Joint ECMWF/OceanPredict workshop on advances in ocean data assimilation, May 2021.

Recommendations from working group 2

Mike Bell (co-chair), Elisabeth Remy (co-chair), Matt Martin (rapporteur)

Participants: M. Athanase, A. Aydogdu, A. Cipollone, E. de Boisseson, F. Counillon, F. Davidson, P. de Rosnay, S. Frolov, A. Geer, K. Mogensen, H. Pohlmann, M. Price, Z. Paschalldi, K. Vasily, A. Vidard, A. Weaver, J. While, H. Zuo.

Balancing model resolution versus ensembles

Global operational ocean and coupled forecasting and reanalysis systems in different centres are run at resolutions including 1/4°, 1/12°, 1/16° and 1/25°. Regional systems are run operationally down to kilometre-scale resolution. The balance of model resolution, ensemble size and the sophistication of the data assimilation method will depend on the application and scale of interest.

R1. The criteria for assessing the benefits of going to high resolution and of adding more ensemble members should be defined separately for short- to medium-range forecasting, for seasonal forecasting, and for reanalysis.

For short- to medium-range forecasting at low- and mid-latitudes, useful information about the benefits of resolution versus ensemble was provided by the talk of S. Frolov who showed that deterministic forecasts at 1/12° and 1/25° do not have very different forecast error characteristics, while a 20-member ensemble at 1/12° resolution gave useful forecasts for much longer lead times. At high-latitudes and in shelf-seas the scales of eddies are much smaller (down to ~1km). There are few observing systems resolving these scales in most regions, and the computational cost of resolving these features in global models would be prohibitive.

R2. For medium-range global forecasting applications, once the ocean resolution reaches 1/12° (or possibly even 1/4°), computational resources should probably be invested in producing ensemble forecasts or increasing the ensemble size and improving the data assimilation method.

R3. For global reanalysis applications (where systems are currently mostly run at 1/4° resolution), focus should be put on improving the data assimilation methods, including the use of time-varying ensemble information, and reducing the long-term biases. An assessment of the impact of increasing model resolution from 1/4° to 1/12° should be made using existing reanalysis datasets (e.g. CMEMS).

Infrastructure

The benefits of sharing infrastructure across different operational and research groups include: having a critical mass of people working on the code; better quality assurance of the code; more research on data assimilation methods; and better access to machine learning/artificial intelligence methods. It would also lead to improved links between operational and academic groups who could more easily use “real world” systems and facilitate the take-up of these for academic work, PhDs and training. Another major benefit is that the same framework could be used for multiple components of the Earth system, thereby enabling research into the application of more sophisticated coupled data assimilation methods.

The deterrents to sharing infrastructure include: a significant overhead in terms of coordination and communication; moving to a new infrastructure is a major undertaking for operational groups even if the resulting methods are what is required; an increase in the complexity of the system which could lead to reduced productivity. It is also the case that approximately 90% of an operational DA system can be implemented using general tools, but the last 10% often needs to be quite specific to a particular system. It was recognised that meeting the needs of academic researchers (flexible, multi-method, easy-to-use infrastructure) with more specific operational needs is difficult to achieve with the same tool.

While there was no recommendation for groups to take up shared infrastructure, for those groups considering doing so the following small steps should be addressed:

R4. There should be more coordination on observation operator code use and development, particularly for use with the NEMO model.

R5. There should be increased collaboration on stochastic methods between modelling groups and developers working on the generation of ensembles for DA systems.

R6. Coordination between JEDI and ECMWF OOPS should be maintained and improved where possible.

R7. For groups considering using JEDI, some initial simple "hands on" tasks should be done to gain familiarity with it, and experiences shared across those groups.

Best practices

It was recognised that there is a lack of shared, documented best practices within the ocean data assimilation community. The dissemination of information on best practices is being developed by the ocean observation community, see for example https://www.oceanbestpractices.org, and they are willing to include modelling and data assimilation best practices.

R8. Observing system experiments (OSEs) with coupled DA systems should be coordinated across multiple groups through relevant projects such as TPOS and SynObs. A data-rich period should be chosen for the OSEs and the adequacy of the (coupled) DA methods should be assessed. Existing best practice guidelines (https://www.aoml.noaa.gov/qosap/osse-checklist and https://www.oceanbestpractices.org) should be followed and updated if necessary for coupled systems.

R9. Best practice guidelines for the development of ocean DA systems for operational applications could be developed with a journal article published, linking with the Ocean Best Practices project (https://www.oceanbestpractices.org) and with the Expert Team on Operational Ocean Forecasting Systems (ETOOFS). This could include a description of methods to assess the fundamental behaviour of the DA systems (e.g. idealised observation experiments, average assimilation increments, increment retention), the definition of a common period and framework for experiments, and the definition of standardised data outputs accessible from the cloud to enable intercomparison.

A specific recommendation was made concerning the evaluation of reanalysis products:

R10. Reanalysis products should be assessed in terms of their fitness for purpose using a range of metrics. These should include the trends and accuracy of important climate phenomena; the consistency of the assimilation (mean assimilation increments, error covariance specifications); the behaviour of explicit estimates of uncertainty.

Data assimilation methods

The tendency is toward the inclusion of more and more earth system components. This requires more coordination between the different communities and means that the DA strategy should be explored further. It was also recognized that there is a lack of expertise/interest in the DA field from young scientists.

R11. A larger community should be built in coupled ocean/atmosphere/sea-ice/land DA with cross-disciplinary expertise including those with knowledge on boundary layer physics, satellite DA, ocean DA, sea-ice DA, land DA and machine learning. This could be done through an on-line seminar series, and through a formal course on coupled DA organised through a university. This would be a way to improve education and attract more students into the field of DA.

R12. The motivations for strongly coupled physical-biogeochemical ocean DA is clear. However, more research would be needed to consolidate the motivation for strongly coupled ocean-atmosphere DA (including the use of coupled forecast error covariances where increments are made to the ocean based on information from atmospheric observations and vice versa), particularly for short-range predictions.

Since the modelling systems now include more of the physics (due to coupled models, increases in resolution and the inclusion of more physical processes), the observation operators and assimilation of particular observation types could be enhanced by assimilating observations with less pre-processing applied.

R13. There are on-going efforts at ECMWF to make better use of surface-sensitive satellite radiance data within the coupled DA framework. This work should be extended to other observations of the ocean which require complex observation operators dependent on atmosphere and ocean information, for example sea level anomaly (SLA) data from altimeters, and satellite ocean colour data. Improvements should be made in the way atmospheric corrections are applied to SLA data by connecting the TAPAS (Tailored Product for Data Assimilation) and satellite radiance communities.

R14. With the data from upcoming satellite missions such as SWOT expected to have significant spatially correlated errors, progress should be made on efficient implementation of methods to model observation error correlations in DA systems. The representation of observation error correlations in vertical profiles, particularly those from high frequency measurements such as gliders, should also be developed.

Improvements to the DA methodology for reanalysis were also discussed and the following recommendation was made:

R15. Improved methods to deal with the issue of model bias in the context of an evolving measurement network, particularly for reanalysis, should be developed. Reanalyses should be viewed as a multi space- and time-scale problem where attention is paid to the different time-scales including long-term biases, slowly varying heat content changes, down to weather time-scales. Multilevel Monte Carlo approaches could be a useful technique to explore.

Machine learning has many potential applications in the DA process that are still to be explored in a more realistic context:

R16. There are large training data sets available and problems that are difficult to model, including: representation of model biases, multivariate balance relationships, preconditioning, estimation of normalisation factors for correlation modelling, reducing the sampling error of small ensembles. Progress should be made in applying machine learning techniques for these applications.