivity-based subsetting has been tested in both an idealized framework and in more operational settings during several years of the NOAA Hazardous Weather Testbed. Here we discuss this effort and its associated results, the technique's current status, and future plans toward ultimate operational implementation.

Andrea Antonini,
LaMMA Consortium – Italy

With L. Fibbi, M. Viti, L. Rovai, S. Melani, A. Ortolani

A ship-based ZPD measurement network over the North-West Mediterranean Sea for improving numerical weather prediction forecasts

An operational network infrastructure of GNSS-meteo receivers has been implemented in the recent years for measuring ZPD from ships travelling mainly between the Italyn and French regions of the upper Tyrrhenian and Ligurian Sea. Some further measurements are available along the Civitavecchia-Barcelona route. Data gathered from each system (GNSS receiver + weather station) are hourly sent to a remote ground processing center, where they are processed through a Precise Point Positioning software (MG-APP). In this work a preliminary validation and some applications of the ZPD product are described. The results of the comparison at synoptic times with the parameters derived from the atmospheric reanalysis model MERRA-2 (Modern-Era Retrospective Analysis for Research and Applications, Version 2), point out the high accuracy of such measurements. As a further step, the observations of that distributed measuring system are ingested into a numerical weather prediction (NWP) pre-operational chain, exploiting high update rate of a 1-hourly cycling in a Rapid Update Cycle (RUC). The added value of such assimilation process is preliminarily evaluated over the Tuscany regional area.

Rossella Arcucci,
Imperial College London – United Kingdom

Data Learning 2.0: integrating Data Assimilation with Machine Learning to deal with limitations in models and data

Computational modelling of real-world systems is intrinsically limited by the many types of imperfections encountered when training systems on real world data. Machine Learning (ML) for physical systems is the approximation of the system based on this data and involves the prediction of physical attributes and dynamics from this data. Data Assimilation (DA) is the correction of the approximation of a physical system by the integration of observations with a dynamic model. This allows for more accurate temporal models to be developed, by correcting the temporal slip and non-physical predictions made by these models. We outline a range of methods which are currently being implemented involving ML and DA and discuss how these mitigate the various problems associated with training models with data. From this, we demonstrate that a variety of methods are currently being used in the field of Data Learning and illustrate improved results in a spread of different real-world applications. We offer a guide to how these methods may be implemented to deal with certain types and limitations in data.

Eviatar Bach,
California Institute of Technology – USA

Filtering the Fokker—Planck equation: Theory and an ensemble method

Data assimilation is the problem of correcting a dynamical system's trajectory from observations. Here we consider instead the problem of correcting a dynamical system's statistical properties, as given by expectations with respect to its time-evolving probability density. This problem can be stated as a Kalman—Bucy (KB) filter applied to the Fokker—Planck equation, with the state space being an L^2 space of probability densities. We propose a finite-dimensional ansatz which matches the KB filter under certain assumptions, and then proceed to find an ensemble method approximating this ansatz using a mean field framework. Numerical experiments show that the ensemble method, the ensemble Fokker—Planck filter (EnFPF), is able to accelerate convergence of a dynamical system to its invariant distribution. We discuss possible applications to climate problems and turbulence modelling.

Lilo Bach,
Geoinformation Centre Bundeswehr,
Deutscher Wetterdienst – Germany

Operational Data Assimilation of SEVIRI Imager Data in Short-Range Forecasting

As of March 2023, visible reflectances from the SEVIRI imager on board MSG are assimilated in the operational limited area model (ICON-D2) at DWD. Visible reflectances are sensitive to the liquid water path, the formation of sub-grid scale clouds and the effective radius of cloud droplets. Modeling these quantities in NWP systems is generally prone to systematic biases which poses challenges in data assimilation. Most notably, our development discovered the entanglement between biases in the sub-grid scale diagnostics and the quality of precipitation forecasts which finally lead to advancements in the ICON physics. Our contribution focuses on the technical and methodological steps undertaken to achieve an overall positive impact of this data on the forecast skill of screen-level observations, radiation and cloud cover in the regional data assimilation system at DWD (KENDA) including among others data reduction, quality control and error modeling.

Peter Bauer,
ECMWF

Destination Earth, digital twins, data assimilation, deep learning

Destination Earth is a European Commission funded flagship activity that develops a new type of information system in support of decision making for climate change adaptation and weather extremes management. The information system builds on the concepts of digital twins known from engineering, which use a digital representation of the real world that is dynamically updated with information from the real world but also feeds information back to the real world to drive actions. For the Earth system, data assimilation methods are an integral part of digital twins but require substantial upgrades in terms of the links to impact areas of weather and climate and in terms of interactivity. Machine learning emerges as a powerful option for accelerating computing and dealing efficiently with many of the known challenges in numerical modelling and data assimilation, which will be even more effective for digital twins. Destination Earth and the momentum developed by big tech companies for machine learning will help implement new ways of creating and interacting with environmental information, but also promise to change the ways operational numerical weather and climate prediction are produced today.

Maxime Beauchamp,
IMT Atlantique - France

Towards a stochastic formulation of neural variational scheme for learning SPDE-based priors and solvers

Over the last few years, learning-based methods have become popular to address important challenges related to the recontruction, forecasting and uncertainty quantification of geophysical variables only available through partial and noisy observations. Among these new approaches, 4DVarnet, an end-to-end neural scheme backboned on a variational data assimilation formulation, appears to be successful in numerous applications. In this framework, both prior models and solvers are jointly learnt to estimate the mean state. In a recent extension of this work, we proposed a first ensemble-based version of this algorithm, denoted as En4DVarNet, that combines both the efficiency of 4DVarNet in terms of computational cost and validation performance with a fast and memory-saving Monte-Carlo based post-processing of the reconstruction, which leads to the estimation of the posterior state pdf. This preliminary setup involved a sampling of independent realizations of the state made among a catalogue of model-based dataset used during the training.
Here, we propose a more general approach: rather than drawing samples in some historical dataset, we propose to formulate the prior as the realization of a Gaussian Process (GP) driven by stochastic partial differential equations with interpretable terms related to advection and diffusion components. Applications are provided on a GP toy model with non-stationary diffusion and on a case study involving an idealized Observation System Simulation Experiment (OSSE) of Sea Surface Height (SSH). This paves the way to stochastic 4DVarNet driven by generative modelling and potential links with full diffusion-based neural models.

Florian Beiser,
Mathematics and Cybernetics, SINTEF Digital and Department of Mathematical Sciences, NTNU – Norway

With K. O. Lye, H. H. Holm

Rank Histogram Estimators for Multi-level Data Assimilation

Multi-level Monte Carlo methods have been established as a tool to obtain uncertainty quantification with lower computational cost at the same accuracy than single-level Monte Carlo. Lately, there has also been an effort to use the same ideas to facilitate multi-level data assimilation. For classical single-level ensemble-based data assimilation, rank histograms have been a standard tool to assess the spread in the posterior ensemble. Since multi-level estimators are based on a standard ensemble estimator on a coarse level, combined with a telescopic sum of differences between identically sampled realisations from increasingly finer level, the concept of a multi-level "rank" is not straight forward. In this talk, we present and discuss various estimators to evaluate rank histograms for multi-level ensembles. We demonstrate the estimators for a test case using the multi-level ensemble Kalman method and a rotating shallow water model.

Michele Bendoni,
CNR ISMAR – Italy

The impact of 4D-Var data assimilation of HF-Radar and SST observations on the surface circulation of the North-Western Mediterranean Sea

We employed a ROMS 4D-Var system to investigate the effect of assimilating surface currents from two HF-Radars (HFR) and satellite SST in a circulation model for the North-Western Mediterranean Sea, on three months period, during fall 2020. The data assimilation (DA) framework consists of a sequence of 3 days windows, during which observations are assimilated (AN), followed by 3 days forecasts (FC). The comparison of observed against simulated velocities from AN and FC shows error reduction and correlation increase with respect to the freerun (FR). Surface velocities from Lagrangian drifters are used to test the impact of DA against independent observations. Within the HFRs area, we obtained an improvement of the FC and AN runs with respect to the FR in terms of error and correlation, whereas outside the HFRs area both improvements and degradation of the solution are present.

Loïk Berre,
Météo - France / CNRS - France

With V. Vogt, E. Arbogast, L. Berre

A 3DEnVar scheme for the operational convective scale NWP system Arome-France

After several years of development, a 3DEnVar version of the assimilation scheme for the Météo-France convective scale model Arome-France is now available for operations to replace the 3dvar scheme used historically. This scheme is based on background error covariances that are fully deduced from the forecast perturbations resulting from an ensemble data assimilation and are therefore completely flow dependent. The first part of this study will describe the different components of the data assimilation system itself and the methodology used to tune them (localization lengthscale, inflation, IAU, ...). The second part will be dedicated to the evaluation of its performances compared to the current operational system during long periods and various meteorological issues (winter storm, fog and convective events, ...). The clear improvements shown allow the introduction of this major change in the test suite that is currently being deployed at MF and thus, possibly in the next upgrade of the MF operational NWP system.

Loïk Berre,
Météo-France / CNRS - France

Formulation and use of 3D-hybrid and 4D-hybrid ensemble covariances in the Météo-France global data assimilation system

The global Data Assimilation (DA) system at Météo-France is currently based on a 4D-Var formulation relying on wavelet-based 3D background error covariances. These covariances are specified at the beginning of the DA window and are implicitly evolved in the DA window through tangent linear and adjoint model integrations. Further research and development steps at Météo-France on data assimilation are conducted in the framework of the Object-Oriented Prediction System (OOPS), which is developed in collaboration with ECMWF.

For instance, 3D background error covariances can be made hybrid through a linear combination between wavelet-based covariances and ensemble-based covariances that are filtered through spatial localisation. This allows covariances to be made more anisotropic in a flow-dependent way, and implementation of this hybridation in the OOPS framework is shown to have general positive impacts on the forecast quality.

This 3D-hybrid approach can also be extended to a 4D-hybrid approach in the OOPS framework : this relies on a linear combination between 4D ensemble covariances on the one hand and 4D linearly propagated covariances on the other hand, corresponding to initial covariances that are more explicitly evolved by tangent linear and adjoint versions of the model. This provides a unifying framework for implementations of DA schemes that correspond to 4DEnVar, 4D-Var and new 4D-hybrid formulations. This is thus considered as a novel way to combine respective features of 4D-Var and 4DEnVar approaches. This leads in particular to a 4D-hybrid formulation of 4DEnVar, whose properties are discussed and whose preliminary implementation in the OOPS framework is presented.

Diksha Bhandari,
University of Potsdam – Germany

Affine Invariant Ensemble Transform Methods to improve predictive uncertainty in ReLU networks

We investigate the effectiveness of Ensemble Kalman Filter (EnKF) based algorithms as methods of Bayesian approximation to improve predictive uncertainty in ReLU networks. Our formulation relies on the application of variants of the Ensemble Kalman-Bucy Filter (EnKBF) and Second-order moment matching methods for Bayesian inference. We examine the application of these ensemble transform methods for derivative-free Bayesian inference in ReLU networks. We demonstrate experimentally that a Kalman-based bayesian approximation over only the network's output layer weights yields improved performance and uncertainty estimates on out-of-distribution data. Our work shows that combining these EnKF based techniques for Bayesian inference with a deterministic ReLU network fixes the problem of arbitrarily overconfident predictions of the network on OOD data, which demonstrates the benefits of partially stochastic networks, in particular Last-layer Bayesian Neural Networks for classification.

Marc Bocquet,
École des Ponts ParisTech – France

With Pierre J. Vanderbecken, Alban Farchi,
Joffrey Dumont Le Brazidec, and Yelva Roustan

Bridging traditional data assimilation and optimal transport

Because optimal transport acts as displacement interpolation in space rather than as interpolation in physical variable space, it can potentially avoid double penalty errors. As such it provides a very attractive metric for physical fields comparison — the Wasserstein distance — which could further be used in data assimilation for the geosciences. The algorithmic and numerical implementations of such distance are however not straightforward. Moreover, a theoretical formulation within typical data assimilation problems is quite contrived. This led to a very limited number of contributions to data assimilation. Here, we pursue this endeavour.

We formulate the problem differently than in previous attempts, and we are able to offer a unified view on both traditional data assimilation and optimal transport. We show that thanks to the McCann interpolant, the optimisation can be reformulated as a simpler two—step problem. In particular, the formalism is very flexible and can account for sparse observations, and non-Gaussian error statistics. Thanks to these theoretical developments, we are able to extend — hopefully elegantly — the traditional 3D-Var/BLUE paradigm at the core of most traditional data assimilation schemes. This is illustrated by simple one— and two—dimensional examples but that show the wealth of new analysis types offered by this unification.

Jelena Bojarova,
Swedish Meteorological and Hydrological Institute – Sweden

Study of the Adjustment Processes and Error propagation in the HybridEnVar Framework in the HARMONIE-Atrome Forecasting System

The optimal length of the data assimilation window is dependent on the predictability scales of the phenomena one wants to model and adjustment scales of the processes involved. Ideally the length of the data assimilation window should be short enough to stay within the predictability limits of the phenomena in order that control of the initial conditions could be possible. At the same time the data assimilation window should to be long enough to allow information from observed qualities to propagate through the model state. Clouds visualise important signatures of convective scale processes. Better understanding of the mechanisms behind adjustment processes on convective scales is required in order to design powerful data assimilation schemes able to efficiently extract an information from cloud measurements. We present results of the sensitivity study where error evolution through the model state components is studied in controlled way in the ensemble environment. We use Hybrid Ensemble Variational Data Assimilation Framework implemented in the HARMONIE-Arome Forecasting System to generate flow-dependent disturbances of flow, mass and moist variables and evaluate their propagation through the time decomposed on the range of scales. The conclusions help to guide the design of the data assimilation system.

Massimo Bonavita,
ECMWF – United Kingdom

Hybrid NWP-Machine Learning or End-to-End Learning?

Machine Learning has proved to be an innovative, disruptive set of technologies whose deployment has revolutionised best practices in various fields of applied science and engineering. ECMWF has started early on the path to introducing ML components in its NWP value chain, with some good results already close to operational implementation. However the speed of evolution of ML applications in NWP/Climate prediction and the appearance on the scene of the big technology companies have caused waves in the previously tranquil NWP/Climate ecosystem and posed some challenging questions on the sustainability of current operational practices in NWP and Climate.

In this talk I will first give an overview of current ECMWF efforts in hybridising our NWP workflow with ML technologies with an emphasis on data assimilation applications. In the second part, I will discuss the potential and limitations of End-to-End Learning of the NWP value chain by looking in more detail at some recent examples of “Earth system simulators” trained on publicly available re-analysis datasets like ECMWF ERA5.

Nicolas Bousserez,
ECMWF – Germany

With Luca Cantarello, Panagiotis Kountouris, Ernest Koffi, Auke Visser, Anna Agusti-Panareda, Massimo Bonavita and Richard Engelen and the CAMS team

A Hybrid Ensemble-Variational Approach for Greenhouse Gas Source Inversions in the Integrated Forecasting System (IFS)

The European Commission has entrusted the European Centre for Medium-range Weather Forecasts (ECMWF) with building an operational global Monitoring and Verification System (MVS) for anthropogenic CO2 and CH4 emissions as part of Copernicus Atmosphere Monitoring Service (CAMS). Research activities to develop this capacity have been carried out by a large consortium of partners under the EU-funded CHE and CoCO2 projects. The consortium will deliver the prototype systems at the required spatial and temporal scales, covering local hotspots to regional and global emission budgets. In this presentation we will describe data assimilation developments at ECMWF to build the global component of the prototype MVS system, which is based on the Integrated Forecasting System (IFS). A key aspect of the methodology is the extension of the current short-window (12-hour) operational 4D-Var analysis to accommodate the optimisation of emissions of long-lived greenhouse gases, such as CO2 and CH4. Efforts to implement a hybrid approach combining ensemble information with the adjoint-based 4D-Var algorithm to build a long-window (order of several weeks) 4D-Var will be outlined. Preliminary results based on case studies of CH4 and CO2 inversions will demonstrate the impact of the long-window 4D-Var inversion method. The limitations of the system as well as possible improvements will be discussed in the context of the long-term developments planned towards the implementation of the operational Copernicus service.

Matteo Broccoli,
Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC) – Italy

Towards an Observation Operator for Satellite Retrievals of Sea Surface Temperature with Convolutional Neural Network

Global ocean numerical simulations typically work with a vertical subsurface resolution of about 0.5m. However, key physical quantities like sea surface temperature (SST) can be retrieved from satellites at a reference depth of a few microns or millimeters below the sea surface. Assimilating such temperatures can lead to bias in the ocean models and it is thus necessary to project the satellite retrievals to the first model level to safely use them in the assimilation process. This projection is non-trivial, since it depends on several factors (e.g., daily cycle, winds, latitude) and it is usually performed either with complex numerical methods or with too simple statistical methods. Here we present an attempt to construct the projection operator with deep learning techniques. We consider a convolutional neural network architecture called U-Net, which was first introduced in the field of computer vision and image segmentation, and it is thus optimal to process satellite retrievals. We train it with subskin SST from satellites and wind speeds from atmospheric analysis to reproduce the SST at 1m and foundation level given by reference models. We also take into account the daily cycle by training the U-Net separately on day- and night-time retrievals. This U-Net can then be used to shape an observation operator to be employed in CMCC’s OceanVar assimilation system.

Jochen Broecker,
University of Reading – United Kingdom

With Tobias Kuna, Lea Oljaca

Exponential stability of the optimal filter for signal processes with hyperbolic dynamics

The optimal filter is a sequence of probability measures which describe the conditional probability of some stochastic process (called the signal process), given the history of another process (called observation process). In typical applications the optimal filter satisfies a dynamical equation, and we investigate the stability of these dynamics with respect to misspecification of the initial distribution. In the present work, the signal process is given by the iterations of a deterministic mapping (uniformly hyperbolic), with only the initial condition being random. Hence we cannot rely on strong randomness of the signal processes, and more dynamical effects (i.e. the chaotic behaviour) of the signal process will need to be exploited.

Mareike Burba,
Deutscher Wetterdienst – Germany

DWD's CONTRAILS project: variational assimilation of satellite reflectances

The German Meteorological Service (DWD) aims to develop a framework that allows assimilating satellite measured reflectances in the global EnVAR system, after reflectances became operational in the regional LETKF system in 2023. In addition to the RTTOV clear-sky input parameters for the infrared or microwave regime, the reflectance observation operator (MFASIS) requires cloud cover, cloud water and cloud ice concentration information (allsky variables). Traditionally, the EnVAR in DWD's DA code (DACE) minimizes virtual temperature, generalized humidity, wind and geopotential. We plan to implement three different versions how the conventional minimization variables Jacobians can be updated including the information from allsky variables:

1. tangent linear model of cloud parametrizations

2. Neuronal network learning cloud parametrizations

3. ensemble covariances similar to an LETKF update

The CONTRAILS project aims to improve the physical understanding, identification, characterization, assimilation and forecast of contrails. The CONTRAILS project partners from research and industry work on ML and/or observational aspects.

Michele Buzzicotti,
University of Rome Tor Vergata and INFN – Italy

Data-Driven Tools: From Ideal models to Geophysical systems

Our ability to collect data is growing rapidly, thanks to computing power and an unprecedented variety of sensors. But big data is not easy to digest. By performing in silico experiments on ideal/complex systems, we investigate the applicability of data-driven tools to extract interpretable information from sparse & heterogeneous observations of complex flows. Key applications are data-driven physics modelling, data augmentation and super-resolution. We compare state-of-the-art purely data-driven tools, such as Generative Adversarial Networks (GAN) [1,2], with purely physics-informed approaches, such as nudging, on the ability to reconstruct missing data. At the same time, we will discuss how the potential of both techniques is combined in the most advanced physics-informed neural networks. Starting from ideal cases such as turbulence on a rotating frame or Rayleigh-Benard convection [3], I will discuss how these tools can potentially lead to a methodological revolution of today's data analysis and data reconstruction techniques in applications such as ocean observations [4,5].

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 882340).

[1] M. Buzzicotti, F. Bonaccorso, P. Clark Di Leoni, L. Biferale. Reconstruc- tion of turbulent data with deep generative models for semantic inpainting from TURB- Rot database. Physical Review Fluids 6 (5), 050503, (2021).

[2] Buzzicotti, Michele. "Data reconstruction for complex flows using AI: recent progress, obstacles, and perspectives." Europhysics Letters (2023).

[3] P. Clark Di Leoni, L. Agasthya, M. Buzzicotti & L. Biferale. (2023). Reconstructing Rayleigh-Benard flows out of temperature-only measurements using Physics-Informed Neural Networks. European Physical Journal E 46(3),16

[4] B.A. Storer, M. Buzzicotti, H. Khatri, S.M. Griffies, H. Aluie. Global energy spectrum of the general oceanic circulation. Nature communications 13 (1), 1-9, (2022).

[5] M. Buzzicotti, B.A. Storer, H. Khatri, S.M. Griffies, H. Aluie. Spatio-temporal coarse-graining decomposition of the global ocean geostrophic kinetic energy. arXiv:2106.04157, (2021).

Edoardo Calvello,
California Institute of Technology – USA

The Mean-Field Ensemble Kalman Filter: Gaussian and Particle Approximations

The ensemble Kalman methodology is an innovative and flexible set of tools which can be used for both state estimation in dynamical systems and parameter estimation for generic inverse problems. It has primarily been developed by practitioners in the geophysical sciences, yet, despite its widespread adoption in fields of application, firm theoretical foundations are only now starting to emerge. We consider a unifying approach to algorithms that rests on transport of measures and mean-field stochastic dynamical systems. The ensemble Kalman methods as implemented in practice rely on projections onto the space of Gaussian measures and particle approximations. With the goal of developing theoretical guarantees for the ensemble Kalman methodology applied to non-linear problems, we discuss the error analysis and ergodicity of the mean-field stochastic dynamical systems arising in ensemble Kalman filtering, along with Gaussian and particle approximations.

Luca Cantarello,
ECMWF – Germany

With N. Bousserez, P. Kountouris, E. Koffi, A. Visser, A. Agusti-Panareda, M. Bonavita, R. Engelen and the CAMS team

Assessing the atmospheric flux inversion system at ECMWF using Observation System Simulation Experiments

As part of the implementation of the global component of the new Copernicus Monitoring and Verification System for anthropogenic CO2 emissions (CO2MVS) and the related activities of the CoCO2 project, the European Centre for Medium-Range Weather Forecasts (ECMWF) has expanded its Ensemble of 4D-Var Data Assimilation algorithm (EDA) to include a new flux inversion system for atmospheric composition species. Moreover, in order to improve the assimilation of long-lived species such as greenhouse gases (CO2 and CH4), a new long-window (i.e., order of weeks) assimilation approach is being implemented.

To evaluate and validate the new flux inversion capabilities within the Integrated Forecasting System (IFS), the EDA has been used to develop an Observing System Simulation Experiment (OSSE) setup, in which the control (unperturbed) member is treated as 'nature run' from which synthetic observations are generated, perturbed, and subsequently assimilated by the ensemble members.

In this presentation we will show the results from a series of OSSEs, in which a number of system components are tested in the form of sensitivity experiments. In particular, we explore: the impact of variance and correlations in the prior fluxes covariance matrix (B); the impact of model bias (e.g. systematic transport error); the preliminary implementation of a new tuning approach for the B matrix.

Elizabeth Carlson,
California Institute of Technology – USA

With Adam Larios, Edriss S. Titi

Continuous Data Assimilation: A Nonlinear Algorithm & Connection to Physics

One of the fundamental challenges of accurate simulation of turbulent flows is that initial data is often incomplete, which for turbulent fluid flows is a strong impediment to accurate modeling due to sensitive dependence on initial conditions. A continuous data assimilation method proposed by Azouani, Olson, and Titi in 2014 introduced a feedback control term to dissipative systems, proving directly how the maximal effectiveness of DA is dependent on the specific physics of the system being investigated. In this talk, we will focus on the insights of a nonlinear version of the AOT algorithm and distinguish the clear connections to the physics of the existing systems.

Jean-Francois Caron,
Environment and Climate Change – Canada

Farewell hybrid covariances: Moving to fully ensemble-derived background-error covariances for NWP at Environment and Climate Change Canada

Since the operational implementation of 4D-ensemble-variational (4D-EnVar) data assimilation at ECCC in 2014, background-error covariances have been represented by a blend of a time-invariant (also called static or climatological) component and a flow-dependent ensemble-derived component. Such hybridization of covariances is a common approach to compensate for deficiencies in the ensemble-derived estimate.

We found that improving the ensemble-derived covariance localization approach can eliminate any benefit from including the static covariances. The localization improvement was achieved by implementing the approach of applying different amounts of horizontal localization to different ranges of background-error covariance horizontal scales as proposed by Buehner and Shlyaeva (2015).

Thus, in December 2021, our global deterministic forecasting system began to rely completely on flow-dependent covariances in the troposphere and the lower stratosphere. In 2024, after revisiting the scaling of the ensemble-derived covariances near the model top, static covariances will longer be used in both the deterministic and ensemble forecasting systems relying on 4D-EnVar for NWP at ECCC.

The experiments that led to the adoption of these changes and their impacts on the forecasts will be presented. It will also be illustrated that relying more on ensemble-derived covariances amplifies the positive impacts when the background ensemble generation strategy is improved.

Diego Saul Carrió Carrió,
Universitat de les Illes Balears – Spain

A New Localization Method for Non-Gaussian Variations of the EnKF: "The Ensemble Squeeze Localization"

In Data Assimilation (DA) it is well-known the relevant importance of localization techniques to effectively ameliorate spurious long-range correlations between the observations and the prior model variables, which are mainly associated with having insufficient ensemble samples. Here we highlight the Observation-Error-Variance Inflation, which is widely used operationally in Local Ensemble Transform Kalman Filter (LETKF) applications to effectively perform real DA. However, it turns out that this localization technique is ill-suited to non-Gaussian variations of the EnKF, such as the Gamma- Inverse-Gamma (GIG) variation for semi-positive-definite variables. In this situation, the impact of assimilating observations from skew distributions does not go to zero as the observation error variance goes to infinity, as it does in Gaussian DA. For effective localization, one needs the impact of the observation on a central analysis point to go to zero as the distance of the observation from that central point goes to infinity. Bearing in mind the widely used of the LETKF, we have developed a GIG version that fits within the LETKF framework, with minimum changes in the code. Then, we introduced a new localization method called the Ensemble Squeeze Localization, that deals with the above-mentioned localization problem by attenuating the ensemble perturbations amplitudes by an observation-error deflation factor. We will show that the Ensemble Squeeze Localization is precisely equivalent to the Observation-Error-Variance Inflation localization through a series of examples using the LETKF applied to a 2-dimensional pseudo Tropical-like cyclone.

Man-Yau (Joseph) Chan,
National Center for Atmospheric Research – USA

Mitigating sampling errors in ensemble data assimilation with Probit-space Ensemble Size Expansion (PESE)

In this talk, I will introduce a method to suppress sampling errors in ensemble data assimilation (EnsDA): the Probit-space Ensemble Size Expansion (PESE). PESE suppresses these errors by generating additional ensemble members through utilizing the users' knowledge of the marginal distributions of forecast variables. Unlike methods that draw additional members from climatology, PESE-generated members have flow-dependent statistics. Preliminary results with the Lorenz 1996 model indicate that using PESE with the Ensemble Adjustment Kalman Filter (EAKF) improves the assimilation of nonlinear observations (signed square-root). Tests with other observation operators and EnsDA filters will also be presented in this talk. Since PESE is computationally efficient and highly scalable, PESE can be utilized with large models in the future.

Yumeng Chen,
University of Reading – United Kingdom

Multivariate state and parameter estimation with data assimilation on sea-ice models using a Maxwell-Elasto-Brittle rheology

The Arctic sea ice is an essential component in the climate systems. Under the context of climate change, the Arctic sea ice requires better initialisation of its state and improved numerical models. Data assimilation (DA) approaches are used operationally to improve the model state by combining information from the model forecast and observations. These approaches can also estimate unobserved model fields and parameters which can be viewed as a data-driven technique for better model representation of the sea ice. In this study, we investigate the fully multivariate state and parameter estimation in a simplified and idealised environment using a novel sea ice rheology, Maxwell-Elasto-Brittle (MEB) where we estimate not only the sea ice concentration, thickness and velocity but also the sea ice damage, stress and cohesion. Specifically, we estimate the air drag coefficient and damage parameter in the MEB model. Mimicking the realistic observation network with different combinations of observation fields, we demonstrate that various issues can potentially arise in complex sea ice models, especially in the cases where the external forcing dominates the model forecast error growth. Even though further investigation will be needed using an operational sea ice model, we show that, with the current observation network, it is possible to improve both the observed and unobserved model state forecast and parameter accuracy.

Ying-Jhen Chen,
Central Weather Bureau – Taiwan

With Guo-Yuan Lien

Comparing the Partial Cycling and Continuous Cycling Data Assimilation Strategies in a High-Resolution Regional Forecast System

The Central Weather Bureau (CWB) of Taiwan operates a regional numerical weather prediction system, named CWBWRF, which is developed based on the Advanced Research Weather Research and Forecasting (WRF-ARW) model and the WRF Data Assimilation (WRFDA). The CWBWRF is configured with 15/3-km resolution nested grids and provides deterministic forecasts for up to 126 hours. The 6-hourly initial conditions are generated using a so-called partial cycling (PC) strategy, by which each run is independently initialized from the NCEP Global Forecast System (GFS) analysis at the previous 12 hours, followed by twice of 6-hour forecast-analysis cycles, and therefore it obtains the initial condition of the 126-h forecast. Hybrid three-dimensional ensemble-variational data assimilation (3DEnVar) using the WRFDA coupled with an Ensemble Adjustment Kalman Filter (EAKF) is conducted in the partial cycling process. In addition, after each cycle's analysis, a "blending" technique is used to replace the large-scale characteristics of the WRFDA regional analysis with the global model analysis, while keeping the small scales from the regional data assimilation. The data assimilation and initialization workflow described above can take advantage of the high-quality global model analysis and still maintain some mesoscale information, which has been shown quite successful at CWB to operationally produce skillful regional forecast guidance.

However, the current data assimilation workflow also has weakness and limitations. The regional DA effect does not have sufficient time to accumulate and is largely obstructed by the global model analysis. To seek a possible alternative workflow, we test a continuous cycling (CC) data assimilation that also equips a similar blending step but it is done only before each 126-h forecast, outside the continuous cycles. It is noted that the same cycling strategy was also proposed and studied by Schwartz et al. (2021, WAF) which demonstrated the benefits of this method in their regional ensemble forecast system. In our study, we show the comparison of the forecast performances of the CWBWRF using the abovementioned PC and CC strategies and discuss their advantages and disadvantages in various aspects of the regional forecast.

Sandy Chkeir,
University of Vienna – Austria

With Martin Weissmann, Philipp Griewank, Florian Meier, Adhithiyan Neduncheran

Preparing AROME assimilation experiments for cloud-affected satellite observations

The direct assimilation of visible and infrared (IR) radiances is a promising on-going research to make the best use of convection-permitting weather models in cloud-affected areas. In this work, we explore the Radiative Transfer for TOVS (RTTOV) operator for simulating the water vapor IR (6.2 µm & 7.3 µm) and the visible (0.6 µm) seviri channels in cloudy conditions in preparation for assimilation experiments with these observations. We try to find the optimal configuration settings to perform our simulations using a convective-scale AROME model. In the initial experiment, we try to test the IR cloudy simulations for a selected convective case over Austria.

Marcin Chrust,
ECMWF – United Kingdom

With A. Farchi, M. Bonavita, P. Laloyaux and M. Bocquet

Hybrid Data Assimilation – Machine Learning for Model Error Estimation and Correction: application to the ECMWF IFS model

In the field of Numerical Weather Prediction (NWP) and more generally in Earth system prediction applications, model error is increasingly viewed as the fundamental obstacle to overcome towards increasing the predictability horizon of current state-of-the-art forecasts and improving their practical utility.
In NWP, a large amount of work has been traditionally devoted to account for the effects of model error on the second order moments of the forecast error distribution, either through sampling and/or explicit modelling of the assumed model error covariances. Less attention has been given to the as prominent issue of estimating and correcting the predictable component of model error, i.e., the model systematic errors. Modern implementations of weak-constraint 4D-Var start to address this issue in a data assimilation (DA) variational framework, with encouraging results. One can view weak-constraint 4D-Var as an online hybrid data assimilation and machine learning system which gradually learns about model errors which are slowly evolving over the length of the assimilation window and uses this knowledge to give improved estimates of the state of the system. The natural question is then if and in what measure we can apply recent machine learning (ML) techniques to further develop the concept of online model error estimation and correction. In this talk we will describe recent progress in developing a hybrid DA/ML system for the ECMWF IFS model. We have supplemented our state-of-the-art physics-based model with a statistical model, implemented via a neural network (NN), that provides flow dependent model error corrections. We have subsequently included the parameters of the model error model in the 4D-Var control vector for joint state-parameter estimation. We will demonstrate that applying flow dependent model error corrections during the medium range forecast has a significant positive impact on the forecast quality.

Adam Clayton,
Korea Institute of Atmospheric Prediction Systems (KIAPS) – South Korea

With I-H Kwon, J-H Kang, K-H Seol

Development of existing and future DA systems at KIAPS

Since April 2020, KMA has been operating a global atmospheric NWP system based on the "Korea Integrated Model" (KIM). Deterministic analyses are provided by a hybrid-4DEnVar system. Initial conditions for a 50-member KIM-based ensemble are updated via an LETKF.

We will explain the impact of a recent improvement to the formulation of the observation operators in the hybrid-4DEnVar system, and our progress in improving the ensemble-based background error covariances, including introduction of a scale-dependent localisation scheme.
We will then explain our progress in developing two new DA systems:

1. A NEMOVAR-based ocean DA system that will be weakly coupled to the existing atmospheric DA system as part of a new Earth System NWP system.
1. An LETKF-based DA system aimed at high-resolution forecasting over Korea, with direct assimilation of radar reflectivities.

Giovanni Conti,
Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC) - Italy

With C. Cardinali

Advanced diagnostic tools to assess the observations impact in the analysis and forecast

Understanding the contribution of the observations to the analysis and the forecast quality is crucial for model improvements and the design of the future global Earth system observation network. Over the last years, advanced diagnostic tools have been developed and used to measure the value of the observations assimilated with respect to the analysis and the forecast quality. In particular, these two measures, the Degree of Freedom for Signal (DFS) and the Forecast Sensitivity to the Observation Impact (FSOI and FSOI-Jo), have been successfully used to diagnose the sub optimality of the assimilation system used and the model and observations deficiencies.

A novel development of the FSOI-Jo tool valid for the ensemble Kalman filter implemented in SPREADS is here presented. FSOI-Jo uses an observation-based objective function (Jo) to compute the short-range forecast error and consequently to compute the observations contribution to reduce such a forecast error. The strictly dependency between the observation impact in the analysis (DFS) and the observation impact in the forecast (FSOI-Jo) is also demonstrated.

Finally, the observation impact in the SPREADS-CAM83L 24 hour forecast error is presented.

Filipe Bitencourt Costa,
REMO - Brazil

With C. A. S. Tanajura

Assessing Impacts of Ensemble Kalman Filter (EnKF) on the Remo Ocean Data Assimilation System (RODAS) Over the South Western Atlantic

This work presents the implementation of the Ensemble Kalman Filter (EnKF) on the REMO Ocean Data Assimilation System (RODAS) with the Hybrid Coordinate Ocean Model (HYCOM) with 1/12° on the western tropical and South Atlantic. The new version of RODAS employs a joint and multivariate assimilation of hydrographic profiles, United Kingdom MetOffice OSTIA Sea Surface Temperature (SST) and AVISO Absolute Dynamic Topography (ADT). Three experiments were performed for six months with assimilation cycle of ten days, (i) Control with no assimilation, (ii) A_EnOI employing Ensemble Optimal Interpolation (EnOI) and (iii) A_EnKF employing EnKF and forced with perturbed atmospheric fields. A_EnKF was successfully implemented as ensemble spread was maintained around 0.35 °C, 0.03 m and 0.05 psu for temperature, Sea Surface Height (SSH) and salinity, respectively. Also, the mean Root Mean Squared Deviation (RMSD) of all ensemble was greater than the RMSD of the mean run for temperature and salinity. The mean correlation of SSH with respect to AVISO was 0.12, 0.33 and 0.31 and the RMSD of SST with respect to OSTIA was 0.92, 0.52 and 0.47 °C for Control, A_EnOI and A_EnKF, respectively. For the subsurface, RMSD with respect to ARGO was 0.22, 0.20 and 0.18 psu for salinity and 1.42, 0.91 and 1.09 °C for temperature for Control, A_EnOI and A_EnKF, respectively. Impacts on the Brazil Current are still been assessed. A_EnOI showed better SSH correlation and smaller temperature error while A_EnKF presented smaller error for SST and salinity. Therefore, A_EnKF shows comparable quality to RODAS previous version. For future works, it is expected with increase in ensemble members, from eleven to thirty, the new version of RODAS should outperform its previous for SSH, SST and subsurface temperature and salinity.

François Counillon,
NERSC/UiB – Norway

With T. Singh, J. Tjiputra

Estimation of Ocean Biogeochemical Parameters in an Earth System Model Using ensemble data assimilation methods

Ocean biogeochemical (BGC) models utilise many poorly-constrained global parameters to mimic unresolved processes and reproduce the observed complex spatiotemporal patterns. Large model errors stem primarily from inaccuracies in these parameters whose optimal values vary in space and time. We aim to demonstrate the ability of ensemble data assimilation methods to provide high-quality and improved BGC parameters within an Earth system model with the Norwegian Climate Prediction Model (NorCPM), which combines the Norwegian Earth System Model and the ensemble data assimilation method. We demonstrate that first in an idealised twin experiment, the dual-one-step-ahead-smoother (DOSA) technique, which can recover the spatially varying parameter values and perform nearly optimally — i.e., produce errors comparable to the model with perfect parameter values. The estimation is quick (1 model year) even with sparse observations. However, in a real framework, the DOSA scheme performs sub-optimally, because state error inherited from the other components (ocean and atmosphere) grows faster than parametric error from the BGC model. We, therefore, train the BGC parameters — global value and spatially varying -- based on the reanalysis performance with the state assimilation in the other components using an ensemble smoother technique. Both global and spatially varying parameters effectively reduce error and also reduce error in non-assimilated quantities (e.g. in air-sea CO_2 flux). While spatially varying parameter performs overall better, it also introduces some large local degradations.

George Craig,
LMU Munich – Germany

The Weak Temperature Gradient approximation as a balance principle for convective-scale data assimilation

Dynamical constraints such as geostrophic and hydrostatic balance are an important component of data assimilation systems for synoptic scales of motion, and a similar balance principle for convective-scale data assimilation would be desirable. The presentation will briefly review the physical basis for balance in mesoscale dynamics, and the problems caused by imbalance in convective-scale data assimilation. Asymptotic analysis is used to show that the WTG approximation is obtained for slow motions on mesoscale length scales, in a manner that is analogous to geostrophic balance on synoptic scales. Unlike geostrophic balance, the WTG approximation includes strong divergent motions driven by diabatic heating and cooling, while still excluding transient gravity waves. Simulations of summertime weather over central Europe at convection-permitting resolution are used to show that WTG balance is approximately satisfied for mesoscale motions, although the approximation is not as accurate as geostrophic balance on synoptic scales. Finally, preliminary results from the ICON-KENDA ensemble data assimilation system of the German Weather Service show that data assimilation increments produce an increased departure from WTG balance that decays over the first half hour of the free forecast.

Sebastiaan (Bas) Crezee,
MeteoSwiss – Switzerland

With B. Crezee, D. Leuenberger, C. Merker, G. Martucci, A. Haefele, and M. Arpagaus

Towards an operational assimilation of RAMAN lidar temperature and mixing ratio profiles with COSMO/KENDA-1

The Raman Lidar for Meteorological Observations (RALMO) is operated at the MeteoSwiss station of Payerne (Switzerland) and provides continuous measurements of temperature and water vapor mixing ratio profiles since the year 2008. The high resolution in time (30 min) and vertical space (30 m) makes these observations very appealing for assimilation into a NWP model.

In this contribution, we show our efforts in improving the robustness and quality of the RALMO observations by comparing them against Payerne radiosonde observations, as well as assimilation experiments with the MeteoSwiss 1-km mesh-size, LETKF-based ensemble data assimilation system COSMO/KENDA 1. In these experiments RALMO observations are assimilated along with its state-dependent estimation of the observation error in order to weight correctly the impact onto the model forecasts.

In a summer experiment, the additional RALMO observations have shown a beneficial impact in the forecasts of high-intensity precipitation and convective clouds, whereas in a winter experiment, we found improvements in the forecasts of fog and low stratus.

Sarah Dance,
University of Reading & National Centre for Earth Observation – United Kingdom

A new computational approach for spatially correlated observation error statistics in data assimilation

Observations with high spatial density are needed to provide information on appropriate scales for convection-permitting weather forecasts. Remote-sensing data are usually thinned to avoid problems with spatial error correlations, so that in some cases only 5% of the observations are used. Explicitly introducing correlated error statistics in the assimilation requires the use of full observation weighting matrices (rather than diagonal), and this can increase the computational costs of the assimilation. We present a new, computationally inexpensive, numerical approximation method, based on a hierarchical data compression using a singular value decomposition (SVD), and a domain localization approach. We applied the approach in a simple variational data assimilation system and found that the computational cost was dramatically reduced while preserving the accuracy of the analysis. This new method has the potential to be used as an efficient technique for practical applications.

Haroldo Fraga de Campos Velho,
INPE: National Institute for Space Research – Brazil

Data Assimilation and Predictability by Machine Learning Approach

Modern numerical weather and climate predictions represent a key scientific advance for society, with very positive impact to many sectors, including prevention and mitigation of natural disasters, agriculture, energy production, tourism and entertainment industries, atmospheric pollutant dispersion, and as an essential tool for the climate change evaluation. Complex physical processes are added to the geophysical dynamical core for producing good prediction. The forecasting is strongly dependent on initial condition. Therefore, better initial condition identification implies in an improved prediction. Data assimilation is a process to combine observations with a previous prediction (background) in the best way to compute the analysis. A kind of Monte Carlo method – ensemble prediction – is the standard way to compute the confidence interval, or predictability (uncertainty quantification), of the calculated prediction.

However, data assimilation and model predictability are computer intensive actions.

The challenge is harder with exponential growth of the observations available, and models with finer computational grid. The talk is to show methods to speed up the data assimilation procedure and the uncertainty quantification addressed by machine learning formulation. Experiments with data assimilation by neural networks are applied to global and limited area atmospheric models. Decision tree is employed to estimate the model predictability for precipitation, but the method can be applied for any meteorological field. The application of machine learning for pre-processing (data assimilation) and post-processing (predictability) can promote a very significant reduction of the computational effort, with similar quality of the analysis and the forecasting confidence interval.

Jana de Wiljes,
Ilmenau University of Technology – Germany

Advances in high dimensional nonlinear filtering – intermediate updates and novel localisation strategies

Extended state and parameter spaces are still an obstacle for consistent particle based systems. Yet the nonlinearity and complexity of the systems of interest require employing methods that do not fully rely on an underlying Gaussianity assumption. Here will propose two approaches that on the one hand increase the accuracy of state-of-the-art methods and on the other hand are designed to improve the computational feasibility.When combined, these strategies hold the potential to significantly advance data assimilation in high-dimensional nonlinear scenarios.

Deepjyoti Deka,
Los Alamos National Laboratory – USA

With Wenting Li

Physics-informed Neural Networks for Localization in Stochastic Power Grids

The growth in distributed energy resources (DERs) is an important step toward solving the global climate crisis. However, many DERs, such as wind and solar power, are intermittent, causing the data in power grids to be perturbed with high uncertainty. Such perturbations degrade the performance of data-driven algorithms introduced in power grids for sensing and control. The performance is further degraded due to low observability and the absence of sufficient labels in historical data.

We propose physics-informed methods to overcome these issues and correct estimation in realistic stochastic power grids. We first develop a novel Physics-Preserved Graph Network (PPGN) that has a unique two-stage architecture, one represents the underlying network using a few measured nodes and another that finds relations between the labeled and unlabeled data samples to further improve the location accuracy. Next, we proposes a novel physics-constrained adversarial training method for robustifying neural networks that uses analytical physical laws satisfied by state perturbations in realistic grids. The numerical experimental in industry-accepted test networks, we validate our physics preserved neural network and adversarial training and show superior performance in accuracy and efficiency, compared to state of the art methods.

Alan Demortier,
Météo – France

Added value of assimilating ground observations from personal weather stations in a convective-scale numerical weather prediction system

The signatures of convective systems can be revealed using dense Personal Weather Station (PWS) surface observations of pressure, temperature and relative humidity. The question here is whether an operational kilometre-scale NWP system such as AROME-France can take advantage of assimilating fine-scale ground information to improve short-range forecasting. This presentation focuses on data pre-processing and discusses the differences obtained with a 3DVar and a 3DEnVar data assimilation system.

Thomas Deppisch,
Deutscher Wetterdienst – Germany

Studying the Interaction between NWP Models and Data Assimilation with Observing System Simulation Experiments — Case Studies with SEVIRI Data in ICON-LAM

All-sky data assimilation of visible reflectances from the SEVIRI instrument (MSG) has recently been operationalized in the limited-area model of DWD. Moreover, data assimilation of thermal radiances at the regional scale and visible radiances at the global scale is currently in the state of development. Implementing these observations into an operational data assimilation system has proven particularly challenging due to the complex information they provide about the atmospheric state, while systematic and conditional biases arising from the interaction of data assimilation and model interactions pose additional problems. In order to study such interactions in a controlled environment, we have developed a setup for Observing System Simulation Experiments (OSSEs) that is close to the operational system used for regional data assimilation system at DWD (KENDA). With these tools at hand, we strive for a better understanding of the atmospheric processes triggered by all-sky assimilation of radiances, especially in the presence of convection.

Mayeul Destouches,
United Kingdom Met Office – United Kingdom

Improving background error covariance estimation with Multilevel Monte Carlo Methods

Multilevel Monte Carlo methods (MLMC) were popularized in the late 2000's in the applied mathematics community. They offer an astute way of combining ensemble forecasts of various resolutions to reduce sampling noise in Monte Carlo estimators, without increasing the computational cost, nor introducing errors from the forecasts at coarser resolution.

We present here how MLMC can be applied to improve the estimation of the background error covariance matrix in ensemble variational data assimilation (DA). The potential of MLMC for ensemble DA is illustrated by experiments with a hierarchy of quasi-geostrophic models.

Care will be taken to explain what distinguish MLMC methods from simple addition of ensemble B matrices at various resolutions. Limitations currently preventing operational implementation and possible workarounds will be discussed.

Theresa Diefenbach,
LMU Munich – Germany

With G. Craig

Measures of imbalance in convective-scale data assimilation

Analyses used as the initial condition for a forecast are not necessarily dynamically consistent with the forecasting model, which manifests itself in imbalance in the model state. Imbalance is associated with gravity wave noise and spurious convective cells, which can deteriorate the forecast with lead time. Therefore, it is important to develop methods to detect and correct for imbalance introduced through data assimilation. While there are effective procedures to reduce imbalance on the synoptic scale, the situation on the convective scale is less clear. Non-linearity as well as non-hydrostaticity play an important role in the formation of convection and there is no clear separation between fast and slowly evolving time scales. Therefore, the development of methods to counteract imbalances in convective-scale data assimilation is subject to current research.

In this study we investigate three different measures of imbalance that are potentially relevant for convective-scale data assimilation:

1. surface pressure tendencies,
2. vertical velocity variance in the vincinity of convective clouds,
3. departures from the vertical velocity that is prescribed by the weak temperature gradient (WTG) approximation.

To test the effectiveness of these measures, we apply them to three different data assimilation algorithms from which we expect different degrees of imbalance in the analysis. The numerical weather prediction system that we use is the ICON-KENDA system of the German Weather Service. Moreover, our experiments involve two different cases with two distinct synoptic situations (weak and strong forcing).
Overall, surface pressure tendencies seem to diagnose a different type of imbalance than the vertical velocity variance and the WTG departure. The latter two measures show very similar results. For reasons of numerical efficiency, we propose the WTG departures over the partitioned vertical velocity variance.

Theresa Diefenbach,
LMU Munich – Germany

With G. Craig, C. Keil, L. Scheck, M. Weissmann

How to use partial analysis increments in an LETKF data assimilation system

Convective-scale data assimilation (DA) faces challenges in effectively incorporating information from complex new sources of weather observations, such as ground-based remote-sensing instruments, satellites or crowd-sourced data. This necessitates a better understanding of the influence of such data in the data assimilation system in order to leverage their informational content in the best possible way and to optimize resource allocation.

In this poster, we present partial analysis increments (PAI) as a computational inexpensive diagnostic to assess the influence of single observations or observations of a particular type on analyses produced in a near-operational set-up of the DA system of the German Weather Service, which is based on the local ensemble transform Kalman filter (LETKF). In particular, the influence of visible satellite observations is compared to that of conventionally assimilated observations. The concepts behind the diagnostic and its relation to single observation experiments are illustrated. Moreover, we use PAI to approximate the sensitivity of the analysis increments to DA parameters, such as the localization length scale and propose an algorithm to optimize the DA system based these results.

Giorgio Doglioni,
University of Trento – Italy

With N. Carlon, L. Cisco, T. Degiacomi, G. Ferrari, A. Chini, D. Zardi

Implementation and testing of a WRF/WRFDA-based operational regional NWP routine for Italy

To obtain optimal initial conditions for initializing regional-scale, high resolution NWPs (Numerical Weather Predictions), it is necessary to derive a realistic estimate of the state of the atmosphere at the initial time. This is operationally done via Data Assimilation.

With the aim of improving the forecasting skill of a regional, 3.5 km resolution NWP chain based on the WRF (Weather Research and Forecasting) model, we designed and tested an initialization routine based on subsequent calls of the 3DVAR algorithm provided within the WRFDA (WRF Data Assimilation system) package alternated to short integrations of the WRF model.

In this presentation we focus on the development and testing of this initialization suite and on the improvement in forecast skill from its operational implementation. Also, the programmed evolution of this initialization suite is discussed along with the expected improvements resulting from these modifications.

Simon Driscoll,
University of Reading – United Kingdom

A data driven emulator of sea ice melt pond processes

Accurate simulation of sea ice is crucial for predicting Arctic sea ice loss. Melt ponds, a key feature of Arctic sea ice, significantly impact the energy budget and climate of the region by altering the ice's albedo. Since melt ponds are small and their evolution depends on various complex factors, their representation in models is parametrized.
We show the level ice melt pond parametrisation in the state of the art Icepack sea ice model, and the Icepack model itself, exhibit substantial sensitivity to the melt pond parameters within their known uncertainty ranges. As a two step process, firstly, we show neural networks can learn and replace the parametrisation itself in the Icepack model, and secondly we learn from observational data the relationship between climate and sea ice state to that of broadband albedo and sea ice melt pond fraction. To achieve this as well as to have the observationally trained emulator run in an online setting, we are exploiting and exploring both machine learning and data assimilation techniques for use where data is sparse and noisy.

Le Duc,
University of Tokyo – Japan

With Y. Sawada

An optimal-transport-based framework for generating analysis ensembles in ensemble filters

An analysis ensemble need be generated in the assimilation step of ensemble filters to continue the next assimilation cycle. In principle, this is equivalent to a sampling procedure from the posterior distribution. Traditional methods are perturbed observations, ensemble transform matrices in ensemble Kalman filters, and resampling in particle filters. In recent years, there are several new appealing methods have been proposed: gain form ensemble transform Kalman filters, quantile-preserving ensemble filters. In this study we show that there exists an elegant theory supported by optimal transport underlying such methods. Given two different probability distributions, optimal transport seeks an optimal transportation map that pushes probabilistic mass from a probabilistic distribution to another. It turns out that the new methods are different implementations of such optimal transportation maps. The underlying principle therefore is to generate an analysis ensemble closest to the corresponding background ensemble.

Connor Duffin,
University of Cambridge – United Kingdom

Statistical finite elements for ocean dynamic processes

I will present a statistical finite element method for nonlinear, time-dependent phenomena. The statistical finite element method (statFEM) is a statistical augmentation of the finite element method that enables model-data synthesis through the admission of model misspecification inside of the governing equations, as represented by a Gaussian process. The method is Bayesian, coherently updates model mismatch upon receipt of observed data, and is applicable to a wide range of problems across science and engineering for which finite element methods are appropriate. Scalability is ensured through making a low-rank approximation to the posterior covariance matrix. I'll first introduce statFEM, before detailing the methodology for nonlinear problems. I will discuss various case studies in oceanography, applying the method to experimental and synthetic data for nonlinear internal waves, and shallow water flow.

Joffrey Dumont Le Brazidec,
École des Ponts and EDF R&D – France

Towards the inversion of plumes from power plants and industrial sites in satellite CO2 images using deep neural networks

Carbon dioxide anthropogenic emissions are the main driver of climate change. Current emission estimates, which are needed to guide reduction policies, rely on statistical data of energy consumption and are subject to important uncertainties.

In order to assess these emissions in an independent, timely and accurate manner, the Copernicus CoCO2 project aims to build a prototype system for a CO_2 emission monitoring service exploiting atmospheric CO_2 measurements, and in particular the XCO_2 images from the future CO2M mission. As part of this project, our goal is to build an inverse modelling system to quantify large local CO_2 sources (large urban areas and industrial sites) based on the spaceborne imagery of the CO_2 atmospheric plumes from these sources. The quantification of such sources depends on the detection of the associated plumes in the satellite images of the CO_2 average column concentrations (XCO_2), which represents a significant challenge. Indeed, the signal of the XCO_2 plumes induced by hotspots emissions rarely exceeds values of a few ppm and its extraction is perturbed by the instrumental noise and variable regional CO_2 background signals in the images.

We explore the use of deep learning techniques to solve the problem of CO_2 plume detection and inversion. Our dataset used to train and test such techniques includes pseudo images based on simulations of hourly XCO_2 fields in the vicinity of various power plants, tracing plume emissions from anthropogenic and biogenic sources.

Specifically, we employ convolutional neural networks (CNN) to analyse the CO_2 plume contour and the associated emissions in a pseudo XCO_2 image. Our findings show that the CNN model outperforms traditional plume inversion approaches such as the cross-sectional method, achieving highly accurate results with a median relative error between 15 and 30%, depending on the power plant studied. Furthermore, our estimations are only slightly affected by the absence of NO2 fields or a detection mechanism. These promising results suggest a high potential of CNNs in estimating local CO_2 emissions from satellite XCO_2 images.

Charlotte Durand,
CEREA, École des Ponts and EDF R&D – France

Deep learning for surrogate modeling to facilitate data assimilation in sea-ice models

A novel generation of sea-ice models can represent the drift and deformation of sea ice with an unprecedented accuracy at the mesoscale. Since these models are computationally heavy, we investigate supervised deep learning techniques for surrogate modeling of neXtSIM from its Arctic-wide simulations. We adapt a convolutional neural network architecture, namely U-Nets, to emulate the sea-ice thickness for a forecast lead time of twelve hours. In our case, the U-Net extracts information at multiple scales and correctly predicts the advection of thickness given several atmospheric forcings. In general, repeatedly applying the neural network performs 36% better than a persistence forecast on a daily timescale, and this gain prevails on monthly timescales with improvements of up to 40%. Still representing the advection of sea-ice, these surrogate models can also handle seasonal forecast of sea-ice thickness with a forecast lead time of up to 5 months. The obtained neural network models allow us to speed-up sea-ice simulations by several orders of magnitude. This allows us to construct larger ensembles to better represent the discrete-continuous behavior of sea ice. Additionally, data-driven sea-ice models can potentially learn a correct representation of small-scale events like cracks and the dynamics of the marginal ice zone. Larger ensembles and a correct representation of the dynamics can facilitate sea-ice data assimilation and could permit the assimilation of yet-unused observations.

Léo Edel,
NERSC – Norway

With Jiping Xie, Calliopé Danton Laloy, Anton Korosov, Julien Brajard, Laurent Bertino

Reconstruction of Arctic sea ice thickness (2000–2010) based on a hybrid machine learning and data assimilation approach

In the Arctic, the sea ice thickness (SIT) remains one of the most challenging parameters to estimate and generally present temporal and spatial discontinuity which are a major difficulty for climate studies. Since 2010, the combined product CS2SMOS enables more accurate SIT retrievals that significantly decrease the SIT errors when assimilated in models (such as TOPAZ4). Can we extrapolate these benefits in the earlier period 2000–2010? In this study, we train a machine learning algorithm to learn the systematic SIT errors between two versions of TOPAZ4 (with and without CS2SMOS assimilation) in 2010–2020, in order to predict the SIT error and extrapolate the SIT prior to 2010. The ML algorithm relies on SIT coming from two versions of TOPAZ4, various oceanographic variables as well as atmospheric forcings (from ERA5). The ML model demonstrates its ability to correct a significant part of the SIT error. We will discuss the sensitivity of the method to the input variables and to different types of ML models. The long Arctic ML-reconstructed SIT record (2000–2022) is validated using in-situ data and earlier satellite data.

Takeshi Enomoto,
Kyoto University – Japan

With S. Nakashita

Assimilation of nonlinear observations using the maximum likelihood ensemble filter with exact Newton optimization

Nonlinear observations can be assimilated with an iterative ensemble-based variational method such as the maximum likelihood ensemble filter (MLEF). However, we find that the conjugate gradient (CG) method sometimes terminates in a few iterations because no cost reduction is expected in the line search. By contrast, the Newton method without line search (exact Newton, EN) continues iterations and converges quadratically. The behaviour of CG and EN is demonstrated in optimization of benchmark functions and assimilation of a single wind speed. Furthermore, EN is found to be more stable in the cycled experiments with a one-dimensional Kortweg—de Vries—Burgers equation model. Finally the localized MLEF with EN is applied to global and regional atmospheric models to improve analysis.

Hassnae Erraji,
Forschungszentrum Jülich GmbH – Germany

High-resolution air quality analyses assimilating unmanned aerial vehicle (UAV) observations

In atmospheric chemistry, a lack of observations exists in the planetary boundary layer (PBL), namely of the vertical distribution of pollutants. In recent years, Unmanned Aerial Vehicles (UAVs) emerged as a promising solution filling this gap by providing in-situ measurements with high temporal and vertical resolution.

This study investigates the impact of assimilating observations from one of the first air quality UAV systems on the analysis using the high-resolution air quality model EURAD-IM. The incorporated 4D-Var system allows for the joint optimization of initial values and emission rates.

The assimilated observations are O3 and NO profiles obtained during a field campaign.

The inversion reveals promising results for the representation of pollutant distributions within the PBL of assimilated and non-assimilated species as well as a valuable information gain for local emission corrections.

Keenan Eure,
Pennsylvania State University – USA

Simultaneous Assimilation of Polarimetric and All-Sky Satellite Data for Ensemble Convection Forecasts

Accurate forecasts of the development and evolution of deep, moist convection in convection-allowing models (CAMs) are both a priority and a challenge of the numerical weather prediction and convective-scale data assimilation communities. Additionally, modeling of the microphysical and internal structures of convection is difficult, as these can affect the storm mode, intensity, and longevity. Underused observations from polarimetric weather radars and all-sky (clear and cloudy affected radiances) satellites have the potential to improve the forecasts of deep convection in CAM ensembles. Since the upgrade to the U.S. national network of WSR-88Ds was completed in 2013, polarimetric radar data offer a wealth of information about the shape, size, and composition of hydrometeors. Several distinct polarimetric signatures in early stages of deep convection have been identified, such as the differential reflectivity (ZDR) column. These columns are vertical protrusions of positive ZDR values above the environmental melting level and can aid significantly in characterizing storm updrafts. Improved information on the updraft has the potential to improve CAM representation of convection. In addition, GOES-16 infrared all-sky brightness temperatures provide complimentary information on cloud structures and cover that S-band radars cannot directly measure. To explore the benefits of both types of data, an ensemble data assimilation approach is used with their simultaneous assimilation. The CAM selected for this study is the Advanced Research version of the Weather Research and Forecasting (WRF-ARW) model with the High-Resolution Rapid Refresh (HRRR) configuration, except for the use of the National Severe Storms Laboratory (NSSL) double-moment microphysics scheme. Observations are assimilated using the Ensemble Kalman Filter (EnKF). Different sets of observations in the experiments are assimilated jointly and separately, and all experiments include conventional observations. Analysis is conducted using a real case to realize the influence of these observations on the prediction of convection initiation, structure, evolution, and its associated hazards. Sensitivities of ZDR assimilation to the quality control, horizontal covariance localization length, and observation error are explored.

Geir Evensen,
NORCE-Norwegian Research Center – Norway

Learnings from petroleum reservoir history matching

This talk explores the dual nature of parameter estimation and recursive ensemble-data-assimilation methods used in petroleum-reservoir history matching and weather prediction. It will shed light on challenges and issues encountered in petroleum applications, such as the impact of ensemble size, the proper specification of measurement error statistics, the additional inclusion of controls or model errors, and the introduction of adaptive correlation-based localization methods. Thus, this talk aims to share experiences and learnings from the petroleum community that uses ensemble data assimilation operationally with the data-assimilation practitioners in the ocean and atmospheric communities.

Luca Facheris,
University of Florence – Italy

With F. Cuccoli, F. Argenti, U. Cortesi, G. Macelloni, A. Antonini, S. Melani, L. Rovai, A. Ortolani

Tropospheric water vapor observations from space through a new measurement concept and the impact on weather forecasts: the SATCROSS project

In a climate change era, measuring atmospheric parameters is fundamental for monitoring the status and initialising the prediction of atmospheric patterns. Among them, the tropospheric water vapor (WV) can contribute to improve the numerical weather prediction (NWP), the assessment of climate feedbacks and the climate projection reliability. A recent project called SATCROSS (Co-rotating satellites for the estimation of water vapor in the troposphere), funded by ASI (Italyn Space Agency) and concluded in 2022, investigated the Normalized Differential Spectral Attenuation (NDSA) method for retrieving WV profiles along atmospheric transects, with a key perspective in the application to small satellites as Cubesats. NDSA measurements are linearly related to the integrated WV along the radio-microwave link paths, that could come from a train of co-rotating LEO satellites. This work aims to illustrate the overall framework of the SATCROSS project with a specific focus on the evaluation of such new measurements in impacting future weather forecasts. A correct initialization of WV patterns in NWP models is indeed a critical aspect, especially when dealing with data assimilation (DA) in high-resolution limited area models, where DA mainly aims at correctly feeding finer scale phenomena, e.g., for enhancing the capability of predicting severe weather events.

Alban Farchi,
CEREA, École des Ponts and EDF R&D – France

Model error correction with data assimilation and machine learning – from theory to the ECMWF forecasting system

Recent studies have shown that it is possible to combine machine learning (ML) with data assimilation (DA) to reconstruct the dynamics of a system that is partially and imperfectly observed. This approach takes advantage of the strengths of both methods. DA is used to estimate the system state from the observations, while ML computes a surrogate model of the dynamical system based on those estimated states. The surrogate model can be defined as an hybrid combination where a physical part based on prior knowledge is enhanced with a statistical part estimated by a neural network. The training of the neural network is usually done offline, once a large enough dataset of model state estimates is available.

Online learning has been investigated recently. In this case, the surrogate model is improved each time a new system state estimate is computed. Although online approaches still require a large dataset to achieve a good performance, they naturally fit the sequential framework in geosciences where new observations become available over time.

Going even further, we propose to merge the DA and ML steps. This is technically achieved by estimating, at the same time, the system state and the surrogate model parameters. This new method has been applied to a two-scale Lorenz system and a two-layer two-dimensional quasi geostrophic model. Preliminary results shows the potential of incorporating DA and ML tightly, and pave the way towards its application to the Integrated Forecast System (IFS) used for operational Numerical Weather Prediction at ECMWF.

Stefano Federico,
CNR-ISAC – Italy

Lightning data assimilation: impact on precipitation and lightning forecast

Lightning is a hazard to life and properties. Lightning is an important issue for electrical companies, forest fires, NOx production, and, in general, outdoor activities. The lightning forecast is important for aviation because, while airplanes are built to withstand lightning strokes, they can suffer structural damage. It is important from a meteorological point of view because it is a clear indication of deep convection. For this reason it can be used to improve the positioning and successive prediction of convection through lightning data assimilation (LDA).

LDA has been used over Italy since many years starting from the work of Federico et al., (2017) and positive impact of LDA on the very short term forecast were reported with both WRF and RAMS@ISAC NWP models. In this work we revise some of the most important results with LDA showing the analysis for a specific case study, the impact of LDA on the precipitation forecast over the sea, and the use of LDA with radar reflectivity data assimilation.

Recently, the possibility to improve the lightning forecast through LDA has also been explored with the WRF model. Results for a two seasons experiment show that LDA can be used to improve the lightning forecast in the 0-3h time range, while the impact is negligible or negative for the 3-6h range.

Reference

Federico, S., Petracca, M., Panegrossi, G., and Dietrich, S.: Improvement of RAMS precipitation forecast at the short-range through lightning data assimilation, Nat. Hazards Earth Syst. Sci., 17, 61–76, doi:10.5194/nhess-17-61-2017, 2017.

Jie Feng,
Fudan University – China

Spatiotemporal estimation of analysis errors in the operational global data assimilation system at the China Meteorological Administration using a modified SAFE method

Quantification of the uncertainties in initial analyses against the real atmosphere ("reality") provides a fundamental reference for the evaluation and development of operational data assimilation (DA) systems. Due to the unknown reality, most existing methods for analysis error estimation use reanalysis datasets or observations as a proxy for reality, which are empirical, nonobjective, and biased. Unlike these methods, our study adopted a modified Statistical Analysis and Forecast Error (SAFE) estimation method to objectively and directly quantify spatiotemporal errors in analyses compared to reality based on unbiased assumptions. In the present study, the SAFE method was first applied to estimate the annual variation and spatial distribution of analysis errors in the Global Forecast System of Global/Regional Assimilation and PrEdiction System (GRAPES_GFS) at the China Meteorological Administration since the beginning of its operational implementation (i.e., 2016—2021).

Qualitative comparison to analysis error estimations in previous studies showed that SAFE can provide more reasonable spatial-mean analysis error profiles than can the estimation with the ERA-5 reanalysis as a reference (the approach hereafter called "ERAv"). Moreover, ERAv overestimates (underestimates) the spatial-mean analysis error below (above) ~500 hPa compared to SAFE because it neglects the uncertainties inherent in reanalysis. Overall, the SAFE estimation reveals that relative reductions of about 12.5%, 29%, and 24.5% were achieved for the spatial-mean analysis errors of wind, temperature, and geopotential height, respectively, in the GRAPES_GFS throughout the six-year study period. These results can largely be attributed to the DA scheme being upgraded from 3D-Var to 4D-Var. SAFE can also provide more reasonable and accurate point-wise analysis errors than ERAv can.

Tobias Finn,
CEREA, École des Ponts and EDF R&D – France

Generating ensembles from single realizations with denoising diffusion models

We present our work on the use of generative deep learning for geophysical systems. Generative modelling has empowered recent breakthroughs in deep learning, like ChatGPT and DALL-E for text and image generation, respectively. As one of the main methods for training deep generative neural networks, denoising diffusion models has been established to generate high-dimensional complex objects. The idea of such denoising diffusion models is simple yet effective: a neural network is trained to denoise and invert a diffusion process in which the training data is noised. After training, the network can be iteratively applied to generate samples of the data distribution from a Gaussian normal distribution; the data generation is inherently stochastic.
We first unveil connections between this generative process and the long-term dynamical properties of chaotic geoscience models. Next, we demonstrate how the generation of objects can be conditioned on a single state realization in denoising diffusion models. With the Lorenz 1963 and 1996 systems, we showcase the conditioning process and evaluate how good denoising diffusion models can represent the true conditional data distribution. Further, we can use the neural network to generate a large ensemble of realizations. This way, the network can represent an informative prior distribution in a Bayesian inference framework, which can facilitate data assimilation in the end.

Steven Fletcher,
Colorado State University – USA

Nongaussian based Kalman Filters

With the advancement of non-Gaussian based variational techniques the need to extend this to hybrid ensemble-variational techniques is the next step towards operational viability. However, there is a problem with the Gaussian assumptions that are made in the derivation of the Kalman filter. In this presentation we shall present a new approach that enables the Kalman filter theory to be applied with: a lognormal, a reverse-lognormal, a Gaussian-lognormal, and a Gaussian-reverse-lognormal distribution. All of there filters are then applied to the Lorenz 1963 model where the z component is known to be have non-Gaussian based errors. All of the filters are compared these against the Extended Kalman Filter to assess improvement over a Gaussian-fits-all approach.

Francesco Fossella,
University of Rome "Tor Vergata" – Italy
Institut Polytechnique de Paris – France

With L. Biferale, A. Carrassi, M. Cencini

Nudging and Ensemble Kalman Filter at high Reynolds numbers

We tested and compared the assimilation techniques of Nudging and Ensemble Kalman Filter (EnKF) with stochastic updating, to a Shell Model for the energy cascade in turbulence. Shell models are known to reproduce the multi-scale and multi-frequencies non-Gaussian fluctuations of 3D turbulent flows at high Reynolds numbers. The aim is to understand how large scale data can be assimilated to improve the instantaneous and statistical prediction of the difficult-to-observe and intermittent small scale variables. We found that, regardless of the observations frequency rate and the set of measured variables, EnKF ensures a more accurate assimilation of the measured variables and a correction of the unmeasured ones that extends over a greater range of scales compared to those achieved by Nudging. Statistical reconstruction is always good and similar among the two approaches.

Alison Fowler,
University of Reading – United Kingdom

With D. Francis, A. Lawless, J. Eyre, S. Migliorini

The importance of anchor observations in data assimilation

Bias correction, via VarBC, is essential for satellite observations to have the dramatic impact on the skill of numerical weather prediction forecasts that they exhibit. To ensure VarBC identifies observation bias and not model bias, unbiased (anchor) observations are necessary. The importance of the precision, positioning and timing of these anchor observations in 4DVar–VarBC is explored theoretically and illustrated with Lorenz 96 model. Anchor observations are shown to be especially helpful in distinguishing model and observation bias when they are collocated with biased observations and occur later in the assimilation window. However, the anchor observations themselves can be a vessel for model bias to contaminate VarBC if the anchor observations observe regions with model bias characteristics different from those of the observations being corrected. Lastly, we look at how metrics of observation impact, such as FSOI, can be used to diagnose the current role of anchor observations and guide network development.

Alison Fowler,
University of Reading – United Kingdom

On the robustness of methods to account for background bias in data assimilation to uncertainties in the bias estimates

Fundamental to the theory of data assimilation (DA) is that the data are an unbiased estimate of the true state. Often this assumption is far from valid and, without bias correction, the resulting analysis will be biased. Two methods to account for biases in the background, b, that do not require a change to the DA algorithm, are compared: explicit bias correction (BC) and covariance inflation (CI). Both methods rely on an estimate of b, which can only be accurately estimated when unbiased observations are available. A lack of unbiased observations means that the estimate of b will exhibit sample errors as well as structural errors due to poor assumptions about the spatial and temporal variability. Given these difficulties in estimating b, the robustness of the two methods in producing an unbiased analysis is studied within an idealised linear system. It is found that: the CI method is much less sensitive to errors in the estimate of b; a smooth estimate of b is crucial to the success of the BC method; the CI method is more sensitive to uncorrected biases in the observations.

Magdalena Fritz,
University of Vienna – Austria

With Stefano Serafin, Martin Weissmann

Parameter estimation for boundary-layer turbulence models over heterogeneous terrain

Accurate representation of complex transport processes in the mountain boundary layer (MoBL) is particularly challenging in numerical weather prediction models. The use of common planetary boundary-layer parameterizations, which invariably assume flat and homogeneous terrain, results in significant systematic errors.

We try to tackle this problem with idealized Observing System Simulation Experiments performed with WRF. Experiments consist of a large-eddy simulation (LES) nature run and a single column model (SCM) ensemble, where the only model error source is the PBL parameterization. In the SCM ensemble, we assimilate surface observations and vertical profiles from the LES simulation with the DART Ensemble-Adjustment Kalman Filter, and we estimate the parameters that affect vertical turbulent mixing. Through optimal, e.g., height-dependent tuning of these parameters, we strive for a reduction of systematic errors.

Kensuke Fujimura,
Chiba University – Japan

Ensemble Kalman Filtering with perturbed rainfall for improving flush flood predictions by a Rainfall-Runoff-Inundation Model

This study aims at applying the ensemble Kalman filter (EnKF) for a rainfall-runoff-inundation (RRI) model to improve flood forecasts with assimilation of operational water-level observations. Data assimilation methods including EnKF usually maintain appropriate forecast ensemble spread based on error propagations of dynamical models. However, forecast ensemble spread decreases in time integration of the RRI model since uncertainty of river routing model is dominated by input rainfall uncertainty. Here we inflate the forecast ensemble spread of RRI model by perturbing the input rainfall intensity and model soil physical parameters for each ensemble member. We compared the RRI simulations with and without EnKF assimilation in Omono River Basin in Japan, and found that the EnKF approach improved predictions of water levels at both observed and unobserved stations.

Yukiko Fujisawa,
Keio University – Japan

With N. Sugimoto, N. Komori, H. Ando, M. Takagi, T. Kouyama, M. Yamamoto, C. Ao, I. Garate-Lopez, S. Lebonnois, AFES-Venus & ALEDAS-V teams

Series of Observation System Simulation Experiments for the Venusian atmosphere

We have developed the first data assimilation system for Venusian atmosphere named "ALEDAS-V" (AFES LETKF Data Assimilation System for Venus) using the atmospheric GCM (general circulation model) and Local Ensemble Transform Kalman Filter (LETKF). Data assimilation on Venus is still in its early stages, and there are few examples of Observation System Simulation Experiments (OSSEs) for the planetary satellite observation missions including Venus. Since planetary satellite missions are limited in time and resources, it would be quite useful to investigate their effectiveness by using OSSEs in advance to optimize their instruments and observation plans. In this presentation, we introduce various OSSEs assuming new and existing satellites observation missions on Venus.

Lilian Garcia-Oliva,
University of Bergen – Norway

With François Counillon, Ingo Bethke and Noel Keenlyside

Intercomparison of initialization methods for Seasonal-to-Decadal Climate Predictions with the NorCPM

The challenge in developing reliable seasonal-to-decadal climate predictions lies in finding the best combination of initialization methods that effectively constrain the atmospheric and/or oceanic states while also addressing the significant biases present in current models through full-field or anomaly initialization.

To address this, we compared several approaches using the Norwegian Climate Prediction Model and the same experimental design: generating reanalysis and seasonal-to-decadal hindcasts for 1980-2010. Specifically, we compared a version with full-field atmospheric nudging (NudF), a version with anomaly atmospheric nudging (NudA), a version that assimilates ocean data with the Ensemble Kalman Filter (ODA), and a version that combines ODA and NudA (ODA-NudA).

Our findings indicate that full-field initialization performs best for short lead times (one month) and over specific regions (such as the tropical Atlantic Niño). However, anomaly initialization is preferable for longer lead times as it helps to limit drift and sustain prediction skill. Overall, ODA outperforms atmospheric initialized versions, but the latter is better at capturing specific events, such as the rapid North Atlantic subpolar Gyre shift. Combining the two methods did not yield optimal results as sustaining the reliability of the ensemble with atmospheric nudging proved challenging, leading to significant degradation of the ODA.

Thomas Gastaldo,
Arpae Emilia-Romagna-SIMC - Italy

With V. Poli, C. Marsigli, E. Minguzzi, D. Cesari., P. P. Alberoni

Assimilation of radar observations in the ICON-KENDA system at Arpae-SIMC

At Arpae-SIMC,the HydroMeteorological and Climate Service of Emilia-Romagna region, in Italy, the ICON model coupled to the KENDA-LETKF data assimilation system is scheduled to replace, by the end of the year, the COSMO-KENDA model chain in providing deterministic and probabilistic weather forecasts over Italy at a horizontal resolution of 2.1 km. The analyses will be generated by assimilating, in addition to conventional data, radar reflectivity and radial wind volumes from the Italyn radar network. Moreover, radar estimated precipitation will be assimilated employing a LHN scheme. The impact on analysis and forecast accuracy of employing radar data in the ICON-KENDA system will be shown. Verification will be performed for precipitation and other near-surface parameters, as well as prognostic variables over a one month period. The impact on case studies of intense precipitation will be also presented.

Hugo Georgenthum,
IMT Atlantique - France

With Lucas Drumetz, Julien Le Sommer, David Greenberg, Ronan Fablet

Short-term forecast of sea surface height from NADIR and SWOT altimetry data with 4DVarNets

Sea surface height forecasting from satellite-derived altimeter observations is a key challenge in operational oceanography. While current operational systems exploit model-driven data assimilation schemes, we investigate deep learning schemes, and more particularly 4DVarNets. They provide neural data assimilation schemes and have been proven successful for the space-time mapping of sea surface height (SSH) maps with NADIR and in a near future SWOT altimetry data. Here, we extend their application to short-term forecasting problems for sea surface dynamics. Formally, we state the short-term forecasting of SSH fields from altimeter data as a space-time interpolation problem. We design an Observing System Simulation Experiment in a Gulf Stream region from high-resolution numerical simulations (NATL60 dataset) with NADIR and SWOT altimeter observations. It supports the relevance of the proposed 4DVarNet schemes to improve the short-term forecasting compared with different baselines. We will discuss further the key features of our approach as well as the relative importance of SWOT data.

Azadeh Gholoubi Khonacha,
NOAA/NWS/NCEP/EMC – USA

Development in All-Sky Microwave Radiance Assimilation over Land in NCEP Global Model

The microwave signal emerging from land surfaces depends on frequency, polarization, and incidence angle. It is also affected by surface types and conditions, making it challenging to model surface emissivity accurately. The complexity of modeling the interaction between all these surface parameters and the microwave radiation has generally restricted the use of observation to temperature and humidity channels, which receive a weak contribution from the surface over land. Therefore, the first step towards all-sky assimilation of microwave observation in the NCEP Global Forecast Model (GFS) is obtaining more reliable surface emissivity estimates. This approach falls into the following steps: evaluating the land surface microwave emissivity and its sensitivity to soil and vegetation types in the Community Radiative Transfer Model (CRTM), collaborating with the CRTM team to update the land surface categories based on the NCEP global surface model,improving the surface emissivity estimation with instantaneous emissivity retrieval from MW surface-sensitive channels along with an emissivity atlas. In this presentation, the progress to all-sky microwave radiance assimilation over land surface will be discussed, and the preliminary results on emissivity retrievals from the Gridpoint Statistical Interpolation (GSI) data assimilation system at NCEP will be demonstrated. In addition, the idea of using machine learning (ML) methods to improve land surface emissivity estimation will be investigated.

Nicole Girardot,
Météo – France

Experimentation of model perturbations in Météo-France global EDA

An Ensemble Data Assimilation (EDA) system has been in operations at Météo-France since 2008 for the ARPEGE global model, in order to estimate and specify flow-dependent background error covariances in 4D-Var. This EDA also provides initial perturbations for the global Ensemble Prediction System (EPS), and perturbed lateral boundary conditions for the AROME mesoscale EDA. The current version includes 50 members with truncation T499 in the forecast and trajectory steps. Observation errors are simulated by random draws of the matrix R, and model errors are currently accounted for through inflation of cycled forecast perturbations. In order to improve the model error simulation, random perturbations applied to model physical parameters are being experimented in the ARPEGE EDA, following previous experimentations conducted in the ARPEGE EPS. The poster will briefly present the method, how it is tested, and some of the results, aiming at a future implementation in operational suite.

Humberto C. Godinez,
Los Alamos National Laboratory – USA

With Anthony DeGenaro, Nishant Panda

Data Assimilation with Spectral Decomposition Methods

In this work we present the use of Koopman eigenfunctions in four-dimensional data assimilation (4D-Var) to simplify the required optimization and avoid the need for adjoint models. Data assimilation are methods that combine information from a model and observations to produce a accurate forecast or prediction of the phenomena of interest. One of the most effective data assimilation methods is the 4D-Var, which uses information available within a time window to calibrate the model inputs (initial conditions, parameters, etc.) in order to improve the forecast of the model. In essence the 4D-Var is an optimization problem that uses a penalizing or cost functional which measures the difference between the model prediction and available observations. The necessary model derivative needed for the optimization is computed through the adjoint model. In this talk we use the Koopman eigenfunctions to approximate the derivative of the model, and reduce the space within which the optimization is performed. We present results with a 2-dimensional shallow water model, where we compare our Koopman eigenfunction approach with the more traditional approach using the adjoint model. The results show that the Koopman eigenfunctions provide a reasonable result in the 4D-Var optimization, and hence is comparable to the use of the adjoint model.

Luis Gustavo Goncalves,
CMCC – Italy

Assimilation of Land Surface States at CMCC using SPREADS/Ensemble Kalman Filter

In this presentation, we will discuss the critical role of land surface states in the initialization of seasonal global numerical prediction. Specifically, we will examine the importance of leaf area index, soil moisture, above ground biomass, and snow cover. We will begin by highlighting the latest research at CMCC, emphasizing the significant impact that accurate land surface states can have on prediction accuracy.

The ongoing research and development using SPREADS with Ensemble Adjustment Kalman Filter (EAKF) focuses on assimilating observations of these surface states into the CLM5-BGC-Crop land surface model, and we will provide details on this work. The results presented here are from 2016 and 2017, during which we conducted offline data assimilation experiments to obtain surface fields using synthetic and real observations. We will discuss the impacts of assimilating each of the surface observations type separately and also all together.

Finally, a prototype for a Land Data Assimilation System (LDAS) at CMCC in support to land surface initialization of ensemble data assimilation for the atmosphere, weakly coupled DA and hydrometeorological reanalysis will be presented.

Olivier Goux,
CERFACS – France

With A. Weaver, S. Gürol, O. Guillet

Accounting for correlated observation error is variational ocean data assimilation

In variational data assimilation, the assumption of uncorrelated observation errors is commonly made to simplify access to the inverse correlation operator. However, this assumption becomes problematic when dealing with certain observation types, such as high-resolution satellite data. Neglecting correlated errors during the assimilation process leads to suboptimal analyses, with observations being overfitted at large spatial scales and underfitted at small spatial scales. To address this issue, diffusion operators provide a cost-effective and flexible framework for modeling the inverse observation error correlation operator with unstructured data, while preserving the conditioning of the inverse problem.

Philipp Griewank,
Universität Wien – Austria

Localizing ensemble observation impact estimates

Ensemble sensitivity can be used to quantify the impact of observations (e.g., E-FSOI) or to determine where to best deploy observations to improve the forecast (e.g., observation targeting and network design). The sensitivity is calculated from the covariances of the initial ensemble to the forecast ensemble, and is accordingly affected by sampling errors. These sampling errors can be reduced via localization. Additionally, the impact estimate needs to account for the localization of the data assimilation system. As there are two different reasons for localization in the calculation of observation impact, finding appropriate localization settings is very challenging in practice, especially when signal propagation is unknown and different for different atmospheric quantities. We detail how the choice of temporal localization affects ensemble sensitivity observation impact estimates for observations with a linear and non-linear operator in a toy model.

Oliver Guillet,
Meteo – France

A C++ implementation of the diffusion method to account for correlated observation errors in OOPS-QG

We present an implementation of the non-diagonal observation error correlation matrix within the OOPS (Object Oriented Prediction System) framework. The code is based on the implicit diusion equation, discretized on the mesh of the observations using a nite element method. The code is developped as a standalone library and has a generic structure which does not depend on the forecast model being used in OOPS.

We show briefly how the code is organised. The main features are highlighted using the quasi-geostrophic toy model available in OOPS. These features include parameters selection for the correlation model, building of the mesh from the observation locations and how to the memory is handled using object-oriented programming.

Lastly, we run a data assimilation experiment on one time-step to showcase the performance of the system in the presence of correlated observation errors. The quality of the analysis is discussed relative to the cost of the inner loop iterations. Pointers on the convergence of the 3DVar are given, with possible ways to further improve the performance of the system.

Selime Gurol,
CERFACS – France

Randomized numerical linear algebra for variational data assimilation

Over the past decade, randomized algorithms have proven to perform well on a large class of numerical linear algebra problems including large-scale problems. Randomized algorithms can significantly accelerate computations and are well suited for exascale platforms. To be able to use these algorithms efficiently, it is crucial to analyze their theoretical convergence properties. In this work, we explored the potential use of randomized linear algebra to accelerate the convergence of the strong constraint variational data assimilation by using different strategies. Numerical experiments are performed by using Object Oriented Prediction System (OOPS) platform.

Soyoung Ha,
National Center for Atmospheric Research – USA

Coupled data assimilation for the operational air quality forecasting in Korea

This talk will introduce the new coupled data assimilation and forecasting system operated by the Korean National Institute of Environmental Research (NIER) using the Modal for Aerosol Dynamics in Europe and the Volatility Basis Set (MADE-VBS) aerosol chemistry in the WRF-Chem model coupled with WRFDA. Along with the recent implementation of aqueous aerosol chemistry, this talk will also present the ongoing effort on the assimilation of aerosol optical depth (AOD) retrieved from the South Korea's Geostationary Environment Monitoring Spectrometer (GEMS) using the updated version of the Community Radiative Transfer Model (CRTM).

Zdenko Heyvaert,
KU Leuven – Belgium

With M. Bechtold, S. Scherrer, W. Dorigo, A. Gruber, S. Kumar, G. De Lannoy

Assessment of multi-layer increment distributions in an EnKF system for land data assimilation

Microwave-based satellite retrievals of surface soil moisture from the combined active-passive ESA CCI Soil Moisture product are assimilated into the Noah-MP land surface model, using a one-dimensional Ensemble Kalman Filter (EnKF) within the NASA Land Information System (LIS). Prior to assimilation, the soil moisture retrievals are rescaled to the model climatology through cumulative distribution function (CDF) matching.
The data assimilation (DA) system produces consistent estimates of geophysical variables over Europe for an 18-year period by updating soil moisture states in each of the four soil layers of the Noah-MP model. This approach marginally improves soil moisture estimates of the model, with the largest improvements found over agricultural areas.
In this study, we analyze the multi-layer increment distributions (MLIDs) which depict how increments of soil moisture between the surface layer and the deeper layers of the model correlate. We show that two distinct regimes emerge from the MLIDs: one showing a strong coupling (high correlation) between increments of different soil layers, and one showing a weak coupling (low correlation).
We assess the impact of meteorological variables (e.g., rainfall), land variables (e.g., root-zone soil moisture), and fluxes (e.g., evapotranspiration) and find that the strong coupling regime in the MLID is typically linked to times and locations with wetter soils and larger hydrological fluxes. Additionally, we look at the impact of three design choices of the DA system on the MLIDs: the size of the perturbations applied to the soil moisture observations, the method to perform CDF matching (per month or over the time series as a whole), and the choice of the meteorological dataset (MERRA-2 or ERA5) used to force the land surface model. Overall, our analysis provides additional insights into the conditions that result in a more substantial impact of surface soil moisture DA on the deeper model layers.

Ieuan Higgs,
University of Reading & National Centre for Earth Observation – United Kingdom

With J. Skakala, A. Carrassi, R. Bannitster and S. Ciavatta

Machine learning for multivariate data assimilation in marine biogeochemistry

We used machine learning to develop a surrogate model that aims to perform a multivariate post-processing step on a univariate data assimilation scheme, in a biogeochemical marine ecosystem model. Our approach aims to reduce the need for large, expensive ensemble methods to estimate the background statistics. The data used for this are obtained from a state-of-the art coupled marine physics-biogeochemistry model simulating the vertical column of a coastal station (L4) in the English Channel across a long time-period (approx. 10 years). The surrogate is trained on the output of a large ensemble free-run with randomly distributed key biogeochemical model parameters. An additional model run served as a synthetic truth for the data assimilation process. We tested a variety of surrogate models with different input features and architecture, optimising the performance while identifying key predictors of the system.

The results of this study show that we can effectively propagate an analysis increment of a single variable in a system where the background statistics can often be poorly estimated. This is a promising step towards using machine learning surrogates as a viable alternative to large, expensive ensembles, or an operational setting where the number of ensemble members can often be a limiting factor.

Helen Hooker,
University of Reading – United Kingdom

Flood inundation forecast improvement by assimilating satellite SAR-derived probabilistic observations

Advances in flood forecasting both at global and local levels link together meteorological and hydrological forecasts that drive the selection of pre-computed flood maps from a simulation library. The simulation library obviates the need to run a hydrodynamic model as part of the forecast process, reducing computation time. Observations of flooding from Synthetic Aperture Radar (SAR) satellite data have previously been used to update flooding extent and depths from hydrodynamic models. However, the accuracy of a simulation library flood forecasting system is largely determined by the flood map selection criteria such as the return period thresholds set. We present a data assimilation framework incorporating probabilistic observations of flooding from SAR into a simulation library flood forecasting system. By utilising SAR data, we aim to improve the flood map selection.

Elias Holm,
ECMWF – United Kingdom

With S. Lang, M. Bonavita, P. Lean, M. Chrust

Soft Re-Centred Ensemble of Data Assimilations

Computational efficiency improvements are central to the development of the Ensemble of Data Assimilations (EDA). One goal for the EDA is to become a reliable estimator of the initial and short-range forecast states. This would avoid the need for ad-hoc spread adjustments like multiplicative inflation in 4D-Var and would enable direct initialisation of the ensemble forecast (ENS) from the EDA. We have gradually improved the accuracy and reliability of the EDA through more representative model and observation uncertainty representations and increase in number and resolution of members enabled through computational efficiency gains. Inspired by Desroziers and Berre (2012) the ECMWF operational EDA uses pre-conditioning vectors from the unperturbed members to speed up convergence of the perturbed minimizations. Building further on their work, the soft re-centred EDA in addition uses the control vector from the unperturbed member's minimization to warm-start the perturbed members' minimizations, which is also conceptually related to the mean-perturbation method of Lorenc et al. (2016). Taken together, the computational optimizations of the EDA, with 50% from last few years and 30% added by the soft re-centring, results in EDA about 1/3 of the cost for same performance and enables improved performance at manageable costs.

The soft re-centring improvement in accuracy and computational efficiency is achieved in two steps.

To improve overall accuracy, the control member of the EDA is run in a 4D-Var configuration closer to the high-resolution 4D-Var (more outer loops at higher minimisation resolutions). The background and first guess from the more accurate control are then used in the perturbed members to: a) re-centre the perturbed members' background on the control's background, which improves the accuracy of the EDA mean analysis and b) to add the control member preliminary analysis increment to perturbed members' background to give a first. As a result, we are able to run the perturbed members in a significantly (approx. 30%) cheaper one outer loop 4D-Var configuration without loss of accuracy and reliability compared with the current two outer-loop configuration. In effect the soft-centring idea integrates the re-centring into the analysis update rather than apply it as a post-processing step after the analysis ("hard" re-centring). As each perturbed analysis is the result of a full 4D-Var analysis update, it is a valid model trajectory and this allows to minimise the initial shock of re-centring.

With soft re-centring there are large improvements in the forecast skill of individual EDA members and an increase in spread of the whole EDA, which are visible when we run an ENS forecast started directly from the EDA initial conditions. Some of these improvements remain visible also in the operational-like ENS setup where 6-hour EDA forecasts are hard re-centred on the control analysis and singular vectors are also added.

We will present results from current work on soft re-centred EDA at same resolution as the operational deterministic analysis at TCo1279/9km and same inner loop resolution.

Daisuke Hotta,
Meteorological Agency – Japan

VAE as a Stochastic Multidimensional Extension to Gaussian Anamorphosis (Part 2)

In the Machine Learning session of ISDA-online in 2021, the author reported that Variational Auto-Encoder (VAE) can be use as a a stochastic multidimensional version of Gaussian Anamorphosis and that this approach allows us to simultaneously account for three difficulties encountered in assimilating image data such as satellite radiances and radar reflectivity: (1) non-Gaussianity, (2) dimensional redundancy, and (3) inter-pixel error correlation. The previous talk focused on the special and simple case where the observations and the model state reside in the same space (i.e., H=id). This presentation extends this approach to cases where the observation operator subsamples and applies nonlinear transformation to the state vector. (739 characters)

Chih-Chi Hu,
Colorado State University – USA

With P. J. van Leeuwen, A. J. Geer

Incorporating non-Gaussian observation errors into variational methods

The observation error has been assumed to be Gaussian in most data assimilation (DA) applications. One of the reasons is that we typically do not have access to the actual shape of the observation error probability density function (pdf). Recently, we have developed a new method, the Deconvolution-based Observation Error Estimation (DOEE), that allows us to derive a fully non-Gaussian pdf even for observations with complicated representation error.

With DOEE, we can now explore the effect of non-Gaussian observation error in the assimilation using non-parametric DA methods, e.g., particle (flow) filters. However, most of the operational weather forecasting centers are still based on variational DA methods, e.g., 4D-Var, in which the DA algorithms are constructed based on Gaussian observation errors. In this presentation, we will propose novel ways to account for non-Gaussian observation errors in the incremental 4D-Var without changing the cost-function, such that it can be easily implemented in a full-scale operational weather forecasting system without adding to the computational cost. We will demonstrate the methodology in the ECMWF system with the assimilation of satellite microwave radiances, which have non-Gaussian observation errors due to the representation errors. The preliminary results indicate promising signs of improvement for the 12-h forecast of the low-level water vapor, cloud and precipitation.

Chih-Chi Hu,
Colorado State University – USA

With P. J. van Leeuwen; J. L. Anderson

Observing system simulation experiments (OSSE) using the Particle Flow Filter (PFF) in a high-dimensional atmospheric model in the Data Assimilation Research Testbed (DART)

The Particle Flow Filter (PFF) is a fully nonlinear and non-Gaussian data assimilation (DA) method that avoids the problem of weight collapse in a standard particle filter setup, and therefore has the potential to be applied in a high-dimensional system. The PFF has been shown to work effectively in high dimensional Lorenz models.

In order to implement PFF for an even higher-dimensional weather forecasting problem, we adapt the PFF to fit in a revised two-step filtering algorithm in the Data Assimilation Research Testbed (DART). In this presentation, we will first introduce the algorithm, and then demonstrate an observing system simulation experiments (OSSE) using a simplified atmospheric general circulation model. Specifically, the performance of the data assimilation using the PFF will be compared with using the Ensemble Adjustment Kalman Filter (EAKF). Discussions will be focused on how the PFF and the EAKF deal with the nonlinear observations differently, and how these differences can affect the model behavior.

Yasutaka Ikuta,
Japan Meteorological Agency – Japan

Generating background error covariances for hydrometeors with conditional generative adversarial networks

To assimilate observations related to hydrometeors, such as radar and brightness temperature, using variational methods, it is important to provide a flow-dependent background error. In this study, a conditional generative adversarial network (CGAN), a type of deep learning network, was used to generate background errors for hydrometeors. The CGAN is trained using the image of background error covariance in advance. In the assimilation, the conditional data for each grid point is created by the first guess and a few ensemble members and is input into the CGAN generator, which generates an image of the background error. The generated images are converted to background error and used for data assimilation. In the presentation, we would like to show the assimilation.

Bruce Ingleby,
ECMWF – United Kingdom

Improved 2m temperature forecasts from assimilation changes at ECMWF

Two metre temperature (t2m) is a key forecast variable and we describe planned assimilation changes at ECMWF to improve short-range forecasts of t2m. Currently screen temperature and humidity from SYNOP and METAR reports are assimilated in a separate surface analysis – primarily to update soil moisture – but only daytime screen humidity is assimilated in the main 4DVar. Tests assimilating t2m in 4DVar gave significant improvements in short-range t2m forecasts but in Winter made tropospheric fields, including T850, somewhat worse. Various remedies were tried: the exclusion of t2m data in cases with a low-level inversion tended to remove both the main benefits and the detrimental side effects.

Reducing the weight given to t2m data with large departures from background was a useful compromise, limiting the assimilation of screen data to the first 6 hours of the 12 hour 4DVar window also reduced the unwanted side effects. Further improvements result from changes to the snow data assimilation as well as to the use of t2m and soil moisture data in the surface analysis. The net result, expected to become operational in 2024, gives improved short-range t2m forecasts particularly in northern hemisphere winter. Investigation of possible further changes to improve t2m analysis/forecast are continuing.

Bruce Ingleby,
ECMWF – United Kingdom

ECMWF use of Mode-S winds and changes to aircraft thinning

High resolution Mode-S aircraft data over Europe were introduced in August 2020, earlier than planned because of the Covid pandemic. Their use was switched off in Nov 2022, because we found that locally the density was too high for the number of 4DVar iterations performed making the forecasts worse.

The problem was first noted via poor observation fit to analysis statistics over central Europe. The ‘track thinning’ used was not sufficient with the many 100s of flights over Europe.

Instead ‘box thinning’ was adapted for aircraft use and tested, initially for Mode-S data and then for all aircraft data. The results are encouraging and the short-range benefits of using Mode-S winds over Europe are again clear in the T+12 fit to European radiosondes. The benefit is largest (up to 6 or 7%) for upper tropospheric wind, but is also substantial for temperatures. We hope to reintroduce Mode-S data operationally in summer 2023.

Tijana Janjic,
KUEI – Germany

With Y. Zeng

Weakly Constrained LETKF for Convective-Scale Data Assimilation

Physical properties of the system that we are modeling dictate plausible values of the initial conditions of the employed numerical models. Unfortunately, state-of-the-art data assimilation techniques such as the ensemble Kalman filter algorithm frequently provide initial conditions that violate essential physical properties. In this talk, we present a new algorithm that weakly but simultaneously preserves mass and non-negativity (Janjic and Zeng, 2021). Our algorithm is a modification of the local ensemble transform Kalman filter and is therefore, easily applicable in high-dimensional systems. The algorithm was tested in idealized experiments that assimilate radar data in a non-hydrostatic, convection-permitting numerical model while updating hydrometeor values. Results confirmed the hypothesis that adding physical constraints when generating initial conditions of models that contain source and sink terms is beneficial. Namely, that physical constraints successfully improve the mass conservation property in analyses and by doing so reduce the increase in integrated mass-flux divergence and vorticity caused by data assimilation, leading to better short-term forecasts.

Tijana Janjic,
KUEI – Germany

With T. Gleiter, N. Chen

Ensemble Kalman Filter based Data Assimilation for Tropical Waves in the MJO Skeleton Model

The Madden-Julian oscillation (MJO) is the dominant component of tropical intraseasonal variability with wide reaching impacts even on extratropical weather and climate patterns. However, predicting the MJO is challenging. One reason are suboptimal state estimates obtained with standard data assimilation (DA) approaches. Those are typically based on filtering methods with Gaussian approximations and do not consider physical properties that are specifically important for the MJO.

In our recent paper (Gleiter et al. 2022), a constrained ensemble DA method is applied to study the impact of different physical constraints on the state estimation and prediction of the MJO with the Skeleton model. The utilized quadratic programming ensemble (QPEns) algorithm extends the standard stochastic ensemble Kalman filter (EnKF) with specifiable constraints on the updates of all ensemble members. This allows to recover physically more consistent states and to respect possible associated non-Gaussian statistics. Our results demonstrate an overall improvement in the filtering and forecast skill when the model's total energy is conserved in the initial condition. The degree of benefit is found to be dependent on the observational setup and the strength of the model's nonlinear dynamics. It is also shown that even in cases where the statistical error in some waves remains comparable to the stochastic EnKF during the DA stage, their prediction is remarkably improved when using the initial state resulting from the QPEns.

Tijana Janjic,
KUEI – Germany

With S. Legler, Y. Ruckstuhl

Learning model parameters from observations by combining data assimilation and machine learning

Parametrization of microphysics as well as parametrization of processes in the surface and boundary layers typically contain several tunable parameters. The parameters are not observed and are only crudely known. Traditionally, the numerical values of these model parameters are chosen by manual model tuning, leading to model errors in convection permitting numerical weather prediction models. More objectively, parameters can be estimated from observations by the augmented state approach during the data assimilation or by combing data assimilation with machine learning (ML).

If the parameters are updated objectively according to observations, they are flexible to adjust to recent conditions, their uncertainty is considered, and therefore the uncertainty of the model output is more accurate. To illustrate benefits of online augmented state approach, Ruckstuhl and Janjic (2020) show in an operational convection-permitting configuration that the prediction of clouds and precipitation is improved if the two-dimensional roughness length parameter is estimated. This could lead to improved forecasts of up to 6 h of clouds and precipitation. However, when parameters are estimated by the augmented state approach, stochastic model for the parameters needs to be pre-specified to keep the spread in parameters. Alternatively, Legler and Janjic (2022) investigate a possibility of using data assimilation for the state estimation while using ML for parameter estimation in order to overcome this problem. We train two types of artificial neural networks as a function of the observations or analysis of the atmospheric state. The test case uses perfect model experiments with the one-dimensional modified shallow-water model, which was designed to mimic important properties of convection. Through perfect model experiments we show that Bayesian neural networks (BNNs) and ensemble of point estimate neural networks (NNs) are able to estimate model parameters and their relevant statistics. The estimation of parameters combined with data assimilation for the state decreases the initial state errors even when assimilating sparse and noisy observations.

Kaushambi Jyoti,
University of Vienna – Austria

Testing Hybrid-3DEnVar in the convective-scale NWP model AROME over Austria

Hybrid 3D ensemble variational (3D-EnVar) data assimilation uses a weighted combination of climatological and ensemble background error covariances. We test a hybrid 3D-EnVar approach for the convective-scale AROME NWP model over Austria for two different summertime convective cases, and compare our results against the operational 3DVar of Geosphere Austria. We use 50 ensemble members, and explore the Valid Time-Shift (VTS) ensemble approach to increase the ensemble size. Furthermore, we explore the localization settings of AROME for vertical localization.

Takuya Kawabata,
Meteorological Research Institute – Japan

With T. Oizumi, P.-Y. Wu, L. Duc, K. Saito

The impact-based forecasting with a large-ensemble DA

Impact-based forecasting (IBF) aims to encourage people's behavioral changes when disasters occur. That is, when meteorological disasters such as floods are predicted, we want to encourage desirable behavior to reduce the damage. For realizing IBF, probabilistic prediction is an important component, because the behavior of MCSs and typhoons is essentially chaotic due to their high nonlinearity. Therefore, accurately estimating the probability of its occurrence is rather important than to accurately predicting meteorological phenomena. We will present a large-ensemble DA system with 1000 members for IBF conducted on the super computer Fugaku and its applications in this symposium.

Fumitoshi Kawasaki,
Chiba University – Japan

Solving Data Assimilation on Quantum Annealing Machines

Data assimilation plays an important role in numerical weather prediction (NWP) to provide optimal initial conditions. Many operational NWP centers use variational methods that iteratively reduce cost functions based on their gradients. Due to the iterations, huge computational resources are consumed for data assimilation in NWP systems. This study proposes solving data assimilation on quantum annealing machines by reformulating the problem into the Ising model. For the quantum annealing machines, we used simulated and D-wave's physical quantum annealing machines. Our preliminary 4DVAR experiments using the 40-variable Lorenz model were promising, showing that the quantum annealers yield analysis whose accuracy was comparable to the quasi-Newton method. The physical quantum machine approach solved the 4DVAR with much shorter time than the other approaches.

Fumitoshi Kawasaki,
Chiba University – Japan

Leading the Lorenz system toward the prescribed regime by model predictive control combined with data assimilation

Typhoons and torrential rain have become stronger and more frequent in recent years. If we could lead the weather toward non-disastrous regimes, such weather-related damage could be considerably reduced. Miyoshi and Sun (2022) proposed the Control Simulation Experiment (CSE), which tries to lead a system toward preferable regimes by adding small perturbations into the nature run. The CSE requires long-term ensemble forecasts to find the necessary perturbations. Alternatively, this study introduces Model Predictive Control (MPC) into data assimilation to lead a system toward preferable regimes. The MPC finds effective control inputs by model-based prediction and optimization. Our preliminary experiments with Lorenz model show that the proposed method successfully leads the system toward the prescribed regime without employing the long forecasts in contrast to the CSE.

Junkyung Kay,
National Center for Atmospheric Research – USA

With J. Pinto, K. Fossell, P. McCarthy, G. de Boer

UAS observation error estimation for data assimilation and its impact on predictive skill

Observation error determines the weights of the observation and background state in data assimilation, and quantifying observation error is critical to the successful assimilation of new observational datasets. Uncrewed Aircraft System (UAS) observations have shown potential benefits in filling observational gaps in the atmospheric boundary layer (ABL). To utilize the UAS observational capabilities in improving numerical weather prediction, we estimate the UAS observation error variance based on the triple collocation diagnostic approach which compares three independent data sets including UAS, rawinsonde, and model simulations. We apply this approaches to data from two experiments: the 2018 Lower Atmospheric Profiling Studies at Elevation — a Remotely-operated Aircraft Team Experiment (LAPSE-RATE) field campaign and the 2021 Atmospheric Radiation Measurement (ARM) Southern Great Plains UAS Intercomparison Project. To assess the impacts of the new error variance on predictive skill, the Data Assimilation Research Testbed (DART) and the Advanced Research version of the Weather Research and Forecasting model (WRF-ARW) are used to assimilate the UAS observations with the newly estimated observation error variances for a case study of terrain-driven flows observed during LAPSE-RATE. Details of the method used to estimate UAS observation error profiles will be presented along with an assessment of the impact of this new estimate of observation error on forecast skill.

Christian Keil,
University of Munich – Germany

With G. Craig, C. Keil, L. Scheck, M. Weissmann

How to use partial analysis increments in an LETKF data assimilation system

Convective-scale data assimilation (DA) faces challenges in effectively incorporating information from complex new sources of weather observations, such as ground-based remote-sensing instruments, satellites or crowd-sourced data. This necessitates a better understanding of the influence of such data in the data assimilation system in order to leverage their informational content in the best possible way and to optimize resource allocation.

In this poster, we present partial analysis increments (PAI) as a computational inexpensive diagnostic to assess the influence of single observations or observations of a particular type on analyses produced in a near-operational set-up of the DA system of the German Weather Service, which is based on the local ensemble transform Kalman filter (LETKF). In particular, the influence of visible satellite observations is compared to that of conventionally assimilated observations. The concepts behind the diagnostic and its relation to single observation experiments are illustrated. Moreover, we use PAI to approximate the sensitivity of the analysis increments to DA parameters, such as the localization length scale and propose an algorithm to optimize the DA system based these results.

Kobra Khosravian,
Deutscher Wetterdienst – Germany

With Klaus Stephan, Jana Mendrok, Alberto De Lozar and Ulrich Blahak

Assimilation of 3D radar information at convective scales at Deutscher Wetterdienst

The study of radar data assimilation in NWP models and its effect particularly on short-term forecast has been intensified recently in DWD. In particular, the seamless Integrated ForecastiNg sYstem (SINFONY) project which leads a short-term forecasting system with a focus on convective events from minutes up to 12 hours shows clearly the benefit of radar data assimilation in improving the short-term forecast.

Besides radar data assimilation is being operationally used in the short-range ensemble numerical weather prediction (SRNWP) system (ICON-D2-KENDA LETKF system) in DWD since 2020 (radial wind from March 2020 and reflectivity from June 2020). This is in addition to the traditional Latent Heat Nudging (LHN) of 2D radar-derived precipitation rates.

We will present the latest results of our research in radar data assimilation in DWD including the results of using the radar data assimilation with a more sophisticated double-moment bulk microphysical scheme in combination with the LHN in the SINFONY forecasting system as well as the latest study of applying the polarimetric radar variables, applying the radar network of other European countries in DWD assimilation system and other ongoing research studies in DWD related to the improvement of radar data assimilation in particular in short-term forecasting system.

Yu-Shin Kim,
University of Oklahoma – USA

With X. Wang, A. Johnson

Impact of Microphysics Parameterization Schemes on the Assimilation of GOES-16 All-Sky Infrared Radiances for a Bow Echo Analysis and Prediction

Assimilating infrared brightness temperatures from the water vapor sensitive channels of the GOES-16 Advanced Baseline Imager (ABI) has been shown by past studies to improve the analysis and prediction of severe weather events. These studies are limited to using a single microphysics scheme. Microphysics schemes are expected to affect bow echo dynamics and brightness temperature. Therefore, this study aims to investigate how assimilating GOES-16 ABI infrared brightness temperatures with different microphysics schemes contributes to the analysis and prediction of the 3 May 2020 bow echo case. The Gridpoint Statistical Interpolation based Ensemble Kalman (GSI-EnKF) system and Weather Research and Forecasting (WRF) model are utilized to conduct data assimilation experiments using Thompson, WDM6, NSSL, and Morrison microphysics schemes.

Correlations structures between brightness temperature and model state variables indicate that assimilating infrared brightness temperatures can adjust bowing MCS dynamics via latent cooling and rear inflow jet. Such corrections enhance the rear inflow jet and bow echo size, primarily for microphysics schemes featuring faster hydrometeor fall velocity and stronger latent cooling. The improved analyses lead to better forecasts of the bow echo's shape, size, timing of the bowing process, and wind speeds. Substituting a larger microphysics-dependent effective radius for a constant default value increases warmer prior brightness temperatures, innovations, and corrections in the rear inflow jet, especially for the WDM6 and Morrison schemes. In the subsequent forecasts, incorporating microphysics-dependent effective radius further improves the experiment using the Morrison scheme but degrades it when using the WDM6 scheme.

Rakesh Teja Konduru,
RIKEN Center for Computational Science – Japan

With J. Liang, and T. Miyoshi

Challenges in Assimilation of High-Frequency Satellite Observations Using NICAM-LETKF in the OSSE framework

We assessed the assimilation of more frequent satellite microwave radiance observations through observing system simulation experiments (OSSEs) by using the global numerical weather prediction system known as the NICAM-LETKF. A nature run was performed by taking the initial condition from the NCEP FNL analysis data. Synthetic conventional observations were generated from nature based on the actual locations and errors from the NCEP PREPBUFR data, and synthetic AMSU-A satellite observations were produced using a radiative transfer model known as the RTTOV. We performed four experiments: one with only conventional observations (NOSAT) and others with additional satellite observations every hour (1H), every 3 hours (3H), and every 6 hours (6H). RMSE of global temperature suggested poor performance of 1H compared to 3H and 6H. Since 1H showed a significantly higher model imbalance, we increased the observation error of the satellite observations and repeated the 1H experiment (1H_R). The imbalance measure was reduced in 1H_R to a similar level as the other experiments, and the RMSE became the lowest.

Shunji Kotsuki,
Chiba University – Japan

Exploring weather control technology to steer the atmosphere towards favorable directions

Japan's Moonshot program, initiated by the Science and Technology Agency, aims to mitigate weather-related disasters by developing weather control technology. Our project explores methods to steer the atmosphere towards favorable directions. Our initial investigations involve analyzing meteorological landscapes that differentiate disaster and non-disaster scenarios. Using operational weather prediction data, our data-driven approach captures Typhoon Dolphin's landscape, revealing two distinct regimes that may be controllable with small manipulations. Additionally, we have begun coupling the model predictive control and data assimilation to find effective manipulations that guide the dynamical system towards predetermined directions. This presentation provides the most recent advancements of our project until the time of the symposium.

Shunji Kotsuki,
Chiba University – Japan

Combining Data Assimilation and Data-driven Sparse Sensing Placement Method For Designing Better Observation Locations for NWP

This study proposes a data-driven approach for determining where to place additional observations to improve numerical weather predictions via the effective use of mobile observations by aircrafts and ships. Here we use the sparse sensor placement (SSP) method which determines the optimal sensor locations so that the selected sensors effectively determine coefficients of proper orthogonal decomposition (POD) modes. We introduced the SSP into an ensemble data assimilation system that incorporates an intermediate global atmospheric model and the local ensemble Kalman filter. This combined approach determines additional observation locations adaptively based on the POD modes of ensemble forecasts. Our preliminary experiment was promising, showing that the SSP-based placement resulted in more accurate analyses than spread-based and uniform placements.

Konstantin Krüger,
German Aerospace Center – Germany

With A. Schäfler, G. Craig and M. Weissmann

The influence of radiosonde observations on the sharpness and altitude of the tropopause

A realistic representation of the extratropical tropopause (TP) in NWP models is critical for reliable weather and climate forecasts. However, current models fail to correctly reflect the sharpness of the TP (i.e., the gradients of wind and temperature). In this study, we investigate whether assimilated radiosonde observations are able to improve the TP structure in models.

This work is based on more than midlatitude 9500 radiosondes that were assimilated by the ECMWF IFS during a one-month period in autumn 2016. In addition, to test whether the diagnosed influence is caused by the assimilated radiosondes, a data denial experiment was conducted, that excluded 500 sondes launched in the framework of the NAWDEX field campaign. Here, we present the departures of a control run (CTR) with all radiosondes assimilated and an identical denial run (DEN) without the NAWDEX radiosondes. The influence is evaluated by comparing temperature, static stability (N^2) and wind profiles of the observations (y), the observation-space background (Hxb) and the analysis (Hxa) in tropopause-relative coordinates.

Temperature is overestimated by the model background near the TP (warm bias, ~1K) and underestimated in the lower stratosphere (cold bias, 0.5) which is also reflected in a clear underestimation of the jump of N^2 at the TP. We show that data assimilation counteracts the background biases and in turn slightly improve the representation of TP sharpness and TP altitude in the analysis. Wind speed is underestimated by the background, especially at and below the TP (~1 m s-1), but the assimilation improves the wind profile across the TP. This positive influence (but also the background bias) is particularly larger in strong wind situations, i.e., the jet stream.

The analysis of the data denial experiment reveals that the diagnosed TP sharpening exists in the CTR, but only to a limited extent in the DEN run, and thus can be attributed to the assimilation of the radiosonde data.

Although the assimilation improves wind and temperature gradients across the TP by pulling the background closer to the observations, the analysis still underestimates the sharpness of the TP which is expected to negatively impact the quality of weather and climate projections.

Monika Krysta,
Bureau of Meteorology – Australia

National Analysis System for Australia

The Australian Bureau of Meteorology is currently developing its future operational National Analysis System. The goal of the system is to generate atmospheric analyses representative of the current meteorological situation. The analyses provide a three-dimensional depiction of the atmosphere, initially on an hourly basis. They are formed in addition to standard analyses by assimilating low latency observations and assigning them higher than typical weights in a 3D-Var formalism. In this presentation we will discuss the concept of the system, introduce its components, describe the implementation with the emphasis on data assimilation, specify the assimilated observations and illustrate the results of subjective and objective verification.

Lukas Kugler,
University of Vienna – Austria

Comparing the assimilation of visible and infrared satellite observations to radar reflectivity for convective-scale numerical weather prediction

Early warning of storms requires early detection and assimilation of convective-scale features including clouds. Visible satellite imagery detects deep convection earlier than radar, but is largely overlooked by research and operational centers (apart from DWD). We estimate the potential impact of assimilating cloud-affected satellite observations in the visible (0.6 µm) and near thermal infrared wavelengths (6.2 µm and 7.3 µm) relative to the impact of assimilating radar reflectivity observations. For that purpose, we performed idealized observing system simulation experiments (OSSE) using the Weather Research and Forecasting model (WRF) at 2-km grid resolution, the radiative transfer model RTTOV/MFASIS, and the Ensemble Adjustment Kalman Filter in the Data Assimilation Research Testbed (DART). The forecast impact was evaluated in two cases: isolated and scattered supercells with different assumed prior uncertainties.

The main result is that satellite observations can be nearly as beneficial as three-dimensional radar reflectivity observations, depending on the prior forecast uncertainty. We conclude that there is a high potential impact of assimilating cloud-affected satellite observations when the location of convection is uncertain and comparatively less potential when the stage of convective development is uncertain. Additionally, we are studying the effect of observation operator nonlinearity on the analysis mean and variance update to investigate undesirable systematic effects.

Martin Lange,
Deutscher Wetterdienst – Germany

With G. Geppert, T. Hüther, S. Hollborn, R. Potthast
Deutscher Wetterdienst, Offenbach, Germany

Development of a 2d-Var system for surface analysis at DWD

DWD is running operationally a global and regional forecast system including weakly coupled Data assimilation for surface variables. To estimate snow depth and sea surface temperature a Cressman analysis is performed, an optimal interpolation scheme (OI) is used for 2m Temperature analysis, and for initialization of soil moisture content a 2d-Var (z,t) method is applied using parameterization of the Jacobians (dT2m/dwb) with T2m as 2m temperature forecast around noon, and wb as initial soil moisture content at 0:00 UTC of the respective top and bottom soil layers.

Separate analysis systems are maintained in addition to the atmospheric analysis which is build in DWD's Data Assimilation Coding Environment DACE.
In order to replace the partly old fashioned data assimilation methods i.e. Cressman and OI, and the respective software code which has grown over the last three decades, new surface analysis programs are developed for snow, SST, and near surface parameter, based on DACE which offers flexible use of variational methods, including options for use of an ensemble based B-matrix.

DACE also provides the fundament to process new observation types and large observation data using parallelism and appropriate forward modeling operators.

The unification of the code base leads to easier maintenance by exploiting synergies and more resilience of the operational framework through broader distributed expert knowledge.

A first implementation of a 2d-Var scheme using horizontal and vertical distance weighted correlations, is done for the 2m temperature analysis which is used as input for soil moisture analysis, and several experiments are performed for optimal configuration.

The presentation will detail the new 2d-Var surface analysis system and shows the impact in comparison with the present OI analysis. An outlook will be given on developments regarding the extension to EnVar method and modifications required for analysis of other NWP relevant surface variables as snow depth and sea surface temperature.

Florian Le Guillou,
ESA - France

With E. Cosme, S. Metref, C. Ubelmann, M-H Rio

Parallel long-window weak-constraint 4D variational assimilation for the reconstruction of ocean surface dynamics

We present a long-window, weak-constraint 4D variational assimilation scheme for dynamically estimating the spatio-temporal trajectory of ocean surface dynamics. The dynamical operator is a quasi-geostrophic (QG) model whose simplicity makes it easy to control over time. The control vector associated to the space-time evolution of the QG model error is expressed in a reduced basis of multiscale wavelet functions localized in space and time. Although the QG model is simple, the reduced basis for the model error is built to represent the statistics of ocean general circulation models. The formulation of the inverse problem enables an easy use of parallelization in the time dimension. The performances of the proposed algorithm are evaluated both with simulated and real data and show a net improvement of the resolved scales compared to operational products. The application of this method is obviously not limited to ocean surface dynamics, but can be applied to a wide variety of problems involving the reconstruction of biogeophysical fields from observations.

Mathieu Le Provost,
MIT - USA

An ensemble filter for heavy-tailed t-distributions

Heavy tails is a common feature of filtering distributions that results from the nonlinearity and stochasticity of the dynamical and observation processes. In these settings, the Kalman filter and its ensemble version - the ensemble Kalman filter (EnKF) - that have been designed under Gaussian assumptions have degraded performance. t-distributions are a parametric family of distributions whose tail-heaviness is modulated by a degree of freedom ν interpolating between a Cauchy distribution for $\nu = 1$ and a Gaussian distribution for $\nu = \infty$. Leveraging ideas from measure transport (Spantini et al., SIAM Review, 2022), we present a generalization of the EnKF whose prior-to-posterior update leads to exact inference for t-distributions. We show that this new filter embeds an adaptive and observation-dependent multiplicative inflation mechanism that alleviate need for any tuning. Moreover, it recovers the Kalman filter for $\nu = \infty$. For nonlinear state-space models with heavy-tailed noise, we propose an algorithm to estimate the prior-to-posterior update from joint samples of the states and observations. We rely on a regularized expectation-maximization algorithm to estimate the mean, covariance, and the degree of freedom of heavy-tailed t-distributions from limited samples (Dogru et al., Commun. Stat., 2018). By sequentially estimating the degree of freedom at each analysis step, our filter has the ability to automatically adapt its prior-to-posterior update to the tail-heaviness of the data. We demonstrate the benefits of this new ensemble filter on challenging Eulerian and Lagrangian filtering problems with heavy-tailed noise.

Mathieu Le Provost,
MIT – USA

Regularization of the ensemble Kalman filter for non-local observations: application to elliptic observations

In many practical settings, we are interested in assimilating non-local observations given by integrals of linear and nonlinear functions of the state, such as radiance measured by satellites, fluxes through surfaces, or solutions of elliptic partial differential equations (PDE). However, the predominant regularization techniques of the ensemble Kalman filter (EnKF), such as distance-based localization, assume that the observations are local, i.e., they are only informative of the physically close state variables. With non-local observations, however, we cannot separate slowly decaying physical interactions from spurious long-range correlations. We propose a regularization of the EnKF that is applicable to generic non-local observation operators, but focus our attention on elliptic PDEs. These non-local inverse problems are usually highly compressible: low-dimensional projections of the observations strongly inform a low-dimensional subspace of the state space. We identify bases for the informative subspaces of the state and observation spaces from the Jacobian of the observation operator. Using these eigenvectors, we propose a low-rank factorization of the Kalman gain that sheds light the inference process: the innovation term is whitened and projected onto the observation eigenvectors, assimilated in the identified informative subspace and lifted to the original space. From the fast spectral decay, a few dominant modes capture the row and column spaces of the Kalman gain leading to a major reduction in variance. We demonstrate the low-rank EnKF (LREnKF) on Lagrangian data assimilation problems encountered in fluid mechanics. We alleviate the need for any tuning of the LREnKF by adaptively identifying the dimensions of the informative subspace.

Lili Lei,
Nanjing University – China

Convolutional Neural Network-based Adaptive Localization for an Ensemble Kalman Filter

Flow-dependent background error covariances estimated from short-term ensemble forecasts suffer from sampling errors, due to limited ensemble sizes and model errors. Covariance localization is often used to mitigate the sampling errors, especially for high dimensional geophysical applications. Most applied localization methods, empirical or adaptive ones, multiply the Kalman gain or background error covariances by a distance-dependent parameter, which is a simple linear filtering model. Here two localization methods based on convolutional neural networks (CNNs) learning from paired datasets are proposed. The CNN-based localization function (CLF) aims to minimize the sampling error of the estimated Kalman gain, and the CNN-based empirical localization function (CELF) aims to minimize the posterior error of state variables. These two CNN-based localization methods can provide localization functions that are spatially and temporally adaptive, non-symmetric with respect to displacement, without prior assumptions for the localization functions and with flexible representation ability. Results using the Lorenz05 model show that CLF and CELF can better capture the structures of the Kalman gain than the optimal Gaspari and Cohn (GC) localization function and the adaptive reference localization method. For both perfect- and imperfect-model experiments, CLF produces smaller errors of the Kalman gain, prior and posterior than the optimal GC and reference localization, especially for spatially-averaged observations. Without model error, CELF has smaller prior and posterior errors than the optimal GC and reference localization for spatially-averaged observations, while with model error, CELF has smaller prior and posterior errors than the optimal GC and reference localization for single-point observations.

César Magno Leite de Oliveira Júnior,
University of São Paulo – Brazil

With A.M. Saraiva, A.C. Botazzo Delbem, H. Fraga de Campos Velho, L.F. Sapucci, J. Gerd Zell de Mattos, C.F. Bastarz, G. Gallarreta Zubiaurre Lemos.

Radiance Data Assimilation by Cellular Neural Networks

Artificial intelligence techniques have good potential for applications in the areas of weather and climate. These new techniques can improve predictions, significantly reducing the computational effort. The investigation of cellular neural networks (CNN) in this context is proposed here. CNN has similarities with cellular automata, resulting in information with the most minor error being propagated across the network. Data assimilation (DA) is an essential procedure in numerical weather forecasting, where previous forecast data ("background") are optimally combined with data from atmospheric observations to generate the initial condition – called "analysis" – for the next prediction cycle. DA is a computationally expensive procedure. Preliminary experiments with a low-order model (chaotic Lorenz system) verified the usefulness of this technique, where the success of the assimilation process with CNN was obtained. Subsequently, a new structure for CNN was developed in order to assimilate data from the AMSU-A sensor, using simulations of the Community radiative transfer model (CRTM) generated from data of the Brazilian Global Atmospheric Model (BAM), used operationally by the INPE (National Institute for Space Research), Brazil. The modified CNN architecture for the BAM, which operates with dynamic filters, is proposed to reproduce the radiance data assimilation process, typically used in global scale numerical prediction models. There are dynamic filters, created from observations and simulations, operating in a cell and its neighborhood, propagating the pattern of the cell that has the smallest error value within the neighborhood.
The analyses obtained in the assimilation process of radiance data by CNN were compared with the analyses used by INPE, which uses the Gridpoint Statistical Interpolation (GSI) coupled in the BAM model to carry out the DA. The GSI framework uses the 3D-Var method for data assimilation. In our research, the results of both assimilation techniques were verified in a time series for the synoptic times of February 2023, comparing the mean global error and the standard deviation for the window channels of the AMSU-A sensor.
The results showed that the proposed CNN reproduces the spatial patterns, maintaining the consistency of the CRTM simulation data, but reducing the magnitude of the global error found and obtaining better values for the standard deviation. The experiments also point out that the CNN-based assimilation can produce more accurate analyzes than the 3D-Var method, as obtained in the whole study period and in all sensor channels.

Hong Li,
Shanghai Typhoon Institute,
China Meteorological Administration – China

Assimilation of Radar reflectivity data for the analysis and forecast of landfalling typhoon: OSSEs and Real data application

Radar reflectivity data are directly assimilated in EnKF with a complex operator based on double-moment (DM) microphysics, with OSSEs for typhoon In-Fa (2021), leading to obvious improvement in intensity and precipitation forecast. Through the cycled assimilation, EnKF adjusts well not only hydrometers related to reflectivity operator but also some other variables indirectly related to reflectivity, especially for temperature and vertical velocity (T and W). The structures and dynamics of the ensemble-based correlations between reflectivity and all analysis variables are examined to show how assimilating reflectivity can improve the analyses and forecast. Besides, sensitivity experiments are conducted for updating different state variables. The results show that updating hydrometers only may improve the precipitation forecast at initial time, but the positive impact vanishes quickly during the forecast, while updating a full set of state variables is helpful to extend the assimilation benefit to 15 hours. Moreover, the results of assimilating real reflectivity data for typhoon Lekima (2019) will also be presented.

Tianyi Li,
Department of Physics and INFN,
University of Rome "Tor Vergata" – Italy

Exploring Generative Models on Lagrangian Trajectories for Navier-Stokes Turbulence

Generative models have made significant advancements in computer vision, inspiring further exploration of their applications in turbulent flow data generation. A Lagrangian trajectory represents the path of a passive particle or drifter, advected by the underlying turbulent flow. Utilizing generative models to produce augmented datasets for single and multi-particle diffusion in complex flows is crucial for enhancing predictive capabilities and understanding various flow phenomena in meteorology and climate sciences.

Our study generates Lagrangian trajectories (three velocity components) in homogeneous isotropic turbulence using two state-of-the-art generative models: Wasserstein Generative Adversarial Networks (WGAN) and the more recent, promising Diffusion Models (DM). Both methods exhibit remarkable performance, with DM outperforming WGAN in probability distributions of velocity components, probability distributions of accelerations, structure functions, and flatness. Notably, WGAN results consistently demonstrate a bottleneck enhancement limitation, particularly when generating three velocity components simultaneously, whereas DM results are satisfactory.

Compared to WGAN, DM's primary disadvantage is slower sampling. However, by adjusting the noise schedule, we can achieve an acceptable computational cost. Additionally, DM is more versatile in conditioning, enabling tasks such as data prediction, completion, and other data assimilation processes essential for understanding and forecasting atmospheric and oceanic phenomena.

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 882340).

References

[1] Buzzicotti, Michele, et al. "Lagrangian statistics for Navier—Stokes turbulence under Fourier-mode reduction: fractal and homogeneous decimations." New Journal of Physics 18.11 (2016): 113047.
[2] Buzzicotti, Michele, et al. "Reconstruction of turbulent data with deep generative models for semantic inpainting from TURB-Rot database." Physical Review Fluids 6.5 (2021): 050503.
[3] Li, Tianyi, et al. "Generative Adversarial Networks to infer velocity components in rotating turbulent flows." arXiv preprint arXiv:2301.07541 (2023).

Chih-Hsin Li,
Central Weather Bureau – Taiwan

Evaluation of an Ensemble Partial Cycle Framework for Use in the Regional Ensemble Prediction System at the Central Weather Bureau of Taiwan

The Central Weather Bureau (CWB) of Taiwan is responsible for issuing official weather forecasts and warnings of hazardous weather. For this purpose, CWB has been developing and operating global and regional numerical weather prediction (NWP) systems and providing both deterministic and ensemble forecast products as the forecast guidance. With the goal to move toward a more strongly linked global and regional NWP system, efforts have been made to connect the two systems. This study proposes a new regional ensemble initialization method named Ensemble Partial Cycle (EPC) to generate regional ensemble initial conditions for CWB's WRF-based regional Ensemble Prediction System (WEPS). The EPC method can be conceptually regarded as a combination of the Ensemble of Data Assimilation and partial cycle approaches. Our early results show that this new method is superior to the method that is adopted in the current operational WEPS based on an Ensemble Adjustment Kalman Filter data assimilation. The superiority is in terms of a better ensemble spread-error relationship and improved forecast performances. With the encouraging results from this newly proposed EPC method and the relevant development in CWB's global prediction system, it becomes possible to establish a better connected global and regional NWP system, which is more suitable for long-term NWP research and development in CWB. In this presentation, results from WEPS with EPC initialization will be further evaluated with several case studies to gain a more thorough understanding of the benefits of the proposed method.

Jianyu (Richard) Liang,
RIKEN Center for Computational Science – Japan

With Koji Terasaki, Takemasa Miyoshi

Developing observation operator based on machine learning model for satellite data assimilation

In satellite data assimilation (DA), the observation operator (OO) is usually based on radiative transfer models such as RTTOV, and it usually takes several years to develop a DA system to include new satellite data. To accelerate this, we developed OO based on machine learning (ML) (ML-OO). We assimilated the conventional observations and AMSU-A brightness temperature (BT) using the RTTOV as OO in a NICAM-LETKF DA system. Next, the model forecast and the BT were used to train the ML model. We tested the performance of the ML-OO. We first assimilated the conventional observations (CONV). We then assimilated additional BT using RTTOV (CONV+BT-RTTOV) and ML (CONV+BT-ML). Evaluated by the RMSD of temperature using ERA-interim, CONV+BT-ML is slightly worse than CONV+BT-RTTOV but better than CONV, which proves the effectiveness of ML-OO. We are currently investigating training the ML-OO without using RTTOV.

Jianyu (Richard) Liang,
RIKEN Center for Computational Science – Japan

With Norihiko Sugimoto, Takemasa Miyoshi

Understanding the Dynamics of Venus' Atmosphere using Bred Vectors

A data assimilation (DA) system for Venus' atmosphere has been developed based on the AFES model and LETKF (AFES-LETKF) (Sugimoto et al. 2017). To better understand the dynamics of Venus' atmosphere and improve the ensemble forecast and DA techniques, this study applies the breeding of growing modes method with the AFES model. Previous studies used this method to study Earth's atmosphere and ocean, as well as the Martian atmosphere. We use a 1-year forecast as a control run. Starting from random perturbed states, we conducted multiple breeding cycles, varying rescaling amplitude (Ramp) and rescaling interval (Rint). The resulting bred vectors (BVs), which are the differences between the control and the perturbed runs, can emphasize growing modes. We discovered several instabilities in Venus' atmosphere using BVs. The next steps include calculating the BV energy equation and the BV dimensions to gain more insight into the dynamics of Venus' atmosphere, which will lead to an improvement of DA.

Guo-Yuan Lien,
Central Weather Bureau - Taiwan

Evaluation of an Ensemble Partial Cycle Framework for Use in the Regional Ensemble Prediction System at the Central Weather Bureau of Taiwan

The Central Weather Bureau (CWB) of Taiwan is responsible for issuing official weather forecasts and warnings of hazardous weather. For this purpose, CWB has been developing and operating global and regional numerical weather prediction (NWP) systems and providing both deterministic and ensemble forecast products as the forecast guidance. With the goal to move toward a more strongly linked global and regional NWP system, efforts have been made to connect the two systems. This study proposes a new regional ensemble initialization method named Ensemble Partial Cycle (EPC) to generate regional ensemble initial conditions for CWB's WRF-based regional Ensemble Prediction System (WEPS). The EPC method can be conceptually regarded as a combination of the Ensemble of Data Assimilation and partial cycle approaches. Our early results show that this new method is superior to the method that is adopted in the current operational WEPS based on an Ensemble Adjustment Kalman Filter data assimilation. The superiority is in terms of a better ensemble spread-error relationship and improved forecast performances. With the encouraging results from this newly proposed EPC method and the relevant development in CWB's global prediction system, it becomes possible to establish a better connected global and regional NWP system, which is more suitable for long-term NWP research and development in CWB. In this presentation, results from WEPS with EPC initialization will be further evaluated with several case studies to gain a more thorough understanding of the benefits of the proposed method.

Guo-Yuan Lien,
Central Weather Bureau - Taiwan

With Z.-M. Huang, T.-Y. Chao, W.-H. Teng, C.-H. Lin, L.-F. Hsiao, J.-H. Chen, D. Kleist

The GFS-GSI based Global Forecast System Adapted at Central Weather Bureau of Taiwan: Data Assimilation Development and Performance Evaluation

In collaboration with the U.S. National Centers for Environmental Prediction (NCEP) since 2016, the Central Weather Bureau (CWB) of Taiwan has been making efforts on adapting the NCEP's FV3 dynamical core based Global Forecast System (GFS) with the GSI-hybrid 4DEnVar data assimilation for operations at the bureau. Development of the first operational version of the CWB-localized Global Forecast System has been completed and is scheduled for transitioning to operations this year. In this system, the deterministic model is run at a horizontal C384 (about 25 km) resolution and the ensemble Kalman filter system is run at a C192 (about 50 km) resolution, both of which are half of the current operational resolution at the NCEP. The observations assimilated in the hybrid data assimilation at CWB are similar but fewer than those assimilated at the NCEP. In this presentation, the CWB-localized Global Forecast System is briefly described, with a focus on several data assimilation developments and studies done at CWB, including the improvement of the observation error specification of the GNSS radio occultation observations, the use of time-lagged ensemble in the hybrid 4DEnVar, and the assimilation of the Himawari-8 AHI radiance observations. The forecast performance of the system in CWB's operational setting will also be shown. In addition, based on the configuration of the system and some recent experiments, the impacts of differences in resolutions and assimilated observations will be discussed.

Sujeong Lim,
Ewha Womans University – Republic of Korea

With Seon Ki Park and Milija Zupanski

Assimilation Impact of Soil Moisture in a Strongly Coupled Atmosphere-Land Data Assimilation System

Soil moisture is important in a coupled atmosphere-land surface model because it propagates to atmospheric variables such as temperature and water vapor mixing ratio in the planetary boundary layer through the latent and sensible heat fluxes. Therefore, soil moisture observations in a coupled atmosphere-land surface data assimilation system can provide useful information for both the land surface and atmospheric systems. In this study, we assimilated the National Aeronautics and Space Administration's Soil Moisture Active Passive (SMAP) soil moisture retrievals, which observe the top 5 cm of soil moisture with a global coverage every 2–3 days and a 1000 km swath width. We interfaced the Maximum Likelihood Ensemble Filter (MLEF), a hybrid ensemble-variational data assimilation system, with Weather Research and Forecasting (WRF). As a strongly coupled data assimilation system, MLEF assimilated both atmospheric and soil moisture observations — the National Centre for Environmental Prediction (NCEP) Prepared Binary Universal Form for the Representation of meteorological data (PrepBUFR) and the SMAP soil moisture retrievals — and simultaneously corrected the atmospheric and land surface variables. Results showed that the coupled atmosphere-land surface data assimilation system can improve soil moisture as well as near-surface atmospheric variables in the planetary boundary layer through cross-covariance between land and atmosphere. To develop a triple-coupled data assimilation system (i.e., a coupled atmosphere-land-chemistry data assimilation system) using the WRF with Chemistry (WRF-Chem), we also preliminary investigated the sensitivity of dust variables propagated from atmospheric and soil moisture observations on Asian dust events.

Shun Liu,
NWS/NCEP/EMC – USA

Data Assimilation System Development and Testing for Rapid Refresh Forecast System

The Rapid Refresh Forecast System (RRFS) is NOAA's next generation regional convection-allowing ensemble forecast system under development for the National Weather Service. RRFS is being developed collaboratively by EMC and GSL along with the larger Unified Forecast System (UFS) community. The RRFS, which covers North America at 3-km horizontal grid spacing, incorporates an hourly cycled deterministic forecast system, an hourly cycled EnKF data assimilation system and a ~12 member time-lagged ensemble forecast system. The deterministic forecast system includes an hourly spin-up cycle and an hourly product cycle. The spin-up cycle starts from a 3-h GFS forecast at 03 and 15z and is cycled for 6 hours. Ensemble members from Global Data Assimilation System (GDAS) are only used in hybrid 3dEnvar in the first spin-up cycles when starting RRFS, otherwise, 30 RRFS members will be used. The product cycle also performs hybrid 3DEnVar assimilation of conventional and radar-reflectivity observations with 30 RRFS 3-km ensemble members. The forecast system provides a deterministic 18 hour forecast each hour and a 60 hour ensemble forecast every 6 hours. The hourly cycled EnKF is re-initialized from ensemble perturbations derived from members of the Global Ensemble twice a day at 06z and 18z and includes assimilation of both conventional and radar reflectivity observations. Toward the implementation of the first version of RRFS, within the constraints of the limited computational resources at NCEP, RRFS will be tested using real-time parallel experiments and retrospective experiments. The forecast performance will be reported at the conference.

Alexander Lobbe,
Imperial College London – United Kingdom

Noise calibration for the stochastic rotating shallow water model

Stochastic partial differential equations have been used in a variety of contexts to model the evolution of uncertain dynamical systems. In recent years, their applications to geophysical fluid dynamics has increased massively. For a judicious usage in modelling fluid evolution, one needs to calibrate the amplitude of the noise to data. In this paper we address this requirement for the stochastic rotating shallow water (SRSW) model. This work is a continuation of a previous study, where a data assimilation methodology has been introduced for the SRSW model. The noise used before was introduced as an arbitrary random phase shift in the Fourier space. This is not necessarily consistent with the uncertainty induced by a model reduction procedure. In this work, we introduce a new method of noise calibration of the SRSW model which is compatible with the model reduction technique. The method is generic and can be applied to arbitrary stochastic parametrizations. It is also agnostic as to the source of data (real or synthetic). It is based on a principal component analysis technique to generate the eigenvectors and the eigenvalues of the covariance matrix of the stochastic parametrization. For SRSW model covered in this paper, we calibrate the noise by using the elevation variable of the model, as this is an observable easily obtainable in practical application, and use synthetic data as input for the calibration procedure.

Andrew Lorenc,
Met Office, United Kingdom – United Kingdom

Designing new DA software for Operational Global NWP

The Met Office's current global NWP system, using the Unified Model and hybrid-4DVar, will be replaced before 2027, on a new supercomputer, using new LFRic model software and JEDI DA software. To minimise risk during this big change, the initial global DA design uses well-proven DA methods, retaining the hybrid-4DVar method which performs well in the current system, while adding an innovative Hybrid statistical-dynamical Tangent Linear Model to facilitate changing model parametrisations.

Requirements are evolving: short-range regional forecasts need a Rapid Update Cycle, and modern operational forecasting needs an ensemble-based system. I will explain how the above system can be adapted to satisfy both these new requirements, without excessive computational cost, using a new "Control-Pert" variational formulation for the ensemble DA.

Cristian Lussana,
Norwegian Meteorological Institute – Norway

With Thomas N. Nipen and Ivar A. Seierstad

MET Nordic dataset: post-processing of model output near-surface fields with unconventional observations

MET Nordic is a dataset of near-surface variables for Scandinavia, Finland, and the Baltic countries at 1 km resolution produced by the Norwegian Meteorological Institute (MET Norway). The dataset goes back to 2012 and it is updated in real time every hour. The MET Nordic dataset consists of post-processed products that (a) describe the current and past weather (analyses), and (b) give our best estimate of the weather in the future (forecasts). The products integrate output from MetCoOp Ensemble Prediction System (MEPS) as well as measurements from various observational sources, including crowdsourced weather stations. These products are deterministic, that is they contain only a single realization of the weather. The forecast product forms the basis for the forecasts on Yr (https://www.yr.no). Both analyses and forecasts are freely available for download.

The near-surface variables included in the dataset are: two-metre temperature, precipitation, air pressure at sea level, relative humidity, wind speed and direction, solar global radiation, long-wave downwelling radiation, cloud area fraction. For temperature and precipitation, the model output is combined with unconventional observations, such as data from citizen weather stations. Their inclusion shows a clear improvement to the accuracy of short-term temperature forecasts, especially in areas where existing professional stations are sparse. In this study, we will summarize the results obtained with the post-processing and we will share the main lessons learned, which can also be useful for systems that want to use these observations for data assimilation.

Citizen weather stations are rapidly increasing in prevalence and are becoming an emerging source of weather information. These low-cost consumer-grade devices provide observations in real time and form parts of dense networks that capture high-resolution meteorological information. Despite these benefits, their adoption into operational weather prediction systems has been slow. However, MET Norway recently introduced observations from Netatmo's network of weather stations in the post-processing of near-surface temperature and precipitation analysis and forecasts. The inverse problem theory offers an ideal framework for the combination of observations with a numerical model background. In particular, despite the final products being deterministic, we have considered a modified ensemble optimal interpolation scheme. Concepts and methods that are usually found in data assimilation are here applied to spatial analysis, where they have been adapted in an original way to represent precipitation at finer spatial scales than those resolved by the background, at least where the observational network is dense enough. The observations are used to continually correct errors in the weather model output caused by unresolved features such as cold pools, inversions, urban heat islands, and an intricate coastline. Integrating citizen observations into operational systems comes with a number of challenges. First, operational systems must be robust and therefore rely on strict quality control procedures to filter out unreliable measurements. Second, post-processing methods must be selected and tuned to make use of the high-resolution data that at times can contain conflicting information. Central to resolving these challenges is the need to use the massive redundancy of citizen observations, with up to dozens of observations per square kilometer, and treating the data source as a network rather than a collection of individual stations.

Anqi Lyu,
Nansen Environmental and Remote Sensing Center – Norway

With F. Counillon

Compare online versus offline assimilation for paleoclimate reanalysis

Much of our understanding of historical climate comes from reanalysis based on instrumental data, but their length limits their use in studying longer-term climate variations. Paleoclimate proxy data are long enough to monitor multidecadal variability but they sparse and heterogeneously distributed. Some model-based paleo proxy reconstruction has been constructed using an offline data assimilating framework, meaning that it linearly combines states from a large dynamical model database a-posteriori. In an online data assimilation approach, observations are assimilated sequentially. The superiority of online approach over offline approach was demonstrated in an idealized framework and with a linear inverse model. Here, we compare the offline and online data assimilation approach using the Norwegian Climate Prediction model (NorCPM), which is based on the Norwegian Earth System Model (NorESM) and the Ensemble Kalman Filter data assimilation method. We assimilate coral data to constrain the ocean component of NorESM. The proxy base model (i.e., the observation operator), is constructed from the instrumental NorCPM reanalysis. We validate the reanalysis produced with online and offline assimilation of coral data for the period 1950–1990. While both approaches can constrain ENSO and the Indian Ocean dipole variability, the online assimilation better captures extreme events. We also compare 1) assimilation in isopycnic versus geopotential depth and 2) the hybrid covariance method where the transient dynamical ensemble is combined with a large static ensemble.

Maud Martet,
Météo – France

New possibilities with AROME 3d-Envar: assimilation of MTG/LI Flash Extend Accumulation (FEA) and direct assimilation of ground-based radar reflectivity

The new 3d-EnVar scheme developed in AROME-France assimilation system allows to add the hydrometeors fields in the control variable. This opens perspectives for the assimilation of observations with information content related to hydrometeors:

- lightning observations from the Lightning Imager (LI) carried on the Meteosat Third Generation (MTG) satellite. These observations are related to the quantity of icy particles, integrated on the vertical.
- reflectivity observed from ground-based radars. Assimilation of relative humidity pseudo-observations retrieved by Bayesian inversion will be replaced by direct assimilation of reflectivity.

These two types of observations are ideal candidates to explore the new possibilities allowed by the 3d-EnVar. Experiments assimilating these additional data show that they provide a better description of convective clouds and precipitation in the first few hours of the forecast.

Cory Martin,
NOAA/NWS/NCEP/EMC – USA

Progress Towards Transition of NCEP's Global Data Assimilation System to JEDI

Currently, the Global Data Assimilation System (GDAS) at the National Centers for Environmental Prediction (NCEP) uses the Gridpoint Statistical Interpolation (GSI) software for computing operational analyses. As part of the transition to the Unified Forecast System (UFS), GSI is slated to be replaced by a Unified Data Assimilation (UDA) framework based on the Joint Effort for Data assimilation Integration (JEDI). Led by the Joint Center for Satellite Data Assimilation (JCSDA), groups from NOAA, NASA, the US Navy and Air Force, and the United Kingdom Met Office all contribute to development and testing of JEDI software. Efforts are underway at NCEP on replacing all GSI-based assimilation components with JEDI-based components in a future implementation of the GDAS. Here we present an overview of initial results from a low-resolution prototype JEDI-based weakly coupled (atmosphere, ocean, sea ice, land, aerosols) cycled GDAS. Technical developments and challenges including workflow and observation processing as well as scientific validation of system components are discussed. In addition to the current progress made, a path forward towards final acceptance testing and transition to operations is described.

Vincent Martinez,
CUNY Hunter College & Graduate Center – USA

Parameter estimation for nonlinear PDEs

In this talk, we will describe a class of algorithms for identifying unknown parameters of nonlinear PDEs. In the absence of observational errors, the convergence of these algorithms can be rigorously established under the assumptions that sufficiently many scales of the solution are observed and that certain non-degeneracy conditions hold, which ensures identifiability of the parameters. This approach to parameter estimation is robust and can be applied not only to recover damping coefficients, but also external driving forces that are unknown apriori. Moreover, it is applicable to a large class of nonlinear equations, including many of those that arise in hydrodynamics, such as the 2D Navier-Stokes equations of incompressible flow, the 2D system for Rayleigh-Benard convection, the 3D primitive equations, or even dispersive-type models such as the 1D Korteweg-de Vries equation or 1D cubic nonlinear Schrodinger equation. We describe the derivation of these algorithms, address their convergence, and showcase the results of several numerical experiments.

Boštjan Melinc,
University of Ljubljana,
Faculty of Mathematics and Physics – Slovenia

With Žiga Zaplotnik

Emulating 3D-Var Data Assimilation using Variational Autoencoder

In numerical weather prediction, data assimilation of atmospheric observations traditionally relies on variational and Kalman filter methods. Here, we propose an alternative machine learning approach with variational autoencoder (VAE) which mimics 3D variational data assimilation (3D-Var), however, the method can be further extended to resemble 4D-Var. The VAE consists of three parts: 1) the input meteorological fields in the gridpoint space enter the encoder, which returns the distributions of the latent vector elements, 2) the latent vector elements are randomly sampled from these distributions, and (3) the latent vector enters the decoder, which finally returns perturbed meteorological fields in the gridpoint space. The 3D-Var cost function is applied to find the latent space vector which optimally assimilates simulated measurements. The background for the assimilation is the encoded short-range forecast based on the persistence model. We observed that the background-error covariance matrix is quasi-diagonal. Furthermore, the analysis increments from the assimilation experiments for the lower-tropospheric temperature indicate that the same set of neural-network-derived basis functions is able to describe both tropical and extratropical background-error covariances.

Richard Ménard,
Environment and Climate Change – Canada

With D. Johnes, M. Bocquet, K. Ide, S. Pathiraja

A critical view on research in chemical data assimilation and inverser modeling. A 5-day workshop summary overview

In march 2023 was held 5-day workshop on Mathematical approaches in chemical data assimilation and inverse modeling. There was 25 participants in person and 12 virtual. The workshop covered mathematical aspects (directly related to CDA, to methods, related to geophysical DA and error estimation), as well as machine learning, parametric Kalman filtering, parameter estimation, and of course, chemical data assimilation and inverse modeling. In this talk we present a critical summary of the main research effort with their specific challenges and and how it grew up specific approaches and realizations. A summary on the coupling with operational numerical weather predictions was also discussed.

Deming Meng,
Nanjing University - China

Synergistic assimilation of cloud and dynamical information based on cloud-dependent background field error covariance

Data assimilation methods based on a three-dimensional variational framework are widely used in weather forecasting operations because they are fast-running and have good assimilation effects. However, the variational scheme has a limitation: it relies on a traditional background error covariance (B) that is highly modelled and does not change with evolving weather situations. This makes it hard to analyze strong convective weather events that have small spatial scales, short lifetimes, strong suddenness, and rapid development and evolution. The static B is even more inadequate for these events. It is well known that clouds and precipitation involve complex physical processes that affect B differently in cloudy and clear areas. Many studies have shown that cloudy areas have larger errors and stronger variable correlations (Montmerle et al., 2010; Michel et al., 2011; Ménétrier et al., 2011). The ensemble Kalman filter assimilation system can produce a flow-dependent B that varies with weather situations, but it also has problems such as limited ensemble samples that make it hard to estimate background errors reasonably or avoid matrix rank deficiency. Therefore, to introduce anisotropic and more realistic weather characteristics into B within a variational framework, this study proposes a new cloud-dependent B strategy by adaptively adjusting the hydrometeor-included B in the cloudy areas according to the cloud index (CI) derived from the satellite-based cloud products. The adjustment coefficient is determined by comparing the error statistics of B for the clear and cloudy areas based on the two-dimensional geographical masks.

To evaluate impacts of cloud-dependent B, two sets of experiments are conducted: The first experiment uses isotropic homogeneous B without adjusting for cloudy areas (EXP_AVE), while another uses cloud-dependent B (EXP_CLD). Both sets of experiments assimilate conventional observational data sets as well as LWP (liquid water path) and IWP (ice water path) produced by NASA LaRC (Langley Research Center). Compared with EXP_AVE experiment, EXP_CLD experiment shows obvious improvements in ETS (equitable threat score) at almost all thresholds except for slight negative contributions (less than 1%) in moderate rain categories at 6-hour or 12-hour forecasts. Average improvement rates for ETS are about 4% for 6-hour forecasts or about 3% for 12-hour forecasts. As thresholds increase, POD (probability of detection) gradually decreases in 6-hour or 12-hour results, but EXP_CLD's POD is generally better than EXP_AVE's. POD has similar improvement rates as ETS, but its values are more obvious, especially in extreme heavy rain cases. Overall, improvement rates for ETS or POD increase gradually with magnitude. This indicates that applying "cloud-dependent" B can more effectively assimilate satellite cloud observations, thus improving precipitation forecast skills, especially for heavy rain or torrential rain.

Furthermore, using "cloud-dependent" assimilation scheme, this paper explores impacts of synergistic assimilation of cloud and dynamic information on typhoon forecasting. Using cross-correlations between hydrometeors or wind fields in cloud-dependent B, clouds and wind fields can affect each other based on implied model constraints when assimilated, making analyses of hydrometeors and dynamic fields more coordinated. For Typhoon Maria (2018), we conducted three sets of experiments: The first experiments only assimilated conventional observations (EXP_Conv), another further assimilated LWP or IWP (EXP_Hydro), or another synergistically assimilated GTS observations, LWP or IWP, or wind field data retrieved by GIIRS/FY4 (EXP_SYN). Results show that compared with EXP_Conv experiment, EXP_Hydro experiment has some improvement on typhoon intensity forecast, but its track forecast has larger errors. Synergistic assimilation of cloud or dynamic information can improve both typhoon intensity and typhoon track forecasts by correcting large-scale environmental fields. In addition, EXP_SYN experiment can provide better precipitation intensity and location, and precipitation scores are significantly improved.

The study confirms that cloud-dependent B allows cloud dependency for multivariate analysis increments, which help to assimilate cloud observations more rationally. Further, this study demonstrates important value of cloud-dependent B in synergistic assimilation of multi-source observational data.

Claire Merker,
MeteoSwiss – Switzerland

Benefit and challenges in assimilating near-surface temperature and humidity observations in complex terrain

The SwissMetNet surface observation network operated by MeteoSwiss provides a dense coverage of near-surface measurements in Switzerland. 2m temperature and humidity measurements from these stations have been introduced in MeteoSwiss' operational data assimilation system in September 2021. This system (KENDA) is based on an ensemble Kalman filter and runs the COSMO model at 1km grid spacing.

In this contribution, we present the benefits and challenges of the 2m temperature and humidity assimilation. The ensemble Kalman filter successfully assimilates the new observations: the temperature and humidity profiles of the near-surface model atmosphere and the representation of fog and low stratus in the COSMO model forecasts are significantly improved. This is in part due to the realistic ensemble correlations provided by the first guess ensemble that takes into account the complex topography. Despite the improvements gained in the forecasts, some challenges remain.

Especially in complex terrain like the Alps in our domain, near-surface measurements are not always representative for their surroundings. This is largely related to the height differences of the model and real-world topography, but also to imperfect parametrizations and other model deficiencies. To mitigate problems arising from height differences in the topography, we discard observations from stations with a height difference of more than 150m. Another issue is the vertical representativeness of near-surface measurements in strong inversion situations. Those situations are quite common on the Swiss Plateau, especially in winter. When the inversion layer is very close to or at the surface, the differences between model data and 2m measurements are only representative for a shallow vertical layer and not valid higher up in the atmosphere. This can cause a degradation of the analysis quality.

We show examples from two years of practical assimilation experience.

Stefano Migliorini,
Met Office – United Kingdom

Assimilation of Transformed Retrievals from IASI radiances at the Met Office

Theoretical studies show that assimilation of Transformed Retrievals (TRs) constructed with the same prior constraints and same observation error as those used for assimilating the same set of radiances leads to equivalent results to assimilating the radiances directly. In this talk we present the results of the first operational-like trial where TRs created from the operational channel selection for IASI radiances are assimilated in addition to the standard set of observations. We show that this trial achieves comparable forecast skill to a control trial where the operational channel selection for IASI radiances are assimilated in addition to the standard set of observations. These results pave the way to the use of TRs to assimilate a much higher information content from advanced satellite sounders into operational NWP models.

Masashi Minamide,
The University of Tokyo / JPL – Japan

Predictability of moist convection through ensemble convective-scale all-sky satellite data assimilation

The development of atmospheric deep moist convection has been a challenging topic for numerical weather prediction, due to its chaotic nature of the development with multi-scale physical interactions. We recently found that greater than 20-km scale (as commonly known as meso-α (2000–200 km) and meso-β (200–20 km) scales) initial features helped to constrain the general location of convective activity with a few hours of lead time, but meso-γ (20–2 km) or even smaller scale features with less than 30-minute lead time were identified to be essential for capturing the spatiotemporal features of individual convection. To examine the potentials of ensemble-based data assimilation in capturing the individual convective development, as well as the subsequent development of severe weather events, we have conducted large ensemble convection-permitting data assimilation experiments with all-sky infrared satellite radiances from the latest-generation geostationary satellites. We found that the greater number of ensembles more effectively suppressed the spurious correlation for convective-scale data assimilation. However, the exact signals of convective development were not clearly captured in covariances even with thousands of ensemble members. These results suggest the potential limitation of the traditional “Eulerian” (i.e. physical grid-based) ensemble approach in convective-scale data assimilation, particularly the potential importance of situation-dependence in addition to the flow-dependence in background covariances to capture and constrain the signals of individual convective activity in numerical weather prediction.

Takemasa Miyoshi,
RIKEN Center for Computational Science – Japan

Toward efficient control of extreme weather events

Since the weather system is chaotic, small differences generally lead to big differences, particularly for storms. We have been exploring controllability through the Control Simulation Experiment (CSE), the same procedure as the well-known Observing Systems Simulation Experiment (OSSE) but with control inputs to the nature run. Our first paper (Miyoshi and Sun, 2022, NPG) proposed the CSE for the first time with proof-of-concept experiments addressing controllability to stay on a chosen regime of the two-regime Lorenz-63 3-variable model. Our second paper (Miyoshi, Sun, and Richard, 2023, NPG) investigated controllability to avoid extreme events with the Lorenz-96 40-variable model. Moreover, we performed CSEs with our global and regional NWP systems for realistic typhoon and heavy rain cases, respectively. In parallel, we applied reinforcement learning to find effective control inputs to a parameter of the Lorenz-63 model. Here, we explored single-sided control inputs to increase the model’s instability, considering realistic situations such that e.g., adding cloud condensation nuclei is more feasible than removing them. In this presentation, we will summarize RIKEN’s activities toward efficient control of extreme weather events.

Takemasa Miyoshi,
RIKEN Center for Computational Science – Japan

PREVENIR: Japan-Argentina Cooperation Project for Heavy Rain and Urban Flood Disaster Prevention

This presentation provides an overall summary of the project PREVENIR and recent activities about data assimilation and numerical weather prediction (NWP) research. PREVENIR is an international cooperation project between Argentina and Japan since 2022 for five years under the Science and Technology Research Partnership for Sustainable Development (SATREPS) program jointly funded by the Japan International Cooperation Agency (JICA) and the Japan Science and Technology Agency (JST). The main goal is to develop an impact-based early warning system for heavy rains and urban floods in Argentina. PREVENIR takes advantage of leading research on Big Data Assimilation (BDA) with the Japan's flagship supercomputer "Fugaku" and its predecessor "K" and develops a total package for disaster prevention, namely, monitoring, quantitative precipitation estimates (QPE), nowcasting, BDA and NWP, hydrological model prediction, warning communications, public education, and capacity building. The total package for disaster prevention will be the first of its kind in Argentina and will provide useful tools and recommendations for the implementation of similar systems in other parts of the world.

Andrew Moore,
University of California Santa Cruz – USA

Weak Constraint 4D-Var Data Assimilation in the Regional Ocean Modeling System (ROMS) using a Saddle-Point Algorithm

The saddle-point formulation of weak constraint 4-dimensional variational (4D-Var) data assimilation has been developed for the Regional Ocean Modeling System (ROMS) and is applied here to the California Current System (CCS). Unlike the conventional primal and dual forcing formulation of weak constraint 4D-Var, the saddle-point formulation can be efficiently parallelized in time which can lead to a substantial increase in efficiency. The performance of the ROMS saddle-point algorithm is assessed here and compared to that of the dual forcing formulation which is the current standard in ROMS. While the rate of convergence of the saddle-point formulation is slower than the dual forcing formulation, the increase in computational speed due to time-parallelization more than compensates for the additional inner-loop iterations required by the saddle-point algorithm in the CCS configuration considered here. Additional increases in performance can be achieved by running the 4D-Var inner-loop iterations at reduced resolution and/or reduced arithmetic precision. The results presented here indicate that in high performance computing environments, the saddle-point formulation of 4D-Var could significantly out-perform the forcing formulation for large data assimilation problems.

Benkiran Mounir,
Mercator-Ocean International - France

Impact of SWOT observations in a global high-resolution analysis and forecasting system

A first attempt was made to quantify the impact of the assimilation of Surface Water Ocean Topography (SWOT) swath altimeter data in a global 1/12° high resolution analysis and forecasting system through a series of Observing System Simulation Experiments (OSSEs). The impact of assimilating data from SWOT and three nadir altimeters was quantified by estimating analysis and forecast error variances for sea surface height (SSH). Wave-number spectra and coherence analyses of SSH errors were also computed. SWOT data will significantly improve the quality of ocean analyses and forecasts. Adding SWOT observations to those of three nadir altimeters globally reduces the variance of SSH and surface velocities in analyses and forecasts by about 30 and 20%, respectively. Improvements are greater in high-latitude regions where space/time coverage of SWOT is much denser. The combination of SWOT data with data from three nadir altimeters provides a better resolution of wavelengths between 50 and 200 km with a more than 40% improvement outside tropical regions with respect to data from three nadir altimeters alone. The study has also highlighted that the impact of using SWOT data is likely to be very different depending on geographical areas. Constraining smaller spatial scales (wavelengths below 100 km) remains challenging as they are also associated with small time scales. Although this is only a first step, the study has demonstrated that SWOT data could be readily assimilated in a global high-resolution analysis and forecasting system with a positive impact at all latitudes and outstanding performances.

Yuka Muto,
Chiba University - Japan

Climatologically augmented local ensemble transform Kalman filter for estimating global precipitation from gauge observations

This study aims to improve the method for estimating global precipitation fields from gauge observations. Unlike the existing estimation methods such as the Optimal Interpolation (OI), we employ the algorithm of the Local Ensemble Transform Kalman Filter that computes analysis precipitation locally by assimilating a subset of surrounding gauge observations. Here we construct an ensemble perturbation matrix from both the flow-dependent and climatological errors for representing a hybrid background error covariance. ERA5's precipitation forecasts are used for the background and the ensembles, and gauge observations from the National Oceanic and Atmospheric Administration (NOAA) are used for the gauge observations. Our estimation showed better agreement with an independent gauge observation product than the NOAA's product based on the OI

Akhilesh S. Nair,
University of Bergen – Norway

With F. Counillon, N. Keenlyside

Improving subseasonal prediction with land surface initialisation in NorCPM

Soil moisture (SM) plays a critical role in land surface dynamics by influencing the exchange of water and energy fluxes between the land and the atmosphere. Its complex spatiotemporal variability, affected by both climate and anthropogenic factors, poses challenges for climate prediction systems. This study emphasises the importance of SM in mid-latitude climate systems and highlights the significance of land surface initialisation for subseasonal-to-seasonal (S2S) predictions. To address this, we introduce the Norwegian Climate Prediction Model Land (NorCPM-Land), a novel global land reanalysis system designed to initialise the land component of the Norwegian Climate Prediction Model (NorCPM). NorCPM-Land assimilates blended SM data from the European Space Agency's Climate Change Initiative (ESA CCI) into a 30-member offline simulation of the Community Land Model (CLM). Assimilating SM data leads to notable improvements in land surface state variability, reducing error by 10.5% when compared to independent SM observations. Moreover, this assimilation approach enhances land surface energy, runoff, and net primary production, highlighting the substantial impact of SM on simulating land surface dynamics.To evaluate the added value of land initialisation for subseasonal predictions, we compare the performance of two sets of hindcasts using NorCPM: one with land initialisation from NorCPM-Land and one without. The hindcasts cover the period from 2000 to 2019 with four initialisations per year. The results demonstrate significant benefits of land initialisation, particularly up to a 3.5-month lead time for soil moisture and a 1.5-month lead time for temperature and precipitation. The most substantial improvements are observed in regions characterised by strong land-atmospheric coupling, including the Central United States, the Sahel, and Central India. Additionally, the hindcasts with NorCPM-Land better capture the severity of high and low-temperature events in parts of Europe, the United States, and Asia, specifically in mid and high latitudes. Overall, our study provides further evidence supporting the significant role of SM content in enhancing the accuracy of S2S predictions.

Saori Nakashita,
Kyoto University – Japan

With T. Enomoto

Observation-space localization methods for the maximum likelihood ensemble filter

Two strategies for localizations in observation space for the maximum likelihood ensemble filter (MLEF) are evaluated in the assimilation of linear observations in a synthetic radiosonde network and additional nonlinear observations such as relative humidity or wind speed with a simplified atmospheric general circulation model. The optimization of the global cost function (LMLEFG) is found to be effective for error reduction in the early cycles and for the use of dense nonlinear observations because information from the surrounding domain is incorporated. By contrast, the improvement is rather limited with the minimization of a local cost function for each domain (LMLEFL) because of little cost reduction. Our results suggest that LMLEFG is suitable for assimilation of observations in various density and types.

Tobias Necker,
University of Vienna – Austria

Empirical optimal vertical localization derived from large ensembles

Localization of error covariances mitigates sampling errors and is crucial for the success of ensemble data assimilation systems. However, finding optimal localization methods, functions, or scales is extremely challenging. To tackle this challenge, we apply a new approach to derive an empirical optimal localization (EOL) from a large ensemble. As presented by Necker et al. 2023, the EOL enables a better understanding of localization requirements and can guide toward improved localization. Our study presents vertical error correlations and optimal localization derived from a 1000-member ensemble simulation (Necker et al. 2020) using a 5-day training period.

In the presentation, we cover both optimal model and observation space vertical localization and discuss: (1) Vertical error correlations and EOL estimates for different variables and settings. (2) The effect and behavior of the EOL compared to common localization approaches, such as distance-dependent localization with a Gaspari and Cohn function. (3) How to achieve positive definiteness of localization. (4) Optimal vertical observation space localization of cloudy infrared and visible satellite observations.

Cloudy infrared and visible satellite observations carry valuable information for the assimilation and prediction of clouds and convective processes. However, proper localization of error covariances between non-local satellite observations and state space is non-trivial and has yet to be studied for the visible spectral range. We apply the EOL approach to evaluate requirements for optimal vertical localization for different variables and spectral channels. Furthermore, we investigate the variability of vertical localization in convection-permitting NWP simulations, which suggests an advantage of using adaptive situation-dependent localization approaches.

Annika Oertel,
Karlsruhe Institute of Technology (KIT) – Germany

The 'Swabian MOSES' field campaign – a testbed for the near real-time assimilation of campaign observations

Forecasting summer-time convective storms in Europe remains a challenge, although convective-scale data assimilation and ensemble forecasting systems provide high-resolution forecasts of such extremes. A key question in this context is whether assimilation of suitable high-resolution observations can improve convective-scale numerical weather prediction (NWP). The 'Swabian MOSES' field campaign from June to August 2023 in Southern Germany provides spatially distributed, high-resolution observations of the dynamics and thermodynamics of the lower troposphere during convective conditions. We aim at assimilating in-situ and remote-sensing campaign observations in near real-time in the quasi-operational convective-scale NWP system of the German Weather Service. Here we provide an overview of assimilated observations and show first results of the impact of the campaign observations.

Andrea Orlandi,
Consorzio LaMMA – Italy

Studies on breeding-driven adaptive Data Assimilation applied to Low Order Models

Low Order Models (LOMs) allow to explore many relevant issues connected with numerical weather and ocean prediction, thanks to their low computational and memory requirements. In this context, we investigated breeding-driven adaptive Data Assimilation (DA) strategies with different DA approaches, exploiting an open-source suite in Python. Two LOMs have been considered: Lorenz '63 (L63) and Lorenz-Emanuel '96 (LE96). In a first phase, mainly on L63, an "operational NWP-like" framework has been implemented, and different assimilation-prediction cycles have been tested. Adaptive DA strategies have been investigated mainly with the spatially distributed LE96. Best performances resulted for the Ensemble Kalman Filter approach, by adding the constraint of not assimilating over the same site in the following time-steps. This suggests the opportunity of including, in further studies, a space-time correlation analysis to improve the adaptive assimilation-sites selection strategy.

Alberto Ortolani,
CNR-IBE; Consorzio LaMMA – Italy

With A. Carrassi, F. Fabiano, U. Cortesi,
J. von Hardenberg, G. Masiello, M. Ridolfi, S. Melani,
S. della Fera, T. Maestri, L. Palchetti

MC-FORUM: the ASI-funded project for the meteorological and climatological exploitation of FORUM

Satellite observations play a major role in operational meteorology, providing homogeneous and global data for the Earth's atmosphere and surface, exploited through Data Assimilation (DA) to produce estimates of the Earth system state. The FORUM (Far-infrared-Outgoing-Radiation Understanding and Monitoring) mission is the ESA 9th Earth Explorer (EE9), scheduled for launch in 2027. FORUM will collect spectrally resolved radiance fields in the Far InfraRed (FIR) and Mid InfraRed part of spectrum, spanning the 100 to 1600 cm-1 band with 0.5 cm-1 un-apodised resolution. The FIR wavelengths are currently unexplored from space despite they constitute a large fraction of the planet's outgoing longwave radiation (OLR). This spectrum embeds the signatures of several climate forcings and related feedbacks, also being highly sensitive to upper tropospheric water vapor and cirrus clouds, and it is crucial for assessing the Earth's radiation balance.

MC-FORUM (Meteo and Climate exploitation of FORUM), a two-year project funded by the Italyn Space Agency (ASI), will start at the end of 2023, with the primary objective of developing tools and skills to exploit FORUM in the meteorological and climate fields. It will study the impact of FORUM data on different spatiotemporal scales and different DA techniques, variational and ensemble-based Kalman filter-like. It will also develop tools for exploiting FORUM measurements for diagnostics and validation of global climate models, for enabling the study of new parameterizations, and the potential of FORUM in the evaluation of radiative forcings and climatic feedbacks.

Mao Ouyang,
Chiba University – Japan

A hybrid data assimilation with reservoir computing to advance the control simulation experiment

Controlling the weather is attractive and has achieved promising progress in small-scale control simulation experiments (CSEs). Two open questions, however, need to be solved for advancing the CSE in high-dimensional systems: one is the computational expensive ensemble forecast for finding the desired ensembles; and the other one is the accuracy of analysis ensembles for better representation of the true trajectory. This research proposes surrogating the physics-based numerical model by the reservoir computing (RC) for reducing the computational cost in CSE, and augmenting ensemble forecasts to improve analysis. Our experiments using Lorenz model show improvements in both DA and CSE, which hint possible applications of RC-based surrogate models to advance the CSE with high-dimensional models such as numerical weather prediction models.

Mao Ouyang,
Chiba University – Japan

Producing balanced analysis ensemble in local particle filter using a differential resampling method

Local particle filter (LPF) is an efficient approach for non-Gaussian data assimilation problems such as vortex alignment through approximating the posterior distributions with the weighted samples. This study investigates better resampling methods for the ensemble-transform-matrix-based LPF in which differentiable transform matrix is important to produce balanced analysis ensemble. Here we investigate the entropy-regularized optimal transport (OT) method, known as Sinkhorn algorithm, for the resampling of the LPF. We conduct a series of LPF experiments by the Rankine vortex with four prior distributions, and two radar observations. Results demonstrate that resampling with entropy-regularized OT produced much more balanced analysis ensemble than the conventional OT owing to differentiability of the transform matrix.

Nishant Panda,
Los Alamos National Lab – USA

Using Koopman and Perron Frobenius Operators For Non-Linear Data Assimilation

From space weather, to climate, considerable efforts are being made to improve the accuracy of the physics models. These models have complex interactions between physical phenomena evolving at different scales. Reliability of such multi-physics models remains a grand challenge. For example, an important research question is how will precipitation evolve over the next 40 years? When physical phenomena interact, the predictions from such models get polluted and the inherent uncertainties are compounded, thus reducing their reliability. To tackle this problem, we look at a novel approach for uncertainty quantification (UQ) in coupled systems by learning the mathematical operators, called the Perron-Frobenius and Koopman operators, that describe the propagation of uncertainties in such models. In particular, we will show how these operators can be used for data assimilation under the 4dVar framework. We also demonstrate a Bayesian framework for data assimilation using the Perron Frobenius operators.

Ivo Pasmans,
University of Reading – United Kingdom

Tailoring data assimilation to discontinuous Galerkin models

Discontinuous Galerkin (DG) methods are rapidly gaining popularity in the geophysical community. In these methods the model solution in each grid cell is approximated as a linear combination of basis functions. Ensemble data assimilation (EnDA) aims to bring the model closer to the truth by combining it with observations using error statistics estimated from an ensemble of model runs. It is known to suffer from several well-documented issues. We have tested whether the DG structure can be exploited to address the following three issues: 1) reduce the need for observation thinning, 2) reduce errors in gradients, 3) produce a more accurate localised ensemble covariance. Using an idealised test setup it is found that strong reduction in error can be realised, especially for high DG orders. However, this does not result in reduction of the error in the gradients. The DG basis is found to be expedient for scale-dependent localisation resulting in an ensemble covariance that is closer to the truth than one created using conventional, non-scale dependent localisation.

Ivo Pasmans,
University of Reading – United Kingdom

Ensemble Kalman filter in latent space using a variational autoencoder pair

Currently, neXtSIMDG, a new sea ice model, is under development and a pairing with a data assimilation system is envisioned. Due to the presence of highly-nonlinear terms in the model equations, bounds on the values of the model fields and plasticity constraints, errors in the model are non-Gaussian. This violates the assumptions underlying the ensemble Kalman filter (EnKF). In this work we propose overcoming this problem by applying EnKF in the latent spaces of two variational autoencoders (VAEs). At the analysis time the VAEs are trained on the model background ensemble and its associated ensemble of innovations respectively. By construction these ensembles follow a normal distribution in latent space and from them a latent analysis ensemble can be produced using EnKF. Using this ensemble and the VAE, the analysis ensemble can be sampled in the model space and propagated forward to the next analysis time by the model. A more detailed outline of this scheme and any preliminary results will be shown.

Tim Payne,
Met Office – United Kingdom

A Hybrid Differential-Ensemble Linear Forecast Model for 4D-Var

For many years the most effective way to assimilate data into a numerical weather prediction model has been four dimensional variational assimilation (4D-Var). One of the difficulties with 4D-Var has been the development and maintenance of the linear model, approximately tangent linear to the full model, for evolving perturbations. The linear model is particularly problematic for physical parameterisations. We present a new method, the hybrid tangent linear model, that solves many of these long-standing issues.

Reference

T.J. Payne, "A Hybrid Differential-Ensemble Linear Forecast Model for 4D-Var", Monthly Weather Review (2021)

Mariam Petrosyan,
Yerevan State University – Republic of Armenia

Assessing Climate Risk in an Urban Environment: Integrating Remote Sensing Data for a Case Study of Yerevan City

This study conducts a comprehensive climate risk analysis in Yerevan City, utilizing remote sensing data and advanced geospatial techniques. The objective of the study is to comprehensively assess the vulnerability of the urban environment to climate-related hazards. Through the integration of high-resolution satellite imagery and geographic information systems (GIS), the research meticulously maps and analyzes land cover changes, temperature fluctuations, and precipitation patterns.

The remote sensing data provides a valuable foundation for characterizing the spatial and temporal dynamics of climate risks. Through extensive spatial analysis, the study identifies and evaluates areas within Yerevan City that are particularly susceptible to various climate-related threats, including extreme weather events. The findings contribute to a deeper understanding of the city's vulnerability profile, aiding in the formulation of effective strategies for urban planning and disaster management.

By harnessing the power of remote sensing data, this research advances the field of climate risk analysis by providing a robust methodology and empirical evidence for the identification and assessment of climate risks in urban areas. The results offer valuable insights for policymakers, urban planners, and other stakeholders, enabling them to make informed decisions and implement targeted measures to enhance the city's resilience to climate change impacts.

The scientific rigor, comprehensive methodology, and practical implications of this study underscore the importance of incorporating remote sensing data in climate risk analysis. This article serves as a significant contribution to the scientific community, showcasing the efficacy of remote sensing techniques for climate risk assessment and providing a replicable framework that can be applied to other cities and regions facing similar challenges.

Flavia Pinheiro,
Oceanographic Modeling and Observation Network (REMO – Petrobras) – Brasil

With P.J. van Leeuwen

Data assimilation for a two-layer quasi-geostrophic model using an Ensemble Synchronization Particle Filter

Particle filters have increasingly shown promising data assimilation solutions to the challenges posed by strongly nonlinear systems. Differently from other existing data assimilation methods and techniques, particle filters make use of state probability density functions (pdfs), by a discrete set of particles.
The concept of proposal density freedom allows a particle filter to work in high-dimensional systems.
In this work, we use the ensemble synchronization framework as a proposal density. The main goal is to synchronize the model with the true evolution of a system using a one-way coupling, via the observations. An extra term is added to the model equations in order to control the growth of instabilities in the synchronization manifold.

To this end, an efficient backward-forward ensemble-based synchronization scheme is used as a proposal density in the improved version of the implicit equal-weights particle filter (IEWPF), which avoids filter degeneracy by construction. We call this methodology an Ensemble Synchronization Particle Filter (ESPF), where synchronization has time to influence the particle trajectories, leading to better filter performance. We analyze the effect of this technique in the two-layer quasi-geostrophic model. By using only 10 members, we observe they all converge very closely to the truth until the end of the window for the streamfunction fields, showing promising results which suggest the Ensemble Synchronization Particle Filter is a promising solution for high-dimensional nonlinear problems in the geosciences, such as numerical weather prediction.

Naila Raboudi,
King Abdullah University of Science and Technology (KAUST) – Saudi Arabia

With B. Ait-El-Fquih and I. Hoteit

Ensemble Kalman Smoothing with Exact Second-Order Observation Perturbations Sampling For Ocean Reanalyses

The stochastic ensemble Kalman filter (EnKF) undersamples the observational errors when the ensemble size is smaller than the rank of the observstional error covariance, which was shown to limit the ensemble spread. Deterministic EnKFs are therefore more commonly used for such applications although they may distort the features of the forecast ensemble distribution by virtue of their deterministic sampling schemes. The observational errors undersampling issue becomes more pronounced in a smoothing context due to the extra updates with the future data. In preparation of generating a high resolution regional ocean reanalysis for the Red Se and the Arabian Gulf, we introduce here a new fixed-lag ensemble Kalman Smoother (FL-EnKS) with exact second-order observation perturbations sampling that resolves the observational error rank while requiring only minor modifications to an exisiting serial FL-EnKS code. The proposed EnKS is evaluated against different EnKS variants and was found to outperform deterministic EnKSs even with relatively small ensembles. We present numerical results from a simple Lorenz-96 model and a reaslistic general circulaiton ocean model.

Brett Raczka,
National Center for Atmospheric Research – USA

With Brett Raczka, Xueli Huo, Andrew Fox, Moha Gharamti, Daniel Hagan, Anthony Holmes, Ying Sun, Lewis Kunik, John Lin, Jeffrey Anderson

Improving Forecasts of Land Surface Carbon Cycling using the Data Assimilation Research Testbed (DART)

The impact of the land upon net carbon uptake from the atmosphere is particularly important for climate projections, yet the representation of carbon cycling in land surface models is often limited because of errors related to initial and boundary conditions, model structure, and parameters. Ensemble data assimilation techniques combined with an expanding network of earth system observations present an opportunity to reduce these errors and improve simulations. First, we describe an approach using NCAR's Community Land Model with DART to constrain carbon cycling simulations through observations of above-ground biomass. Next, we present preliminary results upon the added benefit of including both observations of snow cover and solar-induced fluorescence. Finally, we discuss to what extent additional earth system observations are required, and what role parameter estimation can play to further improve model skill.

Hendrik Reich,
Deutscher Wetterdienst – Germany

With C. Schraff, A. Cress, H. Anlauf

Use of high-density Mode-S aircraft observations in ICON regional and global data assimilation system

Mode-S EHS (Enhances Surveillance) aircraft data of wind and temperature are derived from data collected routinely from all aircraft for the purpose of air traffic control surveillance.

At DWD, Mode-S data have been used operationally in the KENDA data assimilation system for the limited-area model ICON-D2 since 2017. Since October 2021, KNMI has been providing Mode-S data covering a much wider area in Europe, with much higher density.

We present results of assimilation of the new high-density Mode-S data into DWD's regional and global data assimilation systems, which leads to RMSE reductions of up to 8 % for wind and temperature in the upper troposphere. To cope with the high amount of data, a two-step thinning procedure has been developed. Furthermore, the influence of the improved boundary conditions from the global system on the local model is shown.

Maria Reinhardt,
Deutscher Wetterdienst – Germany

Intelligent Camera Cloud Operators for Convective Scale Numerical Weather Prediction

We present an innovational way of assimilating observations of clouds into the weather forecasting model for regional scale: ICON-D2 (ICOsahedral Nonhydrostatic), which is operated by the German Weather Service (Deutscher Wetterdienst, DWD).

A feed forward convolutional neural network and a u-net have been trained to detect clouds in pictures. We use photographs taken by cameras pointed towards the sky and extract the information about clouds by applying the aforementioned networks. The results of both networks are grayscale picture, in which each pixel has a value between 0 and 1, describing the probability of the pixel belonging to a cloud. The performance of both neural networks will be compared.

By averaging over a certain section of the picture one gets a value for the cloud cover of that region. To build the forward operator, which maps an ICON model state into the observation space, we construct a three dimensional grid in space from the camera point of view and interpolate the ICON model variables onto this grid. We model the pixels of the picture as rays, originating at the camera location and take the maximum interpolated cloud cover (CLC) along each ray. CLC is a diagnostic variable of an ICON model state describing the probability of the cloud coverage within the respective grid box.

Monitoring experiments to compare the observations and model equivalents over time have been conducted as well as data assimilation experiments over different time intervals in different months and with different camera locations and amounts of used cameras. The results are presented and evaluated using verification statistics.

Juan Restrepo,
Oak Ridge National Laboratory – USA

With J. Krotz

A Superior Wave Estimator: The Dynamic Likelihood Filter

The dynamic likelihood filter achieves superior results in estimating wave speeds and amplitudes by exploiting the finite-speed of propagation of information in wave problems. We will show that the difference between the dynamic likelihood and other estimation methods is particularly evident in cases where observations are scarce. Data sparsity is common in physical applications. We will also present preliminary results to problems related to transport and advection/diffusion, in the small dispersion/diffusion limit and our plans on how to handle epistemic errors in estimation via machine learning strategies.

Laura Risley,
University of Reading – United Kingdom

On the choice of velocity variables for variational ocean data assimilation

When specifying the background error covariance matrix in variational data assimilation (DA), model variables are transformed into control variables that can be assumed to be approximately uncorrelated. This process is known as the control variable transform. With a view of assimilating ocean surface currents from satellites, we seek to develop a transformation that decorrelates horizontal velocity. Helmholtz theorem can be used to separate horizontal velocity into nondivergent and irrotational components. The transformed variables can then be defined as relative vorticity and horizontal divergence, as is typically done in atmospheric DA. In the NEMOVAR ocean data assimilation system, the horizontal velocity control variables are taken to be the ageostrophic components of the vector. Here, we investigate using alternative velocity control variables based on the ideas above.

Vanya Romanova,
Deutscher Wetterdienst - Germany

Integrated atmosphere-ocean-land data assimilation for climate analysis and seasonal predictions

The DWD ICON-NWP model is used to configure a seamless operational mode for long-term analysis and seasonal predictions. Targeting an integrated data assimilation of the coupled model components (ICON-Atmosphere, ICON-Ocean and JSBACH) and following different assimilation approaches, the model state estimates are simultaneously constrained to the available observational data. The atmosphere component uses the nudging approach for binding the temperature, sea level pressure and velocities to the six-hourly output to the ERA5 reanalysis., The ocean module uses the PDAF Ensemble Kalman Filter (Nerger and Hiller, 2013, Brune et al., 2020) to assimilate ocean temperature and salinity data from the EN4 observational data set. Finally, a land surface analysis will be developed by using a 2D-Var method for assimilating the surfaces variables of snow, soil moisture and in a second step leaf area index from satellite observations.
The assimilation of almost all climate observations on the Earth into a fully coupled climate system is an innovative approach and is a task requiring extensive efforts and resources. As a result, we expect improvement of the predictability skill of the seasonal to climate forecasts. In this study we present our preliminary results and give an outlook for the future plans and strategies.

Reference

Nerger, L., & Hiller, W. (2013). Software for ensemble-based data assimilation systems—Implementation strategies and scalability. Computers and Geosciences,55, 110—118
Brune, S, Baehr, J. (2020). Preserving the coupled atmosphere—ocean feedback in initializations of decadal climate predictions. WIREs Clim Change.; 11:e637. https://doi.org/10.1002/wcc.637

Luca Rovai,
CNR-IBE - Italy

With A.Antonini, L. Fibbi, V.Capecchi, S. Melani, A. Orlandi, A. Ortolani, B. Gozzini

Testing a WRF-based modelling chain for operational forecasting under different data assimilation inputs

Recently, the Italyn meteorological community spent many efforts to implement effective numerical tools for better predicting the extreme phenomena hitting the Mediterranean area. In the present work, we show the results obtained in running sensitivity tests applied to a modelling chain operating at the LaMMA Consortium and based on the Weather Research and Forecasting (WRF) model and its data assimilation (DA) package. Data ingested in the WRF numerical experiments are derived from: automatic weather stations at the regional and national scale, GNSS-meteo networks, radar, and additional sensors managed by LaMMA itself, over the sea too. Such sensitivity studies are aimed at testing the impact (whether beneficial or not) of the various observational datasets, the role of the tuning parameters available in the WRF-DA package and the added valued carried by using background error covariance matrix computed over annual or seasonal time periods. The analysis is made on a single (very critical) case, the 2022 Marche flooding event, but exploiting different initialisations (in ensemble fashion), to evaluate the forecasting accuracy under different initial and DA conditions. The time range of interest is the one of the short-term forecasting (3 -6 hours), according to the spatial domain of model simulation and assimilated data.

Yvonne Ruckstuhl,
KUEI – Germany

Combining the Stochastic Galerkin with data assimilation for parameter estimation

Using data assimilation (DA) for parameter estimation in numerical weather prediction by means of state space augmentation is well covered in literature. It is well accepted that the highly non-linear effect of parameters on model variables can cause sub-optimal performance of standard DA algorithms like the Ensemble Kalman Filter (EnKF), which assume Gaussian error statistics. Various DA algorithms that alleviate or discard this Gaussian assumption are available, but all require accurate higher order error statistics of the model state and parameters. A common approach to approximate these error statistics is to run an ensemble. However, as numerical weather prediction models are computationally expensive, only a modest ensemble size is affordable. The resulting error statistics are therefore contaminated with large sampling errors, rendering the use of higher order statistical moments for DA impractical.

Here, we use the Stochastic Galerkin (SG) based on the generalized polynomial chaos expansion in the stochastic space to approximate the relevant error statistics. The SG is more efficient and accurate in representing uncertainties than an ensemble when the number of independent random state elements is small. We therefore use a standard ensemble to sample the error statistics of the state variables, and apply the SG to obtain accurate, full error statistics involving the parameter. The availability of accurate full error statistics opens the door to DA algorithms that can handle non-Gaussianity, like the quadratic filter or particle filters. We compare the proposed hybrid of SG and DA to the EnKF for parameter estimation on a 2D model that consists of a dynamical core, coupled via a source term to a cloud model for warm cloud physics.

Hamza Ruzayqat,
King Abdullah University of Science and Technology (KAUST) – Saudi Arabia

A Class of Filtering Problems with Unknown Spatial Observations

We address the complex problem of high-dimensional stochastic filtering in state-space models with continuous-time nonlinear dynamics and discrete-time observations, where analytical solutions are often unavailable and numerical approximation methods can be prohibitively expensive as the dimension of the hidden state increases. Although the standard particle filter is a widely used method for low-dimensional nonlinear models, it suffers from weight degeneracy in high dimensions, requiring an exponential increase in the number of samples to address the issue. To overcome this challenge, we revisit a lesser-known and exact computational methodology from Centanni, S. & Minozzo, M. (2006a) designed for filtering of point-processes. We adapt the methodology to propose an algorithm that is capable of handling observations with random spatial locations that are dependent on the hidden state. Of course the algorithm can be still used for models with observations of known deterministic spatial location. The new algorithm is demonstrated to be stable in time and high dimensions, and is tested against a high-dimensional rotating shallow water model (of the order 10^4) with real and synthetic observational data from ocean drifters. We also compare our method to existing ensemble methods where we test them on a linear-Gaussian state-space model and demonstrate a significant improvement in speed and accuracy.

Sami Saarinen,
HPC-consulting Ltd – Finland

SPREADS d4o:– a database for observations

A novel database interface called d4o (in lowercase, and pronounced as dee-for-ou) has been designed for the SPREADS (Dart) code. It is a Fortran interface that accesses underlying SQLite version 3 databases. Observational data has been split between observation report (header) and datum (body) tables. In addition first guess and final departures per ensemble are linked into a separate ensemble table. This format substitutes the Dart observation sequence file, which is both space wasting and does not allow to use data searches through SQL-language. The d4o has also potential to inquire observation information on the fly, by any MPI-task using SQL-queries and propagating information (a matrix) back to the Fortran code. At the moment this feature is less usable due to the complex way observations are expressed in the SPREADS code. As a result all tasks have a copy of all observations! But data parallelization is under way: currently MPI-tasks on each running node can already access just a single copy of the incoming observations per node already saving a huge amounts of memory. At the end of the day we should have a restructured observation organization in the SPREADS allowing to handle 100's of millions of observations across 100's of ensemble members per each 6h assimilation cycle. A bonus feature includes ability to perform post processing & create graphical displays from SQLite databases directly using standard Python interfaces.

Takumi Saito,
Chiba University – Japan

Designing Effective Observing Network for Data Assimilation based on Sparse Sensor Placement Method

Data assimilation (DA) is important in geophysics for combining forecasts states and observations. Despite extensive research on the impact of assimilated observations, few studies have investigated optimal observation placement. Recently, sparse sensor placement (SSP) was proposed in informatics to determine observation placements so as to maximize the determinant of Fisher's information matrix. In order to improve state estimations by DA, this study extends the SSP to identify observation placements that reduce low-dimensional analysis error covariance spanned by left singular vectors of ensemble perturbations. Experiments using sea surface temperature (SST) data revealed that the proposed method provides more accurate spatial patterns of SST from limited observations than the original SSP methods, especially in cases with less training data.

Siva Reddy Sanikommu,
King Abdullah University of Science and Technology – Saudi Arabia

With Mohamad El Gharamti, Yixin Wang, Matthew Mazloff, Ariane Verdy, George Krokos, Rui Sun, Aneesh Subramanian, Benjamin K. Johnson, Angela Kuhn Cordova, Bruce Cornuelle, and Ibrahim Hoteit

A Hybrid Ensemble Biogeochemical Data Assimilation System for the Red Sea: Development, Implementation and Evaluation

A Hybrid ensemble system is implemented for data assimilation (DA) into coupled physical-biogeochemical ocean model of the Red Sea. The system comprises a Massachusetts Institute of Technology general circulation model (MITgcm) coupled with the Nitrogen-version of the Biologeochemistry, Light, Iron, Nutrients and Gases (N-BLING) model, both configured at 4km-resolution. The assimilation is based on the Data Assimilation Research Testbed (DART) and combines a time-varying ensemble generated using the Ensemble Adjustment Kalman filter (EAKF) with a pre-selected quasi-static (monthly varying) ensemble. The system is designed to assimilate observations of both physical (satellite sea surface temperature, altimeter sea surface height, and in situ temperature and salinity) and biological (satellite chlorophyll) variables. Two different assimilation experimental settings are tested: (1) Weakly coupled DA in which the physical and biological observations only update their respective states, and (2) Strongly coupled DA in which both the physical and biological observations are used to update both physical and biology states. Sensitivity experiments are conducted to assess the relative impact of assimilating physical and biology observations. The state estimates are evaluated against independent in situ Glider observations of temperature, salinity, chlorophyll, and oxygen. The results indicate that the strongly coupled DA generally performs better than the weakly coupled DA. The improvements are significant particularly in the subsurface layers. We further conducted identical twin experiments using strongly coupled DA with and without assimilating satellite chlorophyll observations to confirm the positive impact of assimilating chlorophyll observations on the estimation of biogeochemical fields.

Yohei Sawada,
University of Tokyo – Japan

With L. Duc

An efficient estimation of spatio-temporally distributed parameters in dynamic models by an ensemble Kalman filter

The accuracy of Earth system models is compromised by unknown and/or unresolved dynamics, making the quantification of systematic model errors essential. While a model parameter estimation, which allows parameters to change spatio-temporally, shows promise in quantifying and mitigating systematic model errors, the estimation of the spatio-temporally distributed model parameters has been practically challenging. Here we present an efficient and practical method to estimate time-varying parameters in high-dimensional spaces. In our proposed method, Hybrid Offline and Online Parameter Estimation with ensemble Kalman filtering (HOOPE-EnKF), model parameters estimated by EnKF are constrained by results of offline batch optimization, in which the posterior distribution of model parameters is obtained by comparing simulated and observed climatological variables. HOOPE-EnKF outperforms the original EnKF in a synthetic experiment using a two-scale Lorenz96 model. One advantage of HOOPE-EnKF over traditional EnKFs is that its performance is not greatly affected by inflation factors for model parameters, thus eliminating the need for extensive tuning of inflation factors. We thoroughly discuss the potential of HOOPE-EnKF as a practical method for improving parameterizations of process-based models and prediction in real-world applications such as numerical weather prediction.

Bernd Schalge,
Deutscher Wetterdienst – Germany

Towards a coupled convection permitting reanalysis for the European CORDEX domain

For this study the Atmospheric model ICON at 3km resolution is combined with a DA system based on the LETKF. This system is subsequently coupled to the Community Land Model (eCLM). Even though computing resources are much more plentiful today, some compromises are unavoidable to run a full 30-year reanalysis at this scale.
The first issue regards the ensemble size. Usually an ensemble is often run at a coarser resolution to allow for more members, but since this would fall in to the convective gray zone the ensemble has the same resolution as the deterministic run. Furthermore, there is a limitation on ensemble size due to the availability of boundary data for the long period required. It is investigated weather running a 10 member ensemble is sufficient to achieve improved results compared to earlier products. The impact of the large difference in resolution between boundary forcing from the ERA5 ensemble and this simulation is also investigated.

In addition, other issues are tackled as well, for instance optimal settings for shallow convection parameterization or the impact on boundary layer processes by the replacement of the land model.

Leonhard Scheck,
Deutscher Wetterdienst – Germany,
LMU Munich – Germany

A neural network-based forward operator for assimilating near-infrared satellite images

Imagers on geostationary and polar orbiting satellites provide high resolution information on clouds and aerosols that is valuable for numerical weather prediction (NWP). While using thermal infrared images for data assimilation and model evaluation is well-established, and the operational use of visible channels is emerging, near-infrared channels are not yet directly assimilated, mostly because fast and accurate forward operators have become available only very recently. In particular the 1.6 micron channel available on many satellites is interesting for NWP, as this channel is not only much more sensitive to particle radii than visible channels but allows also for distinguishing water from ice clouds. As multiple scattering complicates the solution of radiative transfer problems for near-infrared channels, standard radiative transfer methods are too slow for operational purposes. Here we discuss the design of a fast neural network-based method for generating synthetic images in the 1.6 micron channel that is based on earlier work for visible channels. The method has been implemented in the latest version of the RTTOV forward operator package. Methods to reduce the computational effort and the errors of the method with respect to reference radiative transfer solutions are discussed. We show that the method works well not only for 1.6 microns but for many other solar channels. Reflectance and reflectance error statistics for regional and global NWP model runs will be shown. Finally, we discuss the impact of different choices for the effective water and ice cloud particle radii, including radii derived from a two-moment microphysics scheme.

Nora Schenk,
Deutscher Wetterdienst – Germany

Coupled Data Assimilation at DWD: Development of an Ocean Data Assimilation System

The new project "Earth System Modelling at the Weather scale" (ESM-W) by DWD in cooperation with GeoInfoDienst BW aims to develop a coupled ocean-atmosphere forecasting system based on the ocean model ICON-O and the atmospheric model ICON-NWP. The initialization of the coupled model is based on the operational data assimilation system of DWD (hybrid EnVar+LETKF) for the atmosphere combined with a newly developed 3DVar method for the ocean applied to ARGO float data as well as sea surface products derived from satellite measurements. Following a weakly-coupled assimilation approach, both states are presently treated separately. We describe the components of the coupled system with focus on ocean data assimilation. This serves as preparation for an ensemble-based — and later a strongly coupled — data assimilation framework.

Francine Schevenhoven,
Geophysical Institute, University of Bergen, and Bjerknes Centre for Climate Research, Bergen – Norway

Supermodelling: improving predictions with an interactive ensemble

Supermodelling is a step forward in the multi-model ensemble approach. Instead of combining model data after the simulations are completed, the models in a supermodel exchange information during their simulation, using techniques from data assimilation. Individual model errors can then be reduced at an early stage before they grow and affect other regions or variables. The information exchange, resulting in synchronized models, is trained based on past observations. In this presentation, we introduce supermodelling and demonstrate its potential to improve long-standing biases in climate models. We also focus on the (technical) challenges of the supermodelling technique. Results will be shown based on the combination of the latest versions of the Community Atmosphere Models (CAM) into a supermodel.

Stefano Serafin,
University of Vienna – Austria

With Valentina Hutter, Stefano Serafin, Martin Weissmann, Daniel Leuenberger

Probabilistic observation pre-processing for ensemble-based data assimilation: An application to surface temperature observations in Alpine terrain

Most operational data assimilation algorithms are designed under the assumption that the probability distribution of forecast errors is Gaussian and has zero mean, i.e., there is no systematic error. Over mountainous terrain, oversimplified parameterization schemes (in particular for surface-layer and boundary-layer exchange), combined with large deviations between the modeled and true orography, lead to pronounced model errors. These circumstances ultimately lead to inaccurate estimates of forecast uncertainty and suboptimal use of observations in data assimilation. In this work, we use an ensemble post-processing method known as SAMOS (Standardized Anomaly Model Output Statistics) to correct systematic deviations between observations and first guesses prior to the data assimilation. A 5-year archive of observations and first guess from the MeteoSwiss COSMO-1E analysis ensemble is analyzed, with emphasis on 2-m temperature forecasts. We demonstrate that the SAMOS correction reduces systematic errors both at mountain-top and valley sites, as well as their seasonal and diurnal variations. In addition, SAMOS processing corrects the standard deviation of ensemble first guess. The procedure ultimately results in a better estimate of the forecast uncertainty and in a very nearly Gaussian distribution of forecast errors. Based on these results, we expect that the implementation of SAMOS processing of temperature innovations in the data assimilation cycle can lead to fewer observations being rejected and to an increased weight of the assimilated observations in the analysis.

Daiya Shiojiri,
Chiba University – Japan

Introducing data-driven sparse sensor placement to determine rain gauge locations

This study introduces the data-driven sparse sensor placement method (SSP) to determine rain gauge locations so that spatiotemporal rain fields can be estimated with a smaller number of observations. The SSP determines the rain gauge locations using dominant modes, which are obtained by linear decomposition of spatiotemporal precipitation data. We also propose to estimate spatial precipitation fields from rain gauge data using the algorithm of local ensemble transform Kalman filter (LETKF). Through evaluations with reference to radar-analyzed precipitation, we found that the proposed method can improve the estimation of precipitation compared to the operational rain gauge network known as AMeDAS. We also found that applying a nonlinear decomposition known as non-negative matrix factorization was beneficial for SSP of rain gauges.

Daiya Shiojiri,
Chiba University – Japan

Investigating appropriate inflation methods for assimilating soil moisture data into a land surface model

This study investigates an appropriate inflation strategy for stabilizing our land data assimilation system (LDAS), which incorporates the Integrated Land Simulator and the local ensemble transform Kalman filter. Using the LDAS, we conducted a series of observing system simulation experiments with assimilation of surface soil moisture. Here, we apply the relaxation to prior perturbation (RTPP) method for covariance inflation in addition to the conventionally-used perturbed forcing method. Our results indicate that using the RTPP led to greater stabilization of the LDAS system with reduction of perturbations in the forcing data. Assimilating the surface soil moisture successfully improve the soil moisture, even for unobserved deep layers.

Anna Shlyaeva,
UCAR – USA

With Y. Tremolet, B. Menetrier, C. Gas, F. Hebert, C. Sampson, S. Vahl

Developing and using JEDI for Earth system prediction

Joint Effort for Data assimilation Integration (JEDI) is developed at the JCSDA in collaboration with NOAA, NASA, US Navy, US Air Force, United Kingdom Met Office and NCAR. The goal of the project is developing a cutting-edge data assimilation system that can be used with different prediction systems including atmosphere, land, ocean, sea-ice, atmospheric composition, and coupled Earth system prediction systems. JEDI system provides a capability to use any of the wide set of generic forward operators, background and observation error covariance models and other data assimilation components. This presentation will show the current status and the future plans for the JEDI data assimilation capabilities, including the data assimilation algorithms, background error covariance models and generic development that enables data assimilation for coupled prediction systems with components interfaced to JEDI.

Ehouarn Simon,
Université de Toulouse, INP, IRIT – France

A kernel extension of the Ensemble Transform Kalman Filter

Data assimilation methods are mainly based on the Bayesian formulation of the estimation problem. For cost and feasibility reasons, this formulation is usually approximated by Gaussian assumptions on the distribution of model variables, observations and errors. However, when these assumptions are not valid, this can lead to non-convergence or instability of the methods. We introduce kernel methods in data assimilation to model uncertainties in the data in a more flexible way than with Gaussian assumptions. The aim is to extend the assimilation methods to problems where they are currently inefficient. The Ensemble Transform Kalman Filter (ETKF) formulation of the assimilation problem is reformulated using kernels. Numerical results on the Lorenz 63 model highlight the benefit of the approach in terms of the decrease in root mean square error, especially when the size of the ensemble is small.

Ehouarn Simon,
Université de Toulouse, INP, IRIT – France

Lp-norm regularization – with $1<p<2$ – in variational data assimilation

The L1 and L2 norms have been successful as regularization terms in data assimilation (Freitag et al., 2010). The first one promotes a sparse solution while the second one promotes a smoother solution. The solution may however possess a structure ”in between” that we call ”quasi-sparse”. The Lp-norm with $1 < p < 2$ aims at making a compromise between these 2 norms. Moreover, the L2 and L1 norms introduce oscillations in the solution, and it has been shown that considering the Lp-norm with $1 < p < 2$ can mitigate these oscillations (Schuster et al., 2012).

Finally, the use of the Lp-norm is also motivated by the statistical distribution of physical variables, when it follows a generalized Gaussian distribution instead of the Laplace distribution or the more classical Normal distribution.

Michael Sitwell,
Environment and Climate Change – Canada

An Ensemble-Variational Inversion System for the Estimation of Ammonia Emissions using CrIS Satellite Ammonia Retrievals

Ammonia emissions into the atmosphere are relatively uncertain. An ensemble-variational inversion system was developed to estimate ammonia emissions using ammonia retrievals from the Cross-track Infrared Sounder (CrIS) for use in the GEM-MACH air quality model. Inversions for the monthly mean ammonia emissions over North America increased emissions by 11-41% within the domain. The revised emissions reduced the bias between GEM-MACH and surface observations by as much as 25%.

Sergey Skachko,
Environment and Climate Change Canada – Canada

A new daily SST analysis system at ECCC

A new global daily Sea Surface Temperature (SST) analysis system has been developed at Environment and Climate Change Canada (ECCC) to replace the existing operational system. All components of the new SST analysis system are implemented within the Modular and Integrated Data Assimilation System (MIDAS) software. MIDAS is already used for the data assimilation component of the main operational Numerical Weather Prediction (NWP) systems. The new SST analysis system, integrated together with the global sea-ice analysis, will be part of the combined surface analysis used for all operational prediction systems at ECCC. The data assimilation method used to compute the new SST analyses is 3D-Var with a diffusion operator for representing the horizontal correlations. New algorithms for the data quality control and satellite data bias estimation have also been developed for the new system. The performance of the new system is examined relative to the current operational system by using independent data and the impact of using the new SST analyses within other prediction system is evaluated.

Simone Spada,
National Institute of Oceanography and Applied Geophysics - OGS - Italy

With A. Teruzzi, G. Cossarini

A Gauss-Hermite high-order sampling hybrid filter for data assimilation in geoscience

We propose a novel ensemble algorithm, the Gauss-Hermite High-Order Sampling Hybrid (GHOSH) filter, which we apply in a twin experiment (based on Lorenz96) and in a realistic geophysical application. In the most relevant directions, the GHOSH filter can achieve a higher order of approximation than in other ensemble based Kalman filters (whose approximation order is usually equal to or lower than two). To evaluate the benefits of the higher approximation order, a twin experiment composed by thousands of Lorenz96 simulations has been carried out using the GHOSH filter and a second order ensemble Kalman filter (SEIK; singular evolutive interpolated Kalman filter). In the twin experiment, the comparison between the GHOSH and the SEIK filter has been done varying a number of data assimilation settings: ensemble size, inflation, assimilated observations, and initial conditions. The twin-experiment results show that GHOSH outperforms SEIK in most of the assimilation settings up to a 69% reduction of the RMSE on assimilated and non-assimilated variables. A number of assimilation experiments are performed by a GHOSH parallel implementation in a realistic geophysical application. In each experiment, a year of Mediterranean Sea biogeochemistry is simulated by assimilating satellite surface chlorophyll and varying the assimilation setup. The simulation results are validated using both semi-independent (satellite chlorophyll) and independent (nutrient concentrations from an in situ climatology) observations. Results show the feasibility of GHOSH implementation in a realistic three-dimensional application. The GHOSH assimilation algorithm improves the agreement between forecasts and observations without producing unrealistic effects on the non-assimilated variables. Furthermore, the sensitivity analysis on GHOSH setup indicates that the use of a higher order of convergence sensibly improves the performance of the assimilation with respect to nitrate (i.e., one of the independent observations). In view of potential operational applications of the GHOSH filter, it has to be noticed that GHOSH and SEIK filters have not showed significant differences in terms of time to solution, since, as in all the ensemble-like Kalman filter, the model integration is by far more computationally expensive than the assimilation scheme.

Martin Sprengel,
Deutscher Wetterdienst - Germany

Coupled Data Assimilation at DWD: Use of satellite observations for the ocean

The new project “Earth System Modelling at the Weather scale” (ESM-W) by DWD in cooperation with GeoInfoDienst BW aims to develop a coupled ocean-atmosphere forecasting system based on ICON-O for the ocean and ICON-NWP for the atmosphere. It uses a weakly-coupled assimilation approach combining the operational DA system of DWD (hybrid EnVar+LETKF) for the atmosphere with a newly developed 3DVar system for the ocean.

In the latter, different quantities derived from satellite measurements can be assimilated into the ocean model. We retrieve and monitor the OSTIA sea surface temperature (SST) and the SMOS sea surface salinity (SSS) daily product. These data sets are processed for assimilation and verification. Using our new 3DVar framework, coupled DA experiments assimilating satellite-derived SST and SSS into the ocean model are presented and verification results are discussed.

Thorsten Steinert,
Deutscher Wetterdienst – Germany

Data assimilation for a combined ICON/ICON-ART NWP system at DWD

The global operational NWP system of Deutscher Wetterdienst uses a high-resolution deterministic run and an ensemble at coarser resolution based on the ICON forecast model. We present a planned update that adds a second system using a coarser deterministic run and fewer members, but based on ICON-ART, an extension to ICON that allows the prognostic forecast of aerosol tracers and their interaction with the atmosphere. By combining both ensembles in the data assimilation system, a hybrid method EnVar consisting of a three-dimensional variational analysis (3D-VAR) coupled with an ensemble Kalman filter (LETKF), we can compensate for the reduced ensemble size and achieve analysis and forecast qualities comparable to or better than a full ensemble, while at the same time improving the current operational system. We present results for a three month test period in 2022.

Olaf Stiller,
Deutscher Wetterdienst – Germany

Applying covariance based cross-validation diagnostics for improving the localization of non-local observations

Localization is central to ensemble data assimilation as it not only reduces spurious correlations but also increases the effective size of the ensemble space, which is vital given the small number of ensemble members permitted by today's numerical capabilities. In practice, localization is often performed in observation space requiring, for each assimilated observation, the assignment of an assimilation region which is usually characterized by coordinates of the assimilation location as well as (horizontal and vertical) length scales for the extent of this region. Finding good (optimal?) values for these parameters can be challenging particularly for remote sensing data or observations like surface pressure whose assimilation may affect large parts of the atmospheric column.

This work shows how guidance for the localization can be obtained from novel covariance based diagnostics which have recently been introduced for cross-validating the impact which different observation type or groups have on the analysis or forecast. The proposed method is based on the fact that the success of most modern DA algorithms strongly depends on the consistency between the background error covariance matrix B employed directly or indirectly by the DA system, and the actual covariances between the assimilated observations and model space. The new diagnostics yield a direct test for this consistency with respect to selected observations and the part of model space corresponding to other (statistically independent) observations.

Working with the DWD LETKF it is shown how this can help i) to identify regions where a large assimilation impact would be expected given the employed covariances (derived from the DA ensemble) and ii) to test to which extent these theoretical expectations coincide with results from the observation based covariances.

Andrea Storto,
CNR ISMAR – Italy

Towards coupled air-sea data assimilation in a regional model

Coupled data assimilation in regional climate models is still largely unexplored, besides few pioneering applications, but represents a high potential for regional predictability gain, linked, among several factors, to the correction of imbalances at initial time and/or at the lateral boundaries, and the maximization of the benefits of the regional observing networks.

Here, we present initial works and results relative to extending a variational data assimilation system in the ocean to include atmospheric parameters and observations.

First, the concept is illustrated in an idealized atmosphere-ocean single-column model augmented with data assimilation, which indicates the benefits and advantages of the strongly coupled assimilation paradigm in a controlled environment. Sensitivity to the choice of the coupled background-error covariances is, however, large. Second, we show preliminary results in a realistic regional climate model over the Mediterranean region including state-of-the-art numerical models (NEMO, WRF, HD). Weakly coupled assimilation experiments indicate the importance of ocean data assimilation in the prediction of the intensity of individual medicane events. A roadmap and initial results from the strongly coupled scheme are also presented.

Yi-Jui Su,
Central Weather Bureau – Taiwan

Evaluation of an Ensemble Partial Cycle Framework for Use in the Regional Ensemble Prediction System at the Central Weather Bureau of Taiwan

The Central Weather Bureau (CWB) of Taiwan is responsible for issuing official weather forecasts and warnings of hazardous weather. For this purpose, CWB has been developing and operating global and regional numerical weather prediction (NWP) systems and providing both deterministic and ensemble forecast products as the forecast guidance. With the goal to move toward a more strongly linked global and regional NWP system, efforts have been made to connect the two systems. This study proposes a new regional ensemble initialization method named Ensemble Partial Cycle (EPC) to generate regional ensemble initial conditions for CWB's WRF-based regional Ensemble Prediction System (WEPS). The EPC method can be conceptually regarded as a combination of the Ensemble of Data Assimilation and partial cycle approaches. Our early results show that this new method is superior to the method that is adopted in the current operational WEPS based on an Ensemble Adjustment Kalman Filter data assimilation. The superiority is in terms of a better ensemble spread-error relationship and improved forecast performances. With the encouraging results from this newly proposed EPC method and the relevant development in CWB's global prediction system, it becomes possible to establish a better connected global and regional NWP system, which is more suitable for long-term NWP research and development in CWB. In this presentation, results from WEPS with EPC initialization will be further evaluated with several case studies to gain a more thorough understanding of the benefits of the proposed method.

Norihiko Sugimoto,
Keio University – Japan

With Y. Fujisawa, N. Komori, H. Ando, M. Takagi, H. Kashimura, AFES–Venus & ALEDAS–V teams

AFES (GCM) LETKF Data Assimilation System for Venus

We have developed Venusian GCM (general circulation model) named AFES–Venus (Atmospheric GCM for the Earth Simulator for Venus). Furthermore, in order to make use of observations by the Venus Climate Orbiter "Akatsuki", we have also developed the data assimilation system based on the LETKF (Local Ensemble Transform Kalman Filter) named ALEDAS–V (AFES–LETKF data assimilation system for Venus) for the first time in the world. Here we will introduce important results of AFES–Venus and object analysis produced by ALEDAS–V with Akatsuki horizontal winds assimilation.

Clemente Augusto Souza Tanajura,
Federal University of Bahia (UFBA) – Brazil

With F. B. Costa, L. B. Pires

Observing System Simulation Experiments with SWOT into HYCOM+RODAS over the Southwest Atlantic

Two observing system simulation experiments were realized with HYCOM and the Oceanographic Modeling and Observation Network (REMO) Ocean Data Assimilation System (RODAS) to investigate the impact of SWOT altimetric data in the Southwest Atlantic (34oS–12oS, 32oW–54oW) circulation and thermohaline structure. RODAS is based on EnOI, but the model error covariance matrix is not static and captures subseasonal variability. HYCOM was configured with horizontal resolution of 1/24o and 32 vertical layers nested in a HYCOM simulation with 1/12o in a larger domain. It was forced with NCEP/NOAA CFSR atmospheric forcing and the period of integration was 2 years (2011–2012). A ROMS free run with horizontal resolution 1/24o and 32 vertical levels offered SST, "Argo" T/S profiles and SLA (along–track and SWOT) synthetic data. The latter was processed with the Gaultier et al. (2021) SWOT data generator. In addition to a HYCOM free run (FREE), all observations were assimilated each 3 days (SWOT) and in the other SWOT data was denied (NO SWOT). The mean SSH correlation of FREE was 0.35, while the SWOT and NO SWOT were 0.67 and 0.50, respectively. The mean SST RMSD of FREE was 1.0oC, while the SWOT and NO SWOT were 0.54oC. Another OSSE was recently finished considering 2 SWOT satellites and the results will be assessed and presented.

Qi Tang,
University of Neuchâtel – Switzerland

An ensemble based data assimilation framework for an integrated hydrological model: development and examples

We developed an ensemble based data assimilation (DA) system for an integrated hydrological model to facilitate real-time operational simulations of water quantity and quality. The integrated surface and subsurface hydrologic model HydroGeoSphere (HGS) (Brunner & Simmons, 2012) which simulates surface water and variably saturated groundwater flow as well as solute transport, was coupled with the Parallel Data Assimilation Framework (PDAF) (Nerger et al., 2005). The developed DA system allows joint assimilation of multiple types of observations such as piezometric heads, streamflow, and tracer concentrations. By explicitly considering tracer and streamflow data we substantially expand the hydrologic information which can be used to constrain the simulations. Both the model states and the parameters can be separately or jointly updated by the assimilation algorithm.

A synthetic alluvial plain model set up by Delottier et al., (2022) was used as an example to test the performance of our DA system. For flow simulations, piezometric head observations were assimilated, while for transport simulations, noble gas concentrations (222Rn, 37Ar, and 4He) were assimilated. Both model states (e.g., hydraulic head or noble gas concentrations) and parameters (e.g. hydraulic conductivities and porosity) are jointly updated by the DA. Results were evaluated by comparing the estimated model variables with independent observation data between the assimilation runs and the free run where no data assimilation was conducted. In a further evaluation step, a real-world, field scale model featuring realistic forcing functions and material properties was set up for a site in Switzerland and carried out for numerical simulations with the developed DA system. The synthetic and real-world examples demonstrate the significant potential in combing state of the art numerical models, data assimilation and novel tracer observations such as noble gases or Radon.

James Taylor,
RIKEN Center for Computational Science – Japan

With A. Amemiya, S. Otsuka, T. Honda, Y. Maejima, T. Miyoshi

Improving short range prediction of convective weather systems using a 1000-member ensemble Kalman filter with 30-second update

Short-range prediction of convective precipitating systems remains one of the most challenging tasks in numerical weather prediction (NWP) owing to their highly non-linear evolution, multi-scale processes and fine-scale structures. Most current operational convective-scale NWP systems use update frequencies of 1 hour or greater using observations from conventional parabolic-type radar systems, that typically provide observations on approximately 15-20 vertical levels with scans every 5 minutes, which is insufficient to capture the detailed vertical structures or rapid changes that characterize convective scale weather systems. Other methods of short-range forecasting include nowcasting, which performs a temporal extrapolation of the precipitating system using motion vectors derived from sequential radar images. Nowcasting systems are commonly used by operational weather forecasting centers in conjunction with NWP models as they can provide skillful forecasts at short lead times. In this study we present the results of forecasts for a convective squall line performed using a new experimental real-time convective-scale NWP system that uses a 30-second update with observations from a Phased Array Weather Radar, an advanced type of radar that observes the full sky every 30-seconds with observations on approximately 100 vertical levels. The convective event took place in July 2021 during a month-long demonstration of real-time forecasting system that coincided with the Tokyo Olympic and Paralympic Games. We demonstrate the advantage of the rapid-update NWP system over a nowcasting system that uses PAWR data to generate forecasts at 30-second intervals, showing how the NWP system can successfully predict rapid changes to the convective system's structure and intensity, including the evolution of individual convective clouds, and outperform the nowcast system in forecasts at longer lead times. Additionally, we show how the 30-second update improves conditions for convective growth within the storm environment to enable the model to simulate the development of isolated convective clouds.

Wen-Hsin Teng,
Central Weather Bureau – Taiwan

Evaluation of an Ensemble Partial Cycle Framework for Use in the Regional Ensemble Prediction System at the Central Weather Bureau of Taiwan

The Central Weather Bureau (CWB) of Taiwan is responsible for issuing official weather forecasts and warnings of hazardous weather. For this purpose, CWB has been developing and operating global and regional numerical weather prediction (NWP) systems and providing both deterministic and ensemble forecast products as the forecast guidance. With the goal to move toward a more strongly linked global and regional NWP system, efforts have been made to connect the two systems. This study proposes a new regional ensemble initialization method named Ensemble Partial Cycle (EPC) to generate regional ensemble initial conditions for CWB's WRF-based regional Ensemble Prediction System (WEPS). The EPC method can be conceptually regarded as a combination of the Ensemble of Data Assimilation and partial cycle approaches. Our early results show that this new method is superior to the method that is adopted in the current operational WEPS based on an Ensemble Adjustment Kalman Filter data assimilation. The superiority is in terms of a better ensemble spread-error relationship and improved forecast performances. With the encouraging results from this newly proposed EPC method and the relevant development in CWB's global prediction system, it becomes possible to establish a better connected global and regional NWP system, which is more suitable for long-term NWP research and development in CWB. In this presentation, results from WEPS with EPC initialization will be further evaluated with several case studies to gain a more thorough understanding of the benefits of the proposed method.

Ricardo Todling,
NASA – USA

Is there a relationship between cornered-hat methods and a residual approach to estimate system uncertainty?

Recently a relationship has been established between a now traditional residual diagnostic used to estimate observation, background and analysis error covariances of interest to data assimilation and the three-cornered hat (3CH) method. An existing extension of the traditional residual diagnostic uses residuals from a fixed lag-1 Kalman smoother to retrieve system (model) error covariance. It is thus natural to ask if an additional relationship can be establish between this extended residual method and some form of cornered--hat method. The answer might seem straightforward. Unlike in the standard residual estimation case, attempting to estimate model error amounts to estimating the statistics of a residual quantity itself. As shown in this presentation, in such cases the 3CH method becomes trivial: the sought out uncertainty can be derived directly from the covariance of the residual quantity at hand. In the particular case of estimating model error, one of the corners would have to be composed of vectors providing estimates of model error; these are not typically available in practice. Therefore, at first glance the present work finds no relationship between the lag-1 smoother residual diagnostic for system error estimation and cornered--hat methods. This finds of this work will also be put in the light of recent work by Vogel & Menard (2022) which provides an expanded understanding of three-cornered methods.

Ricardo Todling,
NASA – USA

With D. Holdaway, B. Menetrier, R. Treadon, C. Thomas, D. Kleist and A. Shlyaeva

Preliminary Results Cycling GEOS-JEDI with GSI-based Background Error Covariances

The first phase of transitioning the NASA GMAO GEOS data assimilation capabilities to JEDI involves the replacement of the Gridpoint Statistical Interpolation (GSI) with a corresponding JEDI analysis. This includes taking JEDI's Unifed Observation Operator (UFO), its underlying dependencies, and the JEDI solver that enables a hybrid 4DEnVar strategy similar to what is used in the current GEOS atmospheric data assimilation system.

Variational analysis involves at least two main components associated with the observation and background cost function terms. The first is directly related to the UFO, which is being carefully validated in a joing collaboration between GMAO and NCEP to demonstrate consistency with correspondig results from GSI.

In a hybrid DA system, the second component, the background cost term, involves the ability to set up both a climatologically-based formulation of the background errors and an ensemble-based formulation. Although JEDI provides the features to allow for tuning scales associated with both these formulations, the exercise is laborious and non-trivial. A more direct route has been taken which interfaces the background error covariance machinery of GSI with JEDI. Furthermore, the background error covariance can benefit from a mechanism to provide additional dynamical balance to both its climatological and ensemble components.

In GSI this is accomplished with the Tangent Linear Normal Mode Constraint option. Similarly to the approach taken with background error covariance, JEDI is being interfaced with GSI's TLNMC to provide the required improved balance mechanism.

This presentation will go over the flavors of background error covariance formulation available now in JEDI; it will also show results from the TLNMC interface. This presentation will also show preliminary results from cycling GEOS-JEDI with Hybrid-4DEnVar and comparisons with its GSI-based counterpart.

Maria Toporov,
Institute of Geophysics and Meteorology, University of Cologne – Germany

With Ulrich Löhnert, Vera Schemann, Annika Schomburg, Jasmin Vural, Thomas Deppisch

Assimilation of ground-based microwave radiometer observations into the convection resolving ICON model: observing system simulation experiments

State-of-the-art high resolution, convection resolving NWP models require dense and frequent observations to define the detailed initial conditions. Key variables needed for convection-resolving data assimilation are, among others, the 3-dimensional fields of temperature and humidity. In the boundary layer, both variables are not adequately (vertically, horizontally and temporally) measured by current observing systems. A network of ground-based microwave radiometers (MWR) has the potential to provide real time profile observations. In our study, we perform an Observing System Simulation Experiment and evaluate the benefit and the potential impact of MWR observations on the accuracy of the initial thermodynamic state of the atmosphere. The MWR observations, sensitive to cloud water content, temperature and humidity profiles, are simulated from the Nature Run (simulated "truth") with the radiative transfer model RTTOV-gb and assimilated into the convection resolving ICON-D2 model (2km resolution). In this contribution, we present impact studies of assimilating synthetic observations of a single MWR instrument and extend the approach for evaluating the effect of a network of instruments on the analysis and forecast of thermodynamic fields. We examine the impact of MWR observations in addition to conventional observations by comparison of assimilation experiments to the Nature- and the Control Run (no assimilation).

Victor Trappler,
Inria – France

State-dependent preconditioning for Variational DA

Data Assimilation consists in improving the knowledge of the state of a dynamical system by combining information from a numerical model, from prior modelling, and from available observations. It is especially used to model, understand and forecast complex earth systems.

In Variational Data Assimilation, this is usually done by defining a succession of large and possibly ill-conditioned linear systems, constructed using linearizations of the forward model (the outer loop), and to solve them iteratively (the inner loop). In order to improve the convergence rate in the inner loop and reduce the computational burden, preconditioning techniques are often used to get better-conditioned linear systems, but constructing such preconditioners require additional computations.

The linear system that needs to be inverted depends on the current state of linearization, so we propose to construct a mapping from this state to an approximate inverse of the considered matrix using Deep Neural Networks. Once properly trained, this can be used as a state-dependent preconditioner in a variational data assimilation system, which can be constructed without any additional call to the forward model or linear tangent model.

Michael Tsyrulnikov,
Hydrometcentre – Russia

A locally stationary convolutional ensemble filter

Applications of the ensemble Kalman filter to high-dimensional problems in geosciences are feasible only with a small ensemble size. This makes the sample covariance matrix a poor estimate of the true prior (background-error) covariance matrix and necessitates a kind of regularization of the analysis problem. Existing approaches like covariance localization are largely ad hoc.

In this research we propose a regularization technique that involves a new non-stationary spatial model on the sphere. The model termed the Locally Stationary Convolution Model is a constrained version of the general Gausssian process convolution model. In the new ensemble filter proposed and tested in this study, prior covariances at each analysis step are postulated to obey the Locally Stationary Convolution Model. The spatially variable convolution kernel of the model is estimated online using the prior ensemble. Local stationarity means that the convolution kernel slowly changes in space as compared to its length scale. The estimator of the convolution kernel involves a series of spatial bandpass filters and a neural network. Strategies of training the neural network in the situation in which the truth (true prior covariances) is not available are discussed.

The new filter is tested in numerical experiments in static analysesl on the sphere and in cyclic filtering on the circular domain. The new technique outperformed three alternative methods, which involve specification of prior covariances as (i) localized sample covariances (stochastic ensemble Kalman filter approach), (ii) static (time mean) covariances (variational data assimilation), and (iii) a linear combination of localized sample covariances and static covariances (hybrid ensemble-variational assimilation). The advantage of the new filter turned out to be the largest with small ensembles. The Locally Stationary Convolutional Ensemble Filter does not require ad-hoc remedies like localization, variance inflation, relaxation to prior spread, etc.

Aspects of applications of the proposed approach to practical high-dimensional problems are discussed.

Arianna Valmassoi,
Deutscher Wetterdienst – Germany

With Jan D. Keller, Roland Potthast

Towards a new generation of regional reanalyses for Europe

Regional reanalysis data sets are becoming more and more popular for a broad spectrum of users, from climate adaptation and mitigation, to agricultural and economical applications.

While the development of a reanalysis system mainly relies on an existing numerical weather prediction (NWP) model and data assimilation scheme, it involves a large amount of testing from both a computational and technical perspective, including different spatial resolutions, ensemble sizes and choice of observation types used.

The work here aims to present the results of the development for a new pan-European regional reanalysis at the Deutscher Wetterdienst. The framework involves the ICON NWP model and its operational data assimilation method used for the European domain, namely EnVar for the deterministic run and Localized Ensemble Kalman Filter for the ensemble.

In preparing our next-generation reanalyses, we test varies settings of our system for the year 2022. As reference, we employ the 2023 operational forecast setup, which comprises a 40 member ensemble at 26 km resolution and a deterministic at 13 km run with a refining nest over Europe at 6.5 km, and assimilates all conventional observations, satellite products (e.g. radiation, atmospheric motion vectors) and radio-occultation measurements.

We create 5 experiments using the latin hyper-cube sampling changing the number of ensemble members (i.e. 10, 20, 40) and their resolution (i.e. 26 km, 40 km and 53 km), the deterministic resolution (i.e. 13 and 26 km), and the types of assimilated observations (i.e. removing the satellites, using only conventional observations and using everything as the reference).

Results show that in the European domain, the root mean squared error of the first 8 months' assimilated synop observation does not increase on average above 5\% with respect to the reference run for the 10-meter wind components and 2-meter temperature and relative humidity, but it is up to 10\% for the geopotential height.

Peter Van Leeuwen,
Colorado State University – USA

Continuous nonlinear data assimilation

One of the holy grails of data assimilation is a posterior probability density that does not have large jumps at observation times (filters) or at the start of an assimilation window (smoothers). A solution is a smoother with an ever-extending window. Because observations in the far future will not influence the present state in any significant manner we can use fixed-lag smoothers that update the model trajectory over a finite time window. While every time window will have large overlap with its previous window helping convergence, the problem is still challenging.

By starting the window with the state that is insensitive to the new observation (because of the window length), that state is not updated, and only the realizations of the errors in the model equations are updated. Given that 1) methods to estimate those errors are maturing, 2) the difficulty in generating the background error covariance, and 3) the huge impact this will have on model improvement, this might be a very fruitful way forward. The theory and applications using nonlinear Continuous Particle Flow Smoother will be discussed, including feasibility for geophysical systems.

Peter Van Leeuwen,
Colorado State University – USA

With A. Pein

Estimating Model Error Covariances from weak-constraint variational data assimilation

Weak-constraint variational data assimilation is expected to become mainstream because 1) it incorporates model errors, which are first-order present, 2) it allows for longer time windows and reduce dependence on artificial initial condition assumptions, and 3) it allows for systematic model improvement, one of the holy grails of data assimilation.

For this to happen robust ways to estimate these errors are essential. We have developed a new iterative methodology for model error estimation that can be built directly into any weak-constraint 4Dvar scheme. It has similarities with the Desroziers method, but is not restricted to observation space. We will outline the method, which is surprisingly simple and does not need evaluation of huge matrices, and test the methodology on several simplified geophysical models. It provides robust model error estimates that converge to true values even in sparsely observed settings.

Senne Van Loon,
Colorado State University – USA

Nongaussian Ensemble Data Assimilation

Most standard data assimilation methods are based on the assumption that the background and observational errors are drawn from a gaussian distribution. However, this is rarely the case. To improve the forecasting skill, one can replace the gaussian assumption by allowing some errors to follow lognormal and reverse lognormal distributions. This idea has already been successfully applied to variational data assimilation, and allowed for the development of a (reverse) lognormal Kalman filter. We extend these concepts to ensemble methods, based on the maximum likelihood ensemble filter.

We compare the ensemble data assimilation schemes with mixed gaussian, lognormal, and reverse lognormal distributions to the standard gaussian assumption. Moreover, we allow the underlying error distribution at each assimilation time to change dynamically, and present a machine learning technique that can decide on the optimal distribution to use within the Lorenz-63 model.

Erik S. Van Vleck,
University of Kansas – USA

Adaptive Meshing for Ensemble Based Data Assimilation

The use of adaptive, non-uniform meshes has proven invaluable in the accurate, efficient approximation of solutions to time dependent, nonlinear partial differential equations (PDEs). For ensemble based data assimilation (DA), in addition to the approximation of ensemble solutions, adaptive meshing can take advantage of several factors to enhance the quality of DA results. Primary among these factors is the treatment of the observational data. In this talk we survey recent work on adaptive meshing in DA and outline some of the challenges and opportunities. We present an approach to adaptive meshing in higher space dimensions based on a variational, quasi-Lagrangian meshing technique that makes use of monitor functions (MFs)/metric tensors (MTs)/mesh density functions (MDFs) to determine mesh concentrations. By designing MDFs specific to the observational data, the location and relative importance of the data can be incorporated into adaptive meshes. Some recent progress in this direction is presented that includes the development of robust meshes designed for a range of similar behaviors and the use of non-local convolution type observation operators. Numerical experiments are presented using a discontinuous Galerkin (DG) adaptive mesh technique with application to DA with two space dimension hyperbolic PDEs, an inviscid Burgers' equation and a configuration of a three component Shallow Water Equation.

Martin Verlaan,
Deltares & TU Delft – Netherlands

With T.Zijker, F.Zijl, J.Sumihar

A bias-kalman-filter for operational storm-surge forecasting

Operational forecasts for extreme high-waters at the coast during storms is important for early warning in low lying coastal and delta regions. Data assimilation is an important tool to increase the accuracy of these forecasts and has been operational in various forms since the 1990s. The main source of uncertainty in the Kalman filter for this application has always assumed to be errors in the wind-stress at the sea-surface. However, over time weather forecasts and drag-formulations have become much more accurate over the past decades and now are a significant fraction of forecast errors. The assumptions behind the operational Kalman filter now no long fit very well with the observed forecast errors, which leads to sub-optimal performance.

In this study, we extended an idea for a bias Kalman filter, proposed byAsher et al 2019. The method was based on an ensemble Kalman filter, with a system-error that was introduced as a term similar to the gravity potential, to make sure that the updates would be dynamically consistent and persistent. The measurements were reduced to daily averages, to suppress corrections at sub-daily time-scales. Analysis updates are performed once a day. The resulting filter turned out to be computationally efficient and very robust. The forecast errors are reduced for lead times up to 7 days ahead. Moreover, it turned out to be possible to apply the corrections of the coarse large scale model also to the fine scale model, which is used for the actual forecasts. This is important, because the fine scale model is too expensive to run in an ensemble mode. The resulting forecasts are now being tested and will go operational in September 2023.

Sophie Vliegen,
Alfred-Wegener-Insitut, Bremerhaven – Germany

With Y. Sun, L. Nerger

Assessment of weakly and strongly coupled data assimilation in ocean-biogeochemical modeling

Coupled ocean-biogeochemical models simulate the physical ocean dynamics as well as biogeochemical processes. The coupling is typically one-way so that the physical state influences the ecosystem. There are both satellite observations of the physical properties, e.g. of the sea surface temperature, and the biogeochemistry where observations of surface chlorophyll provide information on plankton concentrations.

We use the coupled ocean-biogeochemical model system NEMO-ERGOM to assess how coupled data assimilation influences the model state. The model is configured as a regional setup for the North Sea and Baltic Sea with a resolution of 1.8 km. The configuration is used in the Copernicus Marine Service CMEMS for the Baltic Sea for operational forecasts and to compute a reanalysis using data assimilation functionality provided by the Parallel Data Assimilation Framework (PDAF, https://pdaf.awi.de). Currently, the operational system uses on offline-coupled (file-based) assimilation setup. In the EU-project SEAMLESS we augmented the NEMO-ERGOM model with online-coupled assimilation for improved compute performance. We assimilate satellite sea surface temperature and chlorophyll observations using an ensemble Kalman filter and assess both weakly and strongly coupled data assimilation with a focus on the biogeochemical variables.

The developed codes for the data assimilation system are in wide parts generic and can also be applied with other model configurations or components. While the developments in SEAMLESS are independent from the BAL-MFC operational developments, it is planned to make them available to the operational service.

Femke Vossepoel,
Delft University of Technology – Netherlands

On Gaussianity and nonlinearity in state- and parameter estimation in earthquake cycle models

The use of data assimilation for earthquake forecasting is still in its infancy. Recent studies have focused on the estimation of stress, strength, and friction parameters of seismic cycle models using Ensemble Kalman filtering (EnKF). While the EnKF successfully reconstructs earthquake cycles based on limited observations of fault slip and subsurface stress, the non-Gaussian distribution of state variables complicates the estimation of the non-linear fault behaviour during seismic slip. This study compares the performance of the EnKF, the particle filter, the adaptive Gaussian mixture filter and the particle flow filter for state- and parameter estimation in earthquake sequences. We test the methods' performance using a simplified Burridge-Knopoff seismic cycle model in periodic and aperiodic conditions. We also evaluate how these methods affect the forecasting skills of the seismic cycle model.

Jasmin Vural,
Deutscher Wetterdienst – Germany

Challenges and benefits of assimilating Doppler-lidar and microwave-radiometer observations

Despite its importance for the numerical weather forecast, the atmospheric boundary layer (ABL) is still undersampled in terms of observations. This observational gap in the ABL is addressed by different types of novel ground-based remote-sensing instruments, which provide vertical profiles of different variables. We implemented observations of two instruments located at the Meteorological Observatory Lindenberg, namely microwave radiometer (MWR) and Doppler lidar (DL), into the ICON/KENDA assimilation system of the DWD. The DL wind observations can be assimilated directly and in the same fashion as the established radar wind profiler (RWP) observations and, most importantly, the RWP provides a direct reference at the same location. We will discuss different aspects of the optimisation process, such as observation error tuning and vertical thinning. The main challenges of the MWR observations are caused by the nature of passive remote sensing instruments as they only provide unlocalised measurements. Proper vertical localisation of both temperature and humidity information turned out to be a key factor for getting a positive impact on the numerical weather prediction (NWP). We show how the channel selection remains a major challenge due to the lack of a proper treatment of observation error covariances in our system. In addition, the lack of a sufficient number of reference observations to verify against makes the evaluation challenging. Nevertheless, we were able to obtain a positive impact on the NWP for both observation systems and will present our final assessment on the use of these instruments for data assimilation.

Shizhang Wang,
Nanjing Joint Institute for Atmospheric Sciences – China

A Local Data Assimilation Method (Local DA v1.0) and its Application in a Simulated Typhoon Case

Integrating the hybrid and multiscale analyses and the parallel computation is necessary for current data assimilation schemes. A local data assimilation method, Local DA, is designed to fulfill these needs. This algorithm follows the grid-independent framework of the local ensemble transform Kalman filter (LETKF) and is more flexible in hybrid analysis than the LETKF. Local DA employs an explicitly computed background error correlation matrix of model variables mapped to observed grid points/columns. This matrix allows Local DA to calculate static covariance with a preset correlation function. It also allows using the conjugate gradient (CG) method to solve the cost function and allows performing localization in model space, observation space, or both spaces (double-space localization). The Local DA performance is evaluated with a simulated multiscale observation network that includes sounding, wind profiler, precipitable water vapor, and radar observations. In the presence of a small-size time-lagged ensemble, Local DA can produce a small analysis error by combining multiscale hybrid covariance and double-space localization. The multiscale covariance is computed using error samples decomposed into several scales and independently assigning the localization radius for each scale. Multiscale covariance is conducive to error reduction, especially at a small scale. The results further indicate that applying the CG method for each local analysis does not result in a discontinuity issue. The wall clock time of Local DA implemented in parallel is halved as the number of cores doubles, indicating a reasonable parallel computational efficiency of Local DA.

Xuguang Wang,
University of Oklahoma – USA

Recent Development of Multiscale Data Assimilation for Numerical Weather Prediction

The emergence of new computing resources and technologies allows for a significant increase in the effective resolution of the next generation numerical models. Such computational resources also streamline the path for observations from a myriad of existing and new in-situ and remote sensing platforms, which observe a variety of scales, to be ingested in the numerical weather prediction (NWP) models. A next generation data assimilation (DA) system is therefore required to effectively analyze the state and its uncertainty across multiple scales, which hereafter is termed as "multiscale data assimilation (MDA)". In this talk, challenges associated with the MDA for NWP are first discussed. A series of recent developments on simultaneous multiscale DA are then discussed. For example, the simultaneous MDA concept is extended to the multi-resolution (MR) ensemble 4DEnVar context and implemented for US NOAA Global data assimilation system (GDAS). It is found that the multi-resolution ensemble 4DEnVar can improve both the global and hurricane track forecasts compared to an equal cost single highresolution 4DEnVar system. Second, a flow dependent vertical localization method is developed for GDAS 4DEnVar and is found to improve tropical cyclone track forecast. Third, the simultaneous multiscale EnVar is further developed with both scale dependent localization (SDL) and variable dependent localization (VDL) and implemented for the NOAA Continental US (CONUS) convection allowing DA system, RRFS and NOAA convection allowing hurricane DA system, HAFS. It is found that this newly developed simultaneous MDA method improves the analysis and prediction of supercells, MCSs and hurricanes. Last, a new simultaneous MDA approach based on the pure ensemble DA, the Multiscale Local Gain Form Ensemble Transform Kalman Filter (MLGETKF) including a scale dependent inflation (SDI) is introduced and tested with a two layer turbulence model.

Anthony Weaver,
CERFACS – France

Scale-dependent background-error covariance modelling, with application to global ocean data assimilation

This presentation describes advances in the use of ensembles in the background-error covariance matrix (B) for improving analyses at multiple spatial scales. The approach taken is inspired by earlier work on Scale-Dependent covariance Localization (SDL). In the current work, the basic formulation of the SDL matrix is used to define a Scale-Dependent covariance Model (SDM). In turn, SDM can be combined with SDL to form a scale-dependent hybrid covariance matrix. Different aspects of the methodology will be discussed including scale separation, scale-dependent covariance parameter estimation, normalization, and computational cost. Results from multi-year experiments with a global ¼ degree ocean model show clear benefits from the multi-scale B formulation compared to a standard "single-scale" B formulation.

Martin Weissmann,
University of Vienna – Austria

The impact of Aolus satellite wind lidar observation in the global NWP system of DWD

This study investigates the impact of wind profiles from the Aeolus satellite in the global NWP system of Deutscher Wetterdienst based on a 3-month data denial experiment. The Aeolus observation led to a reduction of forecast errors of zonal wind, temperature and geopotential up to five days lead time of 2 – 4% on global average. The reduction in forecast error was largest in the tropical upper troposphere and stratosphere as well as in the mid- and upper troposphere of the Southern Hemisphere. The Northern Hemisphere shows a somewhat smaller but still clearly beneficial impact of Aeolus observations.

Overall, Aeolus has a remarkable impact on the forecast skill of global NWP models and we present indication that this large impact of a single satellite instrument is linked to two key features of atmospheric dynamics that are largely unobserved by other systems: Vvertical wind shear in the tropics that is crucial for vertical wave propagation and upper-tropospheric divergence in the extratropics.

Jeffrey Whitaker,
NOAA Physical Sciences Laboratory – USA

Multi-scale localization in the LETKF

Multi-scale covariance localization has shown to improve the background-error covariance estimates in ensemble-variational systems, relative to using a single localization scale. It is now used operationally for atmospheric data assimilation at several operational centers. Recently, a similar approach has been implemented in a local gain formulation of the ensemble transform Kalman filter (MLGETKF) by Wang et al (https://doi.org/10.1175/MWR-D-20-0290.1). The local gain form allows for the use of covariance (B) localization through the use of pseudo-ensemble perturbations through a modulation procedure utilizing the square-root of the (multiscale) localization matrix. Here we test a new method for applying multiple localization scales within the "traditional" LETKF that uses observation-error (R) localization, which is computationally more efficient than the MLGETKF B-localization approach. This method is compared to the MLGETKF approach, and the R-localization approach proposed by Kondo and Miyoshi (https://doi.org/10.2151/sola.2013-038), in a simple non-cycled statistical model as well as a cycled system using a nonlinear surface quasi-geostrophic turbulence.

Nicholas Williams,
Nansen Environmental and Remote Sensing Center – Norway

With Y. Wang, F. Counillon

Enhancing sea ice prediction in NorCPM using assimilation of sea ice thickness from ENVISAT and C2SMOS

We will present the new development of the Norwegian Climate Prediction Model (NorCPM) to improve prediction of Arctic sea ice extent. NorCPM is developed at the Bjerknes Centre for Climate Research in Norway and a combination of the Norwegian Earth System Model (NorESM) and the Ensemble Kalman Filter (EnKF) for the purposes of reanalysis and prediction. The system we use features strongly coupled assimilation of sea surface temperature, temperature and salinity profiles, sea ice concentration and sea ice thickness. The sea ice concentration assimilation scheme has been recently revised to reduce the impact of the degradation of ocean heat content, with the influence of SST near ice covered regions reduced. The assimilation of sea ice thickness is implemented using observations from ESA CCI and C2SMOS and we compare full field and anomaly assimilation for thickness. We find that the assimilation can constrain the error in ice thickness substantially to under 1 m over the ESA CCI period and under 0.4 m over the C2SMOS period. We will also show how sea ice extent prediction is further improved in NorCPM by assimilation of sea ice thickness.

Richard Williams,
Deutscher Wetterdienst – Germany

Coupled Data Assimilation at DWD: Impact of conventional observations on the ocean

The new "Earth System Modelling at the Weather scale" (ESM-W) project at the DWD -- in cooperation with the GeoInfoDienst BW -- aims to develop a forecasting system that couples the atmosphere (ICON-NWP) with the ocean (ICON-O). Ocean observations are assimilated using a newly implemented 3DVar framework that is weakly-coupled to the DWD's operational setup for the atmosphere (hybrid EnVar+LETKF).

Here we focus on in-situ vertical temperature/salinity observations of the ocean, primarily obtained through ARGO profiling floats, and assimilated via 3DVar. We present results of several experiments within our coupled data-assimilation framework, with verification performed by comparison to the available satellite products of the ocean's surface.

Marek Wlasak,
Met Office – United Kingdom

With Benjamin Menetrier, Anna Shlyaeva,
Mayeul Destouches, Stefano Migliorini,
Yannick Tremolet

The development of a static spectral background error covariance model within JCSDA (Joint Centre for Satellite Data Assimilation) JEDI (Joint Effort for Data assimilation Integration) framework

The Met Office is developing a next generation state of the art data assimilation system within JEDI that can be run on both next-generation massively parallel software architectures (for operational usage) and more modest machines (e.g., for academic research).

A part of this effort has been to develop a static background error covariance model that is scientifically inspired by our current operational implementation while being more modular and able to cope with LFRic's (the new forecast model's) cubed sphere geometry. The modular design makes it both easy to change and to assess scientific choices that have been made. The code is open source and is available from JCSDA's SABER (System-Agnostic Background Error Representation) GitHub repository.

This presentation tells the story of how this covariance model was developed and displays the current capabilities of SABER.

Pin-Ying Wu,
Japan Meteorological Business Support Center – Japan

Investigation of error growth and non-Gaussianity in severe weather predictions using large-ensemble DA

We conducted 1000-member ensemble simulations of the typhoon and mesoscale convective system (MCS) that brought record-breaking rainfall in Japan using regional LETKF and EnVar systems. The forecast error growth and probability density distribution (PDF) are estimated and compared among typhoon and MCS events. The results showed that initial error growth is more pronouncedly in the MCSs, while the error from lateral boundary conditions limited the forecast uncertainty in the typhoon. The estimated PDF indicated non-Gaussian distribution related to strong winds brought by the typhoon. The relationship between the PDF shapes and the error growth features is also investigated.

Ting-Chi Wu,
Central Weather Bureau – Taiwan

Evaluation of an Ensemble Partial Cycle Framework for Use in the Regional Ensemble Prediction System at the Central Weather Bureau of Taiwan

The Central Weather Bureau (CWB) of Taiwan is responsible for issuing official weather forecasts and warnings of hazardous weather. For this purpose, CWB has been developing and operating global and regional numerical weather prediction (NWP) systems and providing both deterministic and ensemble forecast products as the forecast guidance. With the goal to move toward a more strongly linked global and regional NWP system, efforts have been made to connect the two systems. This study proposes a new regional ensemble initialization method named Ensemble Partial Cycle (EPC) to generate regional ensemble initial conditions for CWB's WRF-based regional Ensemble Prediction System (WEPS). The EPC method can be conceptually regarded as a combination of the Ensemble of Data Assimilation and partial cycle approaches. Our early results show that this new method is superior to the method that is adopted in the current operational WEPS based on an Ensemble Adjustment Kalman Filter data assimilation. The superiority is in terms of a better ensemble spread-error relationship and improved forecast performances. With the encouraging results from this newly proposed EPC method and the relevant development in CWB's global prediction system, it becomes possible to establish a better connected global and regional NWP system, which is more suitable for long-term NWP research and development in CWB. In this presentation, results from WEPS with EPC initialization will be further evaluated with several case studies to gain a more thorough understanding of the benefits of the proposed method.

Ting-Chi Wu,
Central Weather Bureau – Taiwan

Recent Development and Evaluation of a Global Atmospheric Ensemble Data Assimilation using NICAM global model and Maximum Likelihood Ensemble Filter with State Space Localization

The Maximum Likelihood Ensemble Filter (MLEF) with State Space Localization (MLEF-SSL) was recently developed as a new ensemble data assimilation method that incorporates state space covariance localization, global numerical optimization, and implied Bayesian inference. MLEF-SSL uses random projection to compute the localized forecast error covariance and reduce the analysis dimensions to a manageable space. MLEF-SSL is being applied to a realistic NWP system that includes the Nonhydrostatic Icosahedral Atmospheric Model (NICAM) and assimilation of conventional atmospheric observations and satellite radiances. This so-called NICAM-MLEF-SSL system will be used to explore the performance and effectiveness of MLEF-SSL in realistic and high-dimensional settings. In addition, we will use NICAM-MLEF-SSL to examine the effective ensemble size, which ranges from O(104) to O(105) for high-dimensional applications, by taking advantage of the supercomputing resources of Fugaku.

Earlier results from the NICAM-MLEF-SSL development suggest a satisfactory structure of the forecast error covariance. In this study, we describe further development of the system that includes improved algorithmic solutions to accommodate mathematical operations with high-dimensional matrices, nonlinear minimization, and estimation of posterior error covariance. The NICAM-MLEF-SSL will be evaluated via the assimilation of conventional observations from PREPBUFR data along with satellite microwave radiances from AMSU-A and MHS.

Youlong Xia,
NOAA/NCEP/EMC – USA

With Michael Barlage, Daryl T. Kleist, Clara Draper, Tseganeh Gichamo, Jiarui Dong, and Zhichang Guo

Assessment and application of the UFS Land DA System at NCEP/EMC

The Unified Forecast System (UFS) is a community-based, coupled, comprehensive Earth modeling and prediction system. It provides support for multiple applications of weather and climate forecasts and different forecast durations and spatial domains from regional to global scales. There have been several recent advances in the representation of land processes in the UFS, including the introduction of the Noah Multi-Physics (NOAH-MP) land surface model, and the development of a modern land data assimilation system for global NWP applications. Here, we present an in-depth evaluation of i) the snow simulated by the Noah-MP model, and ii) impact of the snow data assimilation scheme that has been developed for global NWP. The Noah-MP model has evolved through community efforts to pursue and refine a modern-era land model suitable for use in the NCEP operational weather and climate prediction models. It includes multiple-layer snowpack processing, vegetation dynamics, simple groundwater simulation, and updated soil and vegetation parameters. The new snow data assimilation scheme is performed with the JEDI (Joint Effort for Data assimilation Integration) software, and uses the LETKF-OI [Local Ensemble Transform Kalman Filter (LETKF) — and Optimal Interpolation (OI)] to assimilate station snow depth and remotely sensed snow cover. Due to the flexibility of the JEDI system, the LETKF-OI can be identically applied to Noah-MP forecasts generated either offline or within the coupled atmospheric model. Currently, this DA system is being implemented into the NCEP's next version of Global Forecast System (GFSv17).

As a primary step, it is important to assess the impact of assimilating snow depth and snow cover data on snowpack including snow depth, snow water equivalent, and snow cover fraction. An evaluation of the new snow DA applied to the current global NWP system, showed significant improvement of forecast snowpack and a small improvement to near surface temperature. Here, we present a comprehensive assessment of the effect of the new snow DA on surface albedos, water and energy fluxes, and state variables. To achieve this we apply the snow DA within an offline (land-only) modeling framework, using ERA5 meteorological forcing to drive the Noah-MP model, and the JEDI LETKF-OI to assimilate snow depth observed from either Global Historical Climatology Network or Global Telecommunication System. We use both dependent and independent observations to investigate impact of land snow DA on different spatial scales (e.g., single grid point, watershed, river basin, continental, global), different years and seasons, and different FV3 grid configurations (ie., C96, C384, C768). These observations include IMS snow cover fraction, MODIS percent of snow cover fraction and albedos, as well as the observations from the flux towers, SNODAS, SNOTEL, etc. Additionally, the impact of quality analysis and control filtering, GHCN vs GTS, and updated soil color data on land snow DA is also investigated. This presentation reports some preliminary results from these experiments and investigations.

Naicheng Xu,
University of Oklahoma – USA

With Xuguang Wang, Yongming Wang

Scale-Dependent Inflation for Multiscale Ensemble based Data Assimilation

Data assimilation faces great challenges associated with the future landscape of high resolution numerical models and a variety of existing and new in-situ and remote sensing observations. A next generation data assimilation system is required to effectively analyze the state and its uncertainty across multiple scales, which hereafter is termed as “multiscale data assimilation (MDA)” (Wang et al. 2021). A multiscale ensemble-based data assimilation method, MLGETKF (Multiscale Local Gain Form Ensemble Transform Kalman Filter), was introduced and shown to gain 12hour-1day predictability compared to a scale unaware data assimilation approach using a two-layer turbulence model (Wang et al. 2021). However, deficiencies in the background ensemble still exist. Wang et al. (2021) found that these deficiencies can be scale dependent. In this study, several methods are proposed to enable scale dependent inflation (SDI) to address such deficiency. Experiments with the two-layer turbulence model have shown that the proposed scale dependent inflation (SDI) can further improve the accuracy of the multiscale analysis and the subsequent forecast.

Yue (Michael) Ying,
NERSC – Norway

Improving vortex position accuracy with a new multiscale alignment ensemble filter

A multiscale alignment (MSA) ensemble filtering method was introduced by Ying (2019) to reduce nonlinear position errors effectively during data assimilation. The MSA method extends the traditional ensemble Kalman filter (EnKF) to update states from large to small scales sequentially, during which it leverages the displacement vectors derived from the large-scale analysis increments to reduce position errors at smaller scales through warping of the model grid. This study stress-tests the MSA method in various scenarios using an idealized vortex model. We show that the MSA improves filter performance as number of scales N_s increases in the presence of nonlinear position errors. We tuned localization parameters for the cross-scale EnKF updates to find the best performance when assimilating an observation network. To further reduce the scale mismatch between observations and states, a new option called MSA-O is introduced to decompose observations into scale components during assimilation. Cycling DA experiments show that the MSA-O consistently outperforms the traditional EnKF at equal computational cost. A more challenging scenario for the MSA is identified when the large-scale background flow and the small-scale vortex are incoherent in terms of their errors, making the displacement vectors not effective in reducing vortex position errors. Observation availability for the small scales also limit the use of large N_s for the MSA. Potential remedies for these issues are discussed.

Hongqin Zhang,
Institute of Atmospheric Physics,
Chinese Academy of Sciences – China

Integral correction of initial and model errors in system of multigrid NLS-4DVar data assimilation for numerical weather prediction (SNAP)

Model error has a marked influence on forecast error, and integral correction of the initial and model errors in a novel four-dimensional variational data assimilation (4DVar) method (i4DVar) has recently been developed. Unlike traditional 4DVar, in which the analysis increment focuses only on the initial time point, i4DVar introduces an averaged integral correction term to correct the error evolution at multiple times with the same time interval in the assimilation window; that is, the assimilation window is divided into several sub-windows. In this study, we incorporated i4DVar into the System of Multigrid Nonlinear Least-squares-4DVar Data Assimilation for Numerical Weather Prediction (NWP) (SNAP), hereafter referred to as SN-i4DVar, and explored its impacts on NWP. The effectiveness and performance of SN-i4DVar were evaluated using two extreme rainstorm cases assimilating conventional and satellite observations, respectively. The results indicated that integral correction of the initial and model errors reduced analysis error and forecast error growth rates. SN-i4DVar performed better in precipitation forecasting than the original SNAP and exhibited generally better forecasting capability in terms of the horizontal wind speed (u and v), temperature, and humidity variables, indicating that SN-i4DVar simultaneously countered initial model errors to improve NWP capability. Meanwhile, the 4DEnVar method, which considers the optimal analysis in the ensemble space, benefits from the improved ensemble space generated according to the formulas of i4DVar, and the observation information can be more effectively absorbed and a more accurate analysis can be obtained by SN-i4DVar.

Lijian Zhu,
Shanghai Typhoon Institute,
China Meteorological Administration – China

Direct Assimilation of All-Sky GOES-R ABI Radiances in GSI EnKF for the Analysis and Forecasting of a Mesoscale Convective System

In this study, all-sky GOES-R ABI infrared radiances at their native resolution are assimilated using an enhanced GSI ensemble Kalman filter (EnKF) data assimilation (DA) system, and the impacts of the data on the analysis and forecast of a mesoscale convective system (MCS) are explored. Results show that all-sky ABI BT data can correctly build up observed storms within the model and effectively remove spurious storms in model background through frequent DA cycles. Both bias and root-mean-squared innovation of the background and analysis are significantly reduced during the DA cycles, and free forecasts are improved when verified subjectively and objectively against observed ABI BTs and independent radar reflectivity observations. A horizontal localization radius of 30 km is found to produce the best results while 5-min DA cycles improve the storm analyses over 15-min cycles, but the differences in forecasts are small. Further analyses show that the clearing of spurious clouds by ABI radiance is correctly accompanied by reduction in moisture through background error cross covariance, but overdrying often occurs, which can cause spurious storm decay in the forecast. The problem is reduced when the ensemble mean of observation prior instead of observation prior of the ensemble mean state is used in the ensemble mean state update equation of EnKF. The significant difference between the two ways that the ensemble mean of observation prior is calculated when the observational operator is very nonlinear has not been recognized in earlier cloudy radiance DA studies. Furthermore, sensitivity experiments with different data assimilation (DA) configurations are conducted using observing system simulation experiments (OSSEs). The sensitivity of all-sky ABI DA on different data thinning intervals, covariance localization radii and inflation methods, and updated state variables are examined in this study. It is found that the combination of a 6 km data thinning interval for clear-sky BTs, a 30 km horizontal localization radius for both clear and cloudy BTs, RTPP covariance inflation method and updating all state variables is found to provide overall most improvement on the analysis and forecast of the targeted severe weather event, especially the 0–3 h forecasts. Comparison of covariance inflation methods indicates RTPS may introduce more dry bias than RTPP accompanied by the clearing of spurious model clouds, further degrading the forecast. These results can provide guidance for the optimal use of ABI data in the GSI EnKF system.

Milija Zupanski,
Cooperative Institute for Research in the Atmosphere – USA

Addressing Nonlinearity in Data Assimilation

Addressing nonlinearity in data assimilation is critical for making further advances in data assimilation. On one hand, there is an ever-higher model resolution, increased complexity of physical processes, development of coupled modeling systems, all generally contributing to an increased nonlinearity of the prediction models. On the observation side, there is a greater need to assimilate new observations relevant to high-resolution physical processes, such as convection, extreme precipitation, tropical cyclones, as well as land-surface and snow, ice sheets, aerosol, chemistry, many of them employing nonlinear observation operators. Development of non-gaussian data assimilation introduces additional nonlinearities to data assimilation.

We will discuss various approaches to addressing nonlinearity in data assimilation, mostly related to including numerical optimization, but also using retrievals based on machine learning, as well as the development of alternative nonlinear data assimilation approaches.