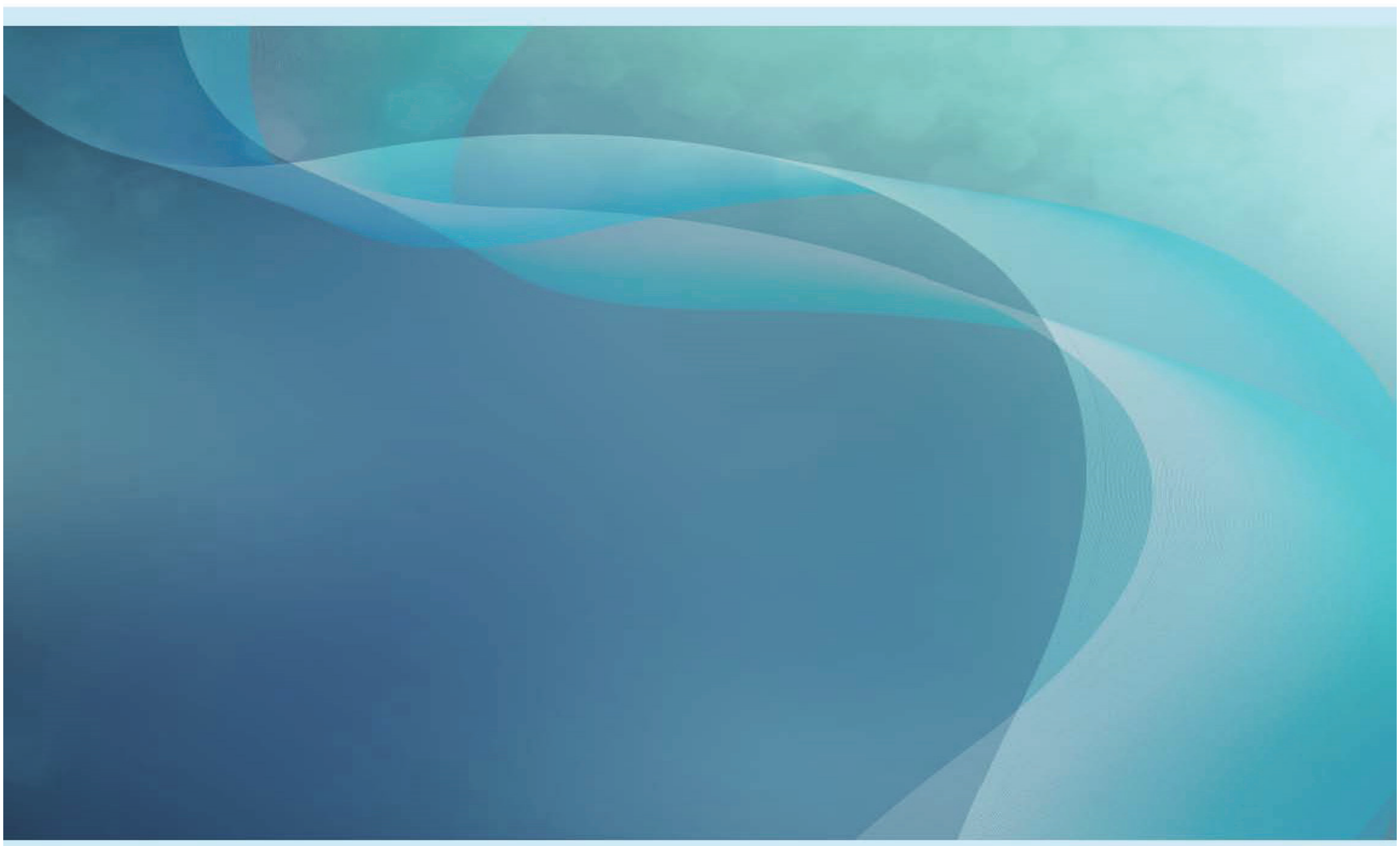
Australian Government
Bureau of Meteorology

Data assimilation – Abstracts of the tenth CAWCR Workshop 5-9 December 2016, Melbourne, Australia

Peter Steinle, Imtiaz Dharssi, Georg Gottwald, Val Jemmeson, Jeffrey Kepert, John Le Marshall, Jin Lee, Terence O'Kane, Pavel Sakov, Yonghong Yin and Keith Day (Eds.)

November 2016



Bureau Research Report: BRR017

Data assimilation – Abstracts of the tenth CAWCR Workshop 5-9 December 2016, Melbourne, Australia

Peter Steinle, Imtiaz Dharssi, Georg Gottwald, Val Jemmeson, Jeffrey Kepert, John Le Marshall, Jin Lee, Terence O'Kane, Pavel Sakov, Yonghong Yin and Keith Day (Eds.)

Bureau Research Report No. 017

November 2016

National Library of Australia Cataloguing-in-Publication entry

Author: Annual Research and Development Workshop; Data Assimilation (2016: Melbourne, Victoria)

Title: Data assimilation – Abstracts of the tenth CAWCR Workshop 5-9 December 2016, Melbourne, Australia / Editor Keith A Day

ISBN: 978-0-642-70683-6

Series: Bureau Research Report - BRR017

Enquiries should be addressed to:

Peter Steinle

Bureau of Meteorology
GPO Box 1289, Melbourne
Victoria 3001, Australia

p.steinle@bom.gov.au

Conference sponsors

We would like to acknowledge and thank the following sponsors for their participation in this conference

Copyright and Disclaimer

Presenters

FOREWORD

The physical modelling of environmental systems has been so successful over the past few decades that it has now become a critical element of modern society. The predictions and outlooks from these systems have a major impact on society, supporting activities such as emergency services, major primary industries, transport, long term risk evaluation and planning, down to every day personal decisions. This success is due to advances in observing technology, modelling science and computing power. Data assimilation/data fusion is central to bringing these three components together, - allowing the new observations to be used by the forecast models, and ensuring that the model variables are properly initialized, and for it too all be done in a timely and scientifically robust manner.

Despite the advances and successes over the past few decades, data assimilation is still far from a solved problem. New observing systems, more complex models incorporating fundamentally different physical processes due to higher resolutions and/or coupling with other models, major changes to computing architecture such as accelerators/GPUs, and of course increased demands in accuracy all combine to drive the demand for further advances in data assimilation techniques and implementations. Addressing these issues is recognized as a high priority to for both the Bureau and the wider, international community including the World Meteorological Organization's World Weather Research Program.

As a result data assimilation and data fusion are expected to continue as very active areas of both theoretical and applied research, and will continue to be strongly supported by many large research organizations.

This workshop brings together a large and diverse group of researchers from universities, operational centres and other research organizations to start outlining where the community is at and what needs to be done to provide data assimilation/data fusion systems suitable for Australia's needs over the next decade or so.

The workshop is organised around nine themes:

- Ensemble DA
- Atmospheric DA
- Satellite DA
- Oceanic DA
- Land DA
- Reanalysis / coupled DA
- Forecast sensitivity DA
- Advanced methods
- Generalised model data fusion

We are pleased to welcome the prominent scientists and experts from overseas, Australian research agencies and universities who have been invited to give presentations. Keynote speakers include:

- **Dr Jeffrey Anderson**
  University Corporation for Atmospheric Research (UCAR), USA
- **Dr Craig Bishop**
  Naval Research Laboratory (NRL), Monterey, USA
- **Dr Mark Buehner**
  Environment and Climate Change Canada
- **Mr Adam Clayton**
  UK Met Office, on secondment to Korea Meteorological Administration (KMA)
- **Dr John Derber**
  National Oceanic and Atmospheric Administration (NOAA), USA
- **Dr Clara Draper**
  National Aeronautics and Space Administration (NASA), Goddard, USA
- **Dr Tijana Janjic Pfander**
  Hans-Ertel-Centre for Weather Research, Deutscher Wetterdienst, Germany
- **Dr Sujay Kumar**
  Hydrological Sciences Laboratory, NASA, USA
- **Dr Stuart Moore**
  National Institute of Water & Atmospheric Research, New zealand
- **Dr Peter Steinle**
  Bureau of Meteorology
- **Dr Michael Tsyrulnikov**
  HydroMeteorological Centre of Russia (HydroMetCenter)

The workshop is hosted by the Bureau of Meteorology (BOM) and is sponsored by BOM, CSIRO, CRAY, and the National Computational Infrastructure. I would like to thank these sponsors for their generous support of the workshop. As chair of the workshop organising committee, I sincerely thank the members of the organising committee: Imtiaz Dharssi, Georg Gottwald, Val Jemmeson, Jeffrey Kepert, John Le Marshall, Jin Lee, Terence O'Kane, Pavel Sakov, Yonghong Yin and acknowledge the kind support provided by Anu Arora and Keith Day.

Dr Peter Steinle
Bureau of Meteorology

THE IMPORTANCE AND FUTURE OF DA AT THE BUREAU OF METEOROLOGY

Peter J. Steinle

Environment and Research Division, Bureau of Meteorology, Melbourne

peter.steinle@bom.gov.au

Introduction

The success of numerical modeling of the physical environment, and the important role played by data assimilation has been well documented around the world. Routinely used for weather, oceanic and seasonal prediction, forecasting of water resources and land surface conditions and for reanalysis of the Earth system, data assimilation has become an essential support for much of modern society. This applies as much to the Australian Bureau of Meteorology (BoM) as anywhere due to the BoM's role in providing information about the past, current and future state of the environment for time scales from minutes to decades. Despite this success, there are increasing demands for further advances in data assimilation. The most obvious reason is that there are still significant errors at times, driving the need for further improvement. Secondly, as society and technology advances, there are increased expectations and requirements for more sophisticated information. Both of these require research into more advanced data assimilation systems.

These issues are all reflected in one of the major international research strategies for environmental prediction: the World Meteorological Organization's World Weather Research Program (WWRP) Implementation Plan (WMO, 2016). While the focus of this is on weather and the atmosphere, the underlying issues extend to all parts of Earth system modeling.

The Implementation Plan aims to advance research to meet the major issues facing weather information providers over the next decade, and provides an internationally agreed approach for research into large areas of environmental modeling, and is therefore very relevant to many of the research issues facing the BoM and its partners. Furthermore, from a modeling perspective, a critical component is research and development in data assimilation – including the collection and use of new and non-conventional observations, advancing assimilation techniques, and the effective implementation of these techniques on current and upcoming high performance computers. The final key part is developing and encouraging new researchers within the field. This also provides the framework for this workshop – covering where the BoM and its research partners are currently placed, the issues we are facing. The involvement of research leaders at many large international institutes in the development of the WWRP implementation plan highlights that there is ample scope for very active and well-supported research careers spanning observing technology, mathematical modeling and high performance computing technology.

Observations

Although remotely sensed data from satellites has been the biggest data contributor to NWP systems for decades, there is still a vast amount of information that is not used. Work on improved forward models, the use of more channels from hyper-spectral infra-red sounders, and the use of low peaking channels over land have all shown promise – and some of this will be discussed in later presentations. While some of these advances have to an extent made their way into operational systems, there is still

considerable room for improvement. On top of this, the latest generation of geostationary satellites provides vastly more data (and information) than is currently used. A situation that will become even more embarrassing with the next generation without further developments. To improve this situation requires additional advances to enable the assimilation of cloud and land surface information. Advances in these areas are nonetheless essential for two reasons. Firstly the direct effects of improving the analysis and of cloud and land surface variables in particular are a clear priority. The second aspect is that improving these parameters should greatly assist in bridging the gap between nowcasting and NWP – one of the major concerns for supporting tactical decision-making during high impact weather such as fires, floods and heavy rain.

The other common source of remotely sensed data is from radars – again the direct use of radar data is generally rather limited, although it is increasing with recent advances in the use of rainfall assimilation, Doppler winds in rain and clear air and of reflectivity. Again some of this will be covered in later presentations. The common features between satellites and radar being that assimilation of data related to cloud, water or land surface variables requires advanced numerical models, background and observation error characterization, assimilation techniques and observation processing. Especially since the errors in both the background and the observations can be significantly non-Gaussian.

There are of course many other, new remote sensing technologies becoming available. Many of these were showcased during the Tokyo Metropolitan Area Convection Study (Nakatani et al. 2015) and the Surface Atmospheric Boundary Layer Exchange campaign (Wulfmeyer et al. 2015). These new instruments are redefining the information available for measuring the atmosphere, particularly in urban areas, however there are various assimilation issues which need to be reduced before this data can be better exploited.

When it comes to urban areas, it is also widely acknowledged that the established observing systems generally do not provide the detail required for urban scale modeling. This leads to consideration of using crowd-sourced data such as from mobile phones (Mass and Madaus, 2014) and various other sources. Characterizing the errors for these observations (including quality control) poses some interesting challenges.

Data Assimilation Techniques

The use of ensemble information to characterize background errors has become standard in oceanography and for global NWP with the extension to limited area NWP systems underway. There will be a number of presentations showing that with existing frameworks (generally variational, ensemble Kalman filter or hybrid) much can be gained by improving the specification of the background error covariances. There are however serious questions to be asked as to when do the assumptions about the error structures break to the extent that more advanced methods are required – particularly as we move to very high resolution systems (e.g. urban) and to greater coupling between the atmosphere-ocean-land-ice and potentially aerosols and chemistry.

High Performance Computing (HPC)

The third critical component of environmental modeling in the future is preparing for the next generation of HPC. The time required to develop state of the art systems means that to exploit the power of upcoming HPC most major centres have already started developing new software systems. However there are still significant challenges for numerical prediction models ahead as outlined in Kellie (2014) and Bauer (2016) and for data assimilation the problems are at least as challenging.

Summary

Advances in observations, data assimilation and HPC have provided enormous increases in the value society receives from environmental prediction models. With the continued demand for more detailed information capturing more complex features and interactions, more sophisticated observing and assimilation systems are called for. Given the importance of environmental information to modern society this combination of needs and challenges will continue to require a very active and long-term engagement across the research community. Many aspects of these challenges are covered in this workshop, which is expected to provide an additional stimulus to data assimilation research within Australia.

References

Bauer, P. 2016: Today's weather forecast: Good with a strong chance of improvement. http://www.earthmagazine.org/article/todays-weather-forecast-good-strong-chance-improvement

Kellie, A. 2014: Meeting the HPC Challenges of Atmospheric and Related Sciences, an NCAR View. WWRP World Weather Open Science Conference, Montreal, Canada. https://www.wmo.int/pages/prog/arep/wwrp/new/wwosc/documents/WWOSCFinalPresentationVersionWedAug20.pdf

Mass, C.F. and Madaus, L.E. 2014: Surface Pressure Observations from Smartphones: A potential Revolution for High-Resolution Weather Prediction?, Bull. Am. Met. Soc. http://dx.doi.org/10.1175/BAMS-D-13-00188.1

Nakatani, T., Misumi, R., Shoji, Y., Saito, K., Seko, H., Seino, N., Suzuki, S., Shussse, Y., Maesaka, T. and Sugawara, H. 2015: Tokyo Metropolitan Area Convection Study for Extreme Weather Resilient Cities, Bull. Am. Met. Soc. http://dx.doi.org/10.1175/BAMS-D-14-00209.1

WMO. 2016: Catalysing Innovation: WWRP Implementation Plan 2016-2023 https://www.wmo.int/pages/prog/arep/wwrp/new/documents/ WWRP_ImplementationPlan_Part1.pdf

Wulfmeyer, V. and the SABLE Team, 2015: New Concepts for Studying Land-Surface-Atmosphere Feedback Based on a new Lidar Synergy and Grey Zone Simulations. Geophys. Res Abst. EGU2015-5054
http:// meetingorganizer.copernicus.org/EGU2015/EGU2015-5054.pdf.
.

BUILDING STATE-OF-THE-ART FORECAST SYSTEMS WITH THE ENSEMBLE KALMAN FILTER

Jeffrey Anderson

Institute for Mathematics Applied to Geosciences
National Center for Atmospheric Research, USA

jla@ucar.edu

This talk provides a comprehensive introduction to an ensemble Kalman filter data assimilation system, the Data Assimilation Research Testbed (DART). DART can produce high-quality weather predictions but can also be used to build a comprehensive forecast system for many other climate system models and observations. A description of the basic ensemble Kalman filter algorithm is followed by a discussion of algorithmic enhancements, in particular localization of observation impacts and inflation of prior ensembles, that are essential for efficient implementations for large prediction models. Several example applications in geosciences will be used to examine additional capabilities of modern ensemble prediction systems. This talk will provide background for subsequent talks by other speakers that will explore the newest developments in ensemble filtering.

STOCHASTIC AND DOUBLY STOCHASTIC SPATIO-TEMPORAL FIELD MODELING FOR DATA ASSIMILATION

Michael Tsyrulnikov, Dmitry Gayfulin and Alexander Rakitko

HydroMetCentre of Russia

michael.tsyrulnikov@gmail.com

Introduction

First, a limited area Stochastic Pattern Generator (SPG) is presented and its properties discussed. The SPG is designed to be used as a building block in various model tendency error simulation techniques in ensemble data assimilation and ensemble prediction. The SPG relies on the "proportionality of scales" property (Tsyrulnikov 2001) of the spatio-temporal field to be generated. This property implies that larger (shorter) spatial scales are associated with larger (shorter) temporal scales.

Second, two doubly stochastic models of "truth" are introduced. One model is a single-variable time discrete model and the other one is an evolutionary model for spatio-temporal pseudo-random fields on the 1D spatial domain (the circle). The doubly stochastic models can produce complicated and highly variable covariances of the "truth" and, importantly, error covariances of any filter in question. The unique feature of the doubly stochastic models is that the above covariances can be estimated not only in the time mean sense (which is possible with singly stochastic models of the truth) but also individually for any pair of points in space-time. Besides, the models are linear, allowing the use of the exact Kalman filter as a benchmark in testing new filters.

1. Stochastic Pattern Generator

1.1. Model (tendency) errors

The model error is, by definition, the difference between the model tendency $F(x^t)$ (the model operator F evaluated at the true system state x^t), and the true tendency dx^t/dt. The true model error is normally unknown. Its spatio-temporal probability distribution is mostly unknown either. But the model error is a very important source of the forecast error, so a plethora of ad-hoc techniques has been developed in ensemble applications. These techniques can be classified as non-stochastic (the main techniques in this class are known as multi-model, multi-physics, and multi-parameter) or stochastic (additive perturbations, SPPT, SKEB, and SPP). All the existing stochastic model-error models require a stochastic spatio-temporal pattern generator.

1.2. The SPG equation

This development is motivated by the fact that the existing pattern generators produce fields with *separable* spatio-temporal correlations, that is, without any space-time interactions, specifying the same temporal scale for all spatial scales. We show that these interactions in the model-error spatio-temporal field are important because their specification has significant impact on the structure of the resulting *forecast* errors. We argue that the above "proportionality of scales" is ubiquitous in geosciences and other fields and require it to be satisfied by our SPG. We start with the general Markov model

$$\frac{\partial \xi(t,\boldsymbol{s})}{\partial t} + A\xi(t,\boldsymbol{s}) = \alpha(t,\boldsymbol{s}),$$

where t is time, $\boldsymbol{s}$ = (x; y; z) is the spatial vector, α is the driving white noise, A is the spatial operator, and ξ is the output random field. To get the spatial isotropy, we specify $A = P(-\Delta)$, where Δ is the spatial Laplacian and P is the polynomial or order q. We further simplify P to be of the form

$$A = \mu(1 - \lambda^2\Delta)^q,$$

where μ and λ are the scalar parameters. Imposing the "proportionality of scales" ($\tau_k \sim k^{-1}$ as $k \to \infty$, where k is the total spatial wavenumber and τ_k is the temporal length scale associated with k), we eventually arrive at the SPG equation

$$\left(\frac{\partial}{\partial t} + \mu\sqrt{1 - \lambda^2\Delta}\right)^3 \xi(t, \boldsymbol{s}) = \sigma\alpha(t, \boldsymbol{s}),$$

where the parameter σ controls the variance, λ controls the spatial length scale, and μ controls the temporal length scale. The spatial domain is the 3D or 2D cube with cyclic boundary conditions. The numerical scheme is spectral in space and finite-difference in time. Properties of the generated fields are revealed and illustrations are given.

2. Doubly stochastic models of "truth"

2.1. How to build a doubly stochastic model of "truth"?

Here is the recipe:

1. Take a linear evolutionary model (non-stochastic).
2. Force it with the white noise, getting a singly stochastic model.
3. Make the coefficients of the singly stochastic model random, satisfying their own singly stochastic models with constant coefficients.

2.2. Example

The one-variable doubly stochastic model of "truth" is described in (Tsyrulnikov and Rakitko 2016). Here we outline the model defined on the circle. Following the recipe given above, we start with the non-stochastic advection-diffusion-decay model

$$\frac{\partial x}{\partial t} + U\frac{\partial x}{\partial s} + \rho x - \nu\frac{\partial^2 x}{\partial s^2} = 0,$$

where s is the spatial coordinate on the circle, U is the advection velocity, ρ is the decay coefficient, and ν is the diffusion coefficient. We force this model with the white noise α, getting the singly stochastic model

$$\frac{\partial x}{\partial t} + U\frac{\partial x}{\partial s} + \rho x - \nu\frac{\partial^2 x}{\partial s^2} = \sigma\alpha(t, s), \qquad (1)$$

where $\sigma = e^{\Sigma}$ is the intensity of the driving white noise. Finally, we postulate that the coefficients $U(t,s)$, $\rho(t,s)$, $\nu(t,s)$, and $\Sigma(t,s)$ (or some of them) are spatio-temporal random fields by themselves satisfying the singly stochastic model Eq.(1) (with their own constant coefficients U, ρ, ν, and σ), getting the doubly stochastic model

$$\frac{\partial x}{\partial t} + U(t,s)\frac{\partial x}{\partial s} + \rho(t,s)x - \nu(t,s)\frac{\partial^2 x}{\partial s^2} = e^{\Sigma(t,s)}\alpha(t,s).$$

Once generated, the coefficient fields $U(t,s)$, $\rho(t,s)$, $\nu(t,s)$, and $\Sigma(t,s)$ are fixed, determining the time and space specific statistics of the "truth" (and the error statistics of any filter in question).

2.3. Capabilities of the doubly stochastic models of "truth"

We show that the doubly stochastic models:

- are capable of generating complicated spatio-temporal variability,
- permit computing not only the "true" field but also its "true" time and space specific spatio-temporal statistics (local in space and time variances, length scales, etc.), and
- allow the use of the exact KF (as a benchmark), which facilitates filters' comparisons.

Note that the filter's "statistics of the day" can also be obtained by assuming that the model of "truth" is deterministic, whereas the forecast model is stochastic (Bishop and Satterfield 2013). As compared with that technique, the doubly stochastic approach offers, in addition, the capability to estimate the local statistics of the "truth" and to study the filter's behaviour in different and controlled field regimes.

Conclusions

1. The Stochastic Pattern Generator (SPG) produces pseudo-random spatio-temporal Gaussian fields on 2D and 3D spatial domains. It is based on a linear third-order in time stochastic model driven by the white in space and time Gaussian noise. The spatial operator of the stochastic model is designed to ensure that the generated pseudo-random fields satisfy the "proportionality of scales" property. The Fortran code of the SPG is freely available from https://github.com/gayfulin/SPG.

2. The doubly stochastic models of "truth" are linear evolutionary differential (or difference) models with random forcing and coefficients being random fields by themselves. These models allow studying the performance of any filter in question while knowing the "true" time and space specific statistics of the filter's errors and of the "truth".

References

Bishop, C.H. and Satterfield, E.A. 2013: Hidden error variance theory. Part I: Exposition and analytic model, Mon. Wea. Rev., 141, 1454-1468.

Tsyroulnikov, M.D. 2001: Proportionality of scales: an isotropy-like property of geophysical fields, Quart. J. Roy. Meteorol. Soc., 127, 2741-2760.

Tsyrulnikov, M.D. and Rakitko, A.S. 2016: A Hierarchical Bayes ensemble Kalman Filter, Physica D (Nonlinear Phenomena), in press, doi:10.1016/j.physd.2016.07.009.

OCEAN MODEL, ANALYSIS AND PREDICTION SYSTEM VERSION 3: OPERATIONAL GLOBAL OCEAN WEATHER FORECASTING

Gary Brassington, Paul Sandery, Pavel Sakov, Justin Freeman, Prasanth Divakaran, Duan Beckett, Aihong Zhong, Xinmei Huang, Leon Majewski and Mikhail Entel

Bureau of Meteorology, Melbourne

g.brassington@bom.gov.au

The Ocean Model, Analysis and Prediction System version 3 (OceanMAPSv3) is a near-global (75S-75N; no sea-ice), uniform horizontal resolution (0.1°x0.1°), 51 vertical level ocean forecast system producing daily analyses and 7 day forecasts. This system was declared operational in April 2016, upgraded to include ACCESS-G APS2 in June 2016 and ported to the Bureau's new supercomputer in Sep 2016. This system realises the original vision of the BLUElink projects (2003-2015) to provide global forecasts of the ocean geostrophic turbulence (eddies and fronts) in support of Naval operations as well as other national services. The analysis system has retained an ensemble-based optimal interpolation method with 144 stationary ensemble members derived from a multi-year hindcast. However, the BODAS code has been upgraded to the ENKF-C (Sakov, 2014). A new strategy for initialisation has been introduced leading to greater retention of analysis increments and reduced shock. The analysis cycle has been optimised for a 3-cycle system with 3 day observation windows. The sea surface temperature and sea surface height anomaly analysis errors in the Australian region are 0.34 degC and 6.2 cm respectively an improvement of 10% and 20% respectively over version 2. In addition, the RMSE of the 7 day forecast has lower error than the 1 day forecast from the previous system (version2). International intercomparisons have shown that this system is comparable in performance with the two leading systems and is often the leading performer for surface temperature and upper ocean temperature. In this talk we will present an overview of the system, in particular the data assimilation and initialisation, demonstrate the performance and outline future directions.

ASPECTS OF SUB-MESOSCALE OCEAN ANALYSIS AND FORECASTING

Paul Sandery and Pavel Sakov

Environment and Research Division, Bureau of Meteorology, Melbourne

p.sandery@bom.gov.au

Sub-mesocale resolving ocean models contain dynamical features that appear to be qualitatively similar to patterns seen in satellite imagery. In theory they should offer improvements as they resolve more and parameterise less of the physics. The use of higher resolution models, however, does not necessarily improve forecast skill. Typical observations often used to constrain mesoscale features in eddy resolving models may not be sufficient to constrain sub-mesoscale features. In this context, the benefits of using data assimilating sub-mesoscale forecasting systems remains unclear. Here we carry out side by side reanalysis experiments with a 2.5 km and 10 km resolution regional models to investigate advantages and shortcomings of the different resolutions in ocean forecasting. We find that for the higher resolution system, the mesoscale features can be constrained to the observations whilst permitting sub-mesoscale features to evolve. Counter-intuitively, the higher resolution system could not match the skill of the lower resolution system in forecasting the mesoscale circulation. Whilst predictability at these scales from the higher resolution model appears to be lower, one advantage is an ability to include more information content from higher resolution observations in the analysis. This is shown using qualitative comparisons with AVHRR sea surface temperature and MODIS ocean colour satellite imagery.

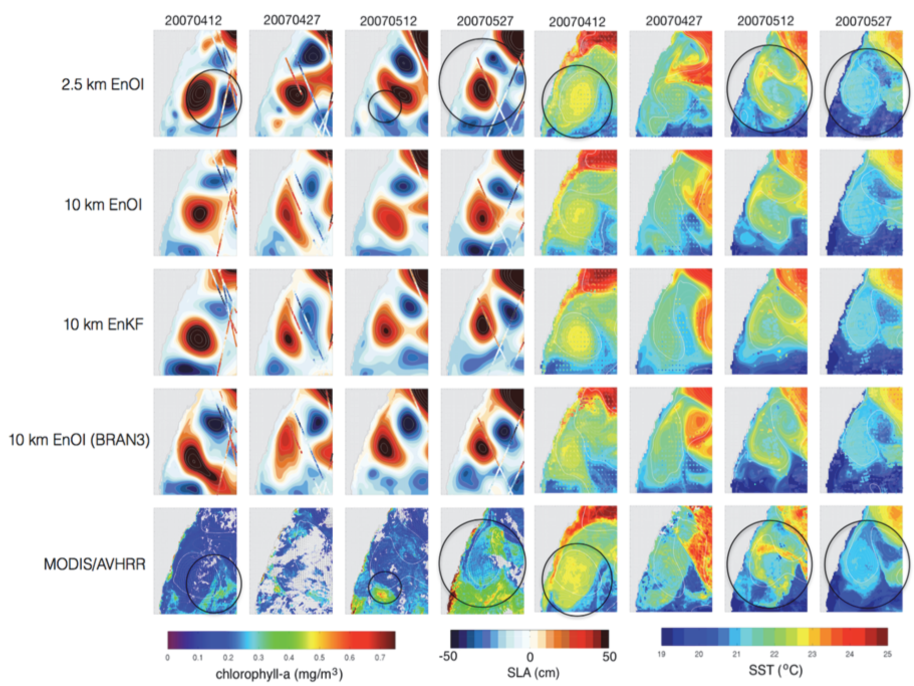


Figure 1: Daily mean sea surface temperature (SST), sea-level anomaly (SLA), Advanced Very High Resolution Radiometer (AVHRR) SST and Moderate Resolution Imaging Spectroradiometer (MODIS) chlorophyll-α at 15 day intervals for a selected period in the reanalysis. EnOI - Ensemble Optimal Interpolation, EnKF - Ensemble Kalman Filter. BRAN - Bluelink Reanalysis.

A HIGH-RESOLUTION REANALYSIS OF THE EAST AUSTRALIAN CURRENT USING ROMS AND 4D-VAR: SYSTEM EVALUATION AND OBSERVATION IMPACT

Colette Kerry[1,] Moninya Roughan[1], Brian Powell[2] , Peter Oke[3]

[1] *Coastal and Regional Oceanography Lab, School of Mathematics and Statistics, UNSW Australia, Sydney*
[2] *Department of Oceanography, School of Ocean and Earth Sciences, University of Hawaii at Manoa, Honolulu, HI, United States*
[3] *CSIRO Marine and Atmospheric Research, Hobart*

c.kerry@unsw.edu.au

As with other Western Boundary Currents globally, the East Australian Current (EAC) is highly variable making it a challenge to model and predict. We have configured a numerical ocean model of the EAC region and combined it with an unprecedented observational data set to generate a high-resolution ocean state estimate over a 2-year period (Jan 2012 - Dec 2013). The numerical model is configured using the Regional Ocean Modelling System (ROMS 3.4), has variable (2.5-6km) horizontal resolution and takes boundary forcing from the BlueLink ReANalysis (BRAN3). In addition to the traditional data streams (satellite derived SSH and SST, Argo profiling floats and XBT lines) we exploit newly available IMOS observations. These include velocity and hydrographic observations from a deep-water mooring array (the EAC transport array, 27.5°S) and several moorings on the continental shelf, high-frequency (HF) radar observations (at Coffs Harbour, 30°S), and ocean gliders.

For the assimilation, we use a time-dependent variational scheme (4D-Var) that uses the model physics to compute increments in the initial conditions, boundary and surface forcings such that the difference between the modelled time-evolving flow and the observations is minimised over a time window. Results show that the reanalysis represents both assimilated and independent (non-assimilated) observations well. Comparison with independent shipboard CTD cast observations shows a marked improvement in the representation of the subsurface ocean, indicating that information is successfully propagated from observed variables to unobserved regions as the assimilation system uses the model dynamics to adjust the model state-estimate.

In solving this state-estimation problem with 4D-Var, we compute the dynamical covariance between the observations and the model that allows us to directly compute the impact of each observation on the circulation estimate. For the reanalysis, we investigate the impact of each data stream on estimates of volume transport in the EAC, focussing on 4 shore normal sections between 27.5 °S and 36 °S. Significantly, we find that the most influential observation platforms are the HF radar off Coffs Harbour and the full depth EAC mooring array, with satellite-derived SSH and SST dominating in the absence of radar and moored observations. Not only do the HF radar observations have high impact on transport estimates at 30°S, they also have significant impact both up and downstream of the radar location. Likewise, the impact of the EAC array is far reaching, contributing to transport estimates hundreds of kilometers downstream of its location at 27.5°S. The observation impact of deep gliders deployed into EAC eddies is also high.

As an extension of this work we have developed a 750-m resolution model of the Sydney-Newcastle region (nested within the EAC model described here) that we will use to investigate the predictability

of shelf dynamics. The prediction of fine scale coastal processes presents additional challenges, compared to the mesoscale, as the circulation is likely to be more rapidly decoupled from the initial state and depend strongly on surface and boundary forcings and model parameters.

ON ASSIMILATING REFLECTANCE INTO MARINE BGC MODELS

Emlyn M. Jones, Mark E. Baird and Mathieu Mongin

CSIRO Oceans and Atmosphere, Hobart

Emlyn.Jones@csiro.au

Skillful marine biogeochemical (BGC) models are required to understand a range of coastal and global phenomena such as changes in nitrogen and carbon cycles. The refinement of BGC models through the assimilation of variables calculated from observed in-water inherent optical properties (IOPs), such as phytoplankton absorption, is problematic. Empirically-derived relationships between IOPs and variables such as Chlorophyll-a concentration (Chl-a), Total Suspended Solids (TSS) and Color Dissolved Organic Matter (CDOM) have been shown to have errors that can exceed 100% of the observed quantity. These errors are greatest in shallow coastal regions, such as the Great Barrier Reef (GBR), due the additional signal from bottom reflectance. Rather than assimilate quantities calculated using error-prone IOP algorithms, this study demonstrates the advantages of assimilating quantities calculated directly from the less error-prone satellite remote-sensing reflectance.

The assimilation of a directly-observed quantity, in this case remote-sensing reflectance, is analogous to the assimilation of temperature brightness in Numerical Weather Prediction (NWP), or along-track sea-surface height in hydrodynamic models. To assimilate the observed reflectance, we use an in-water optical model to produce an equivalent simulated remote-sensing reflectance, and calculate the mis-match between the observed and simulated quantities to constrain the BGC model with a Deterministic Ensemble Kalman Filter (DEnKF). Using the assumption that simulated surface Chl-a is equivalent to remotely-sensed OC3M estimate of Chl-a resulted in a forecast error of approximately 75%. Alternatively, assimilation of remote-sensing reflectance resulted in a forecast error of less than 40%. Thus, in the coastal waters of the GBR, assimilating remote-sensing reflectance halved the forecast errors.

When the analysis and forecast fields from the assimilation system are compared with the non-assimilating model, an independent comparison to in-situ observations of Chl-a, TSS, and dissolved inorganic nutrients (NO_3, NH_4 and DIP) show that errors are reduced by up to 90%. In all cases, the assimilation system improves the result compared to the non-assimilating model. This approach allows for the incorporation of vast quantities of remote-sensing observations that have in the past been discarded due to shallow water and/or artefacts introduced by terrestrially-derived TSS and CDOM, or the lack of a calibrated regional IOP algorithm.

HIERARCHICAL BAYES ENSEMBLE KALMAN FILTERING

Michael Tsyrulnikov* and Alexander Rakitko*

**HydroMetCentre of Russia*

michael.tsyrulnikov@gmail.com

Motivation

In the Kalman filter (KF), the analysis is computed by applying the very well known Kalman gain matrix

$$\boldsymbol{K}=\boldsymbol{BH}^{\mathrm{T}}(\boldsymbol{HBH}^{\mathrm{T}}+\boldsymbol{R})^{-1} \qquad (1)$$

to the innovation vector. The resulting analysis is mean-square optimal as long as $\boldsymbol{B}$ is the *true* background error covariance matrix. The ensemble Kalman filter (EnKF) relies on the same Kalman gain matrix, with $\boldsymbol{B}$ replaced by the ensemble (sample) covariance matrix $\boldsymbol{S}$. In high-dimensional applications, there are two issues that make $\boldsymbol{S}$ significantly different from the true $\boldsymbol{B}$:

1. Sampling noise (due to the inevitably small ensemble size).

2. Systematic errors (due to time accumulated filter's imperfections, with respect to the true time specific $\boldsymbol{B}$).

The first issue is addressed in any EnKF application by introducing a kind of localization to the sample covariances (in a large variety of techniques in model space, in observation space, in ensemble space, in inverse space, etc.) or by mixing the sample covariances with static ones (in the EnVar) or by spatially filtering/smoothing the sample covariances.

Concerning the second issue, by systematic errors we mean the inevitable (for any sub-optimal filter) difference between the covariance matrix of the probability distribution the forecast ensemble members are actually drawn from (at each specific time instant) and the true $\boldsymbol{B}$ at the same time instant. This issue has attracted much less attention: to our knowledge, there is only one device in common use: covariance inflation.

What is important for us in this research is that

1. All the above covariance "pre-processing" (covariance regularization) techniques are not based on a theoretic optimality criterion targeting the accuracy of the resulting analysis.

2. The remaining errors in the "pre-processed" $\boldsymbol{S}$ make the application of the Kalman gain matrix sub-optimal.

The idea and the ultimate goal of this research is to extend the existing optimality criterion, which optimizes the weights of the background and the observations under the tacit assumption that the "pre-processed" covariances are exact, to the criterion in which this latter assumption is not made and the handling of the covariances is determined in the resulting optimization process.

We can approach this idea from another perspective. Over years and decades, it has become evident

that the most fruitful approach to data assimilation has been the Bayesian paradigm. According to this approach,

1. The true state $\mathbf{x}$ is assumed random, and a prior distribution of $\boldsymbol{x}$ given the forecast (background) $\mathbf{x}^{\mathrm{f}}$ is introduced. In the KF and EnKF, this distribution is multivariate Gaussian: $\mathrm{p}(\boldsymbol{x}|\boldsymbol{x}^{\mathrm{f}}) \sim \mathbf{N}(\boldsymbol{x}^{\mathrm{f}},\boldsymbol{B})$.

2. The observation likelihood $\mathrm{p}(\boldsymbol{y}|\boldsymbol{x})$ is specified.

3. The posterior probability density $\mathrm{p}(\boldsymbol{x}|\boldsymbol{x}^{\mathbf{f}},\boldsymbol{y})$ is computed, from which the optimal analysis $\mathbf{x}^{\mathrm{a}}$ and the analysis ensemble are derived.

As the background covariance matrix $\boldsymbol{B}$ is significantly in error, a switch from the Bayesian paradigm to the *hierarchical* Bayesian paradigm is justified. According to the hierarchical Bayesian approach, the parameters of the prior distribution, in our case, the mean and the covariance matrix $\boldsymbol{B}$, are also assumed to be random and may, and in our view, should be subject to a Bayesian update along with the state. This means that the so-called secondary filter, which treats the covariances, is also to be based on the Bayesian estimation.

First steps towards this new paradigm were done by Myrseth and Omre (2010), who explicitly assumed in their Hierarchical EnKF (HEnKF) that $\boldsymbol{B}$ is a random matrix and updated its prior probability distribution using the ensemble. Bocquet (2011) treated $\boldsymbol{B}$ as a random nuisance parameter, whose role is to change the Gaussian prior distribution of the state $\mathbf{x}$ to a more realistic continuous mixture of Gaussians. Here, we present the next step: our Hierarchical Bayes Ensemble Filter (HBEF), which treats the ensemble members as generalized observations on the covariances and allows observations to influence the covariances.

1. Setup

We formulate the HBEF for the linear dynamics and linear observations. We split $\boldsymbol{B}_{\mathrm{k}}$ (k is the time index) into the model error covariance matrix $\boldsymbol{Q}_{\mathrm{k}}$ and the predictability error covariance matrix $\boldsymbol{P}_{\mathrm{k}} = \boldsymbol{F}_{\mathrm{k}}\boldsymbol{A}_{\mathrm{k-1}}\boldsymbol{F}^{\mathrm{T}}_{\mathrm{k}}$ (where $\boldsymbol{F}_{\mathrm{k}}$ is the forecast model operator and $\boldsymbol{A}_{\mathrm{k-1}}$ is the previous-cycle analysis error covariance matrix). The reason for such splitting is the fundamentally different nature of model errors (which are external to the filter) vs. predictability errors (which are internal, i.e. determined by the filter). This splitting is found to be beneficial in numerical experiments (not shown). Correspondingly, we split the forecast ensemble into the model error ensemble and the predictability ensemble.

Observation errors are assumed to be Gaussian. Other settings come, mainly, from the formulation of conditions under which the EnKF actually works in geophysical applications:

1. The ensemble size is too small for the sample covariance matrices to be accurate estimators.
2. The direct computation of the predictability covariance matrix $\boldsymbol{P}_{\mathrm{k}}$ is not feasible.
3. The model error covariance matrix $\boldsymbol{Q}_{\mathrm{k}}$ is temporally variable and explicitly unknown.

We also hypothesize that

4. Conditionally on $\boldsymbol{Q}_{\mathrm{k}}$, the model errors are zero-mean Gaussian.
5. We can draw independent pseudo-random samples from $\mathrm{N}(\mathbf{0},\boldsymbol{Q}_{\mathrm{k}})$ with the true $\boldsymbol{Q}_{\mathrm{k}}$.

Under these assumptions, the KF theory cannot be applied. In this research, we propose a theory and

design a filter (the HBEF) that acknowledge in a more systematic way than this is done in the EnKF that the covariance matrices $\boldsymbol{P}_k$ and $\boldsymbol{Q}_k$ are substantially uncertain. We regard $\boldsymbol{P}_k$ and $\boldsymbol{Q}_k$ as additional (to the state $\boldsymbol{x}_k$) random matrix variables to be estimated along with the state following the hierarchical Bayes paradigm.

2. Formulation of the HBEF

We describe how the prior distributions for the $\boldsymbol{P}_k$ and $\boldsymbol{Q}_k$ matrices are specified. We discuss the choice of the so-called inverse Wishart matrix variate distribution. We show that the conditional Gaussian distribution of the state $\mathbf{x}_k$ is preserved in the course of the filtering.

Then, we demonstrate that ensemble members can indeed be regarded as generalized observations on the respective covariance matrices. This is because the conditional Gaussianity of the ensemble members implies the existence of the probability density of the ensemble members given the covariances, i.e the likelihood of the covariances needed for their Bayesian update. We address the problem of the fundamental imperfection of the predictability ensemble.

After that, we write down the posterior probability density of the extended control variable ($\boldsymbol{P}_k$,$\boldsymbol{Q}_k$,$\boldsymbol{x}_k$). The conditional Gaussianity of $\boldsymbol{x}_k$ given $\boldsymbol{P}_k$ and $\mathbf{Q}_k$ greatly simplifies the resulting analysis equations. In their non-approximated Monte-Carlo based form, the mean-square optimal posterior estimates (i.e. the deterministic analyses) of $\boldsymbol{P}_k$, $\boldsymbol{Q}_k$, and $\boldsymbol{x}_k$ are computed using importance sampling. In the simplest approximated form of the posterior, the deterministic analyses of $\boldsymbol{P}_k$ and $\boldsymbol{Q}_k$ are computed as linear combinations of the respective prior covariance matrices $\boldsymbol{P}_k^f$ and $\boldsymbol{Q}_k^f$ and the respective sample covariances:

$$\boldsymbol{P}_k^a = \frac{\chi \boldsymbol{P}_k^f + N\boldsymbol{S}_k^{pe}}{\chi + N} \qquad and \qquad \boldsymbol{Q}_k^a = \frac{\phi \boldsymbol{Q}_k^f + N\boldsymbol{S}_k^{me}}{\phi + N} \tag{2}$$

where N is the ensemble size, χ and ϕ are the scalar parameters, $\boldsymbol{S}_k^{pe}$ is the predictability ensemble covariance matrix, and $\boldsymbol{S}_k^{me}$ is the model error ensemble covariance matrix. The prior covariance matrices are propagated from the previous analysis cycle using persistence: $\boldsymbol{P}_k^f = \boldsymbol{P}_{k-1}^a$ and $\boldsymbol{Q}_k^f = \boldsymbol{Q}_{k-1}^a$. Note that the linear combinations of the prior and ensemble covariances in Eq.(2) closely resemble the mixing of climatological and ensemble covariances in the EnVar. Equation (2) also implies that the simplest version of the HBEF still does require the covariance regularization.

The deterministic analysis of the state $\boldsymbol{x}_k^a$ is computed using the KF analysis equation with the gain matrix Eq.(1) computed with $\boldsymbol{B} = \boldsymbol{P}_k^a + \boldsymbol{Q}_k^a$. The analysis ensemble generation technique is borrowed from the stochastic EnKF.

3. Numerical experiments

We formulate two doubly stochastic linear models of "truth". Their general form is

$$\boldsymbol{x}_k = \boldsymbol{F}_k \boldsymbol{x}_{k-1} + \sigma_k \boldsymbol{\varepsilon}_k, \tag{3}$$

where $\boldsymbol{\varepsilon}_k$ is the discrete-time driving white noise and $\boldsymbol{F}_k$ and σ_k are modeled to be random by themselves. One such model is a one-variable (i.e. scalar) model and the other one is a model defined for a field $\boldsymbol{x}$ on the circle.

Importantly, with the model of truth Eq.(3), the true prior covariances $\boldsymbol{B}_{\mathrm{k}}$ and the signal covariances $\boldsymbol{V}_{\mathrm{k}}$ can be estimated as accurately as needed. This is achieved by averaging over L independent runs, in which the sequences of $\boldsymbol{F}_k$ and σ_k (as well as the sequence of the observation operators) are the same (thus preserving the specificity of each time instance), whereas the "true" and simulated model and observation errors are simulated in each run randomly and independently from the other runs.

3.1. Verifying the primary filters

Figure 1 shows the analysis RMSEs obtained with the model of "truth" on the circle for the simplest version of the HBEF, the variational filter (Var), and the stochastic EnKF. It is seen that the HBEF is by far the best filter. For small $N \leq 8$, the Var filter becomes more competitive than the EnKF, but remains worse than the HBEF.

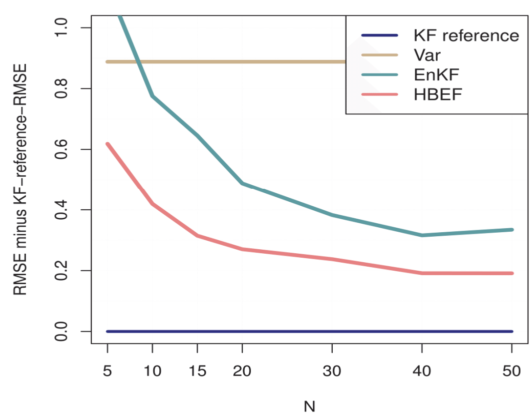


Fig. 1: Analysis RMSEs (with the reference-KF analysis RMSEs subtracted) as functions of the ensemble size *N*.

3.2. Verifying the secondary filters

Each filter produces estimates of its own $\boldsymbol{B}$, which can be compared with the "true" $\boldsymbol{B}$ for this specific filter at each time instant. The resulting RMS errors in the background error variance with the one-variable model of "truth" are shown in Fig.2 for the EnKF and the simplest version of the HBEF. The almost uniform and substantial superiority of the HBEF is evident.

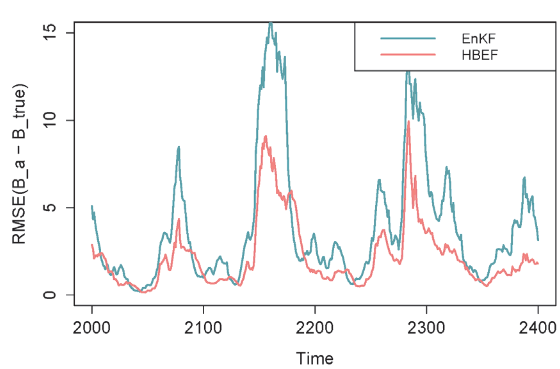


Fig. 2: RMSEs in the background error variances produced by the EnKF and the HBEF.

Conclusions

- We have acknowledged that in most applications, the EnKF works with: (i) the explicitly unknown and variable model error covariance matrix $\boldsymbol{Q}_\mathrm{k}$, (ii) the partially known (through ensemble covariances) background error covariance matrix. Under these explicit restrictions, we have proposed a new Hierarchical Bayes Ensemble Filter (HBEF) that optimizes the use of observational and ensemble data by treating $\boldsymbol{Q}_\mathrm{k}$ and the predictability covariance matrix $\boldsymbol{P}_\mathrm{k}$ as random matrices to be estimated in the analysis along with the state. The ensemble members are treated in the HBEF as generalized observations on the covariance matrices.
- With the new HBEF filter, in the course of filtering, the prior and posterior distributions of the state remain conditionally (given $\boldsymbol{P}_\mathrm{k}$ and $\mathbf{Q}_\mathrm{k}$) Gaussian provided that: (i) it is so at the start of the filtering, (ii) observation errors are Gaussian, (iii) the dynamics and the observation operators are linear, and (iv) model errors are conditionally Gaussian given $\boldsymbol{Q}_\mathrm{k}$. Unconditionally, the prior and posterior distributions of the state are non-Gaussian.
- The HBEF is thoroughly tested with a one-variable doubly stochastic model of truth. The model has the advantage of providing the means to assess the instantaneous variance of the truth and the true filter's error variances. The HBEF is found superior the EnKF and the HEnKF under most regimes of the system, most data assimilation setups, and in terms of performance of both primary and secondary filters. Similar results are obtained with a doubly stochastic model of "truth" on the circle.
- It is shown that the HBEF's feedback from observations to the covariances can be beneficial.
- The simplest version of the HBEF is designed to be affordable for practical high-dimensional applications on existing computers. To this end, the $\boldsymbol{P}_\mathrm{k}$ and $\mathbf{Q}_\mathrm{k}$ matrices need to be propagated from one analysis time to the next; this can be achieved either by storing the localized covariances on a coarse grid or by using an (estimated at each analysis time) parametric model for the covariances.

References

Bocquet, M. 2011: Ensemble Kalman filtering without the intrinsic need for inflation, Nonlin. Proc. Geophys., 18, 735-750.

Myrseth, I. and Omre, H. 2010: Hierarchical ensemble Kalman filter, SPE Journal, 15, 569-580.

Tsyrulnikov, M.D. and Rakitko, A.S. 2016: A Hierarchical Bayes ensemble Kalman Filter, Physica D (Nonlinear Phenomena), in press, doi:10.1016/j.physd.2016.07.009.

SATELLITE SST ASSIMILATION INTO AN OCEAN MODEL (SHOC) USING 4D-VAR

Chaojiao Sun[1], Peter Oke[2], Alex Kurapov[3], and Lars Nerger[4]

[1]*Oceans and Atmosphere, CSIRO, Perth, Western Australia*
[2]*Oceans and Atmosphere, CSIRO, Hobart, Tasmania*
[3]*College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon, United States*
[4]*Alfred Wegener Institute, Helmholtz Center for Polor and Marine Research, Bremerhaven, Germany*

Chaojiao.Sun@csiro.au

Satellite sea surface temperature (SST) observations off the Bonney Coast in South Australia were assimilated using a coastal ocean data assimilation (DA) system. The coastal region off the Bonney Coast has a vigorous seasonal upwelling system that sustains a productive ecosystem that attracts blue whales and supports rich fisheries and Australian fur seals. We have developed the 4D-Var ocean DA system to improve the simulation and forecasts of these upwelling events by assimilating SST observations. Our DA system is unique in that we use a set of stand-along adjoint (ADJ) and tangent linear (TL) model codes that are dynamically consistent with an ocean model, and a different ocean model for the nonlinear model for free-run and forecasts. The results show that the assimilation improves the accuracy of analyses and forecasts of observed upwelling events along the Bonney Coast. This study thus demonstrates that variational data assimilation is feasible for an ocean model that does not have its own TL and ADJ codes, using variational codes from another ocean model with similar physics.

Observation and method

We assimilate the UK Met Office (UKMO) Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) Level 4 (gap-free) global SST daily product at about 5 km resolution. OSTIA SST is determined by assimilating SST data from several satellites provided by the Group for High Resolution Sea Surface Temperature Project (GHRSST) and in-situ observations using a variant of optimal interpolation (OI), blending several Level 2 products together while correcting bias and reducing noise (Martin et al., 2007).

In our DA system, we use the Sparse Hydrodynamic Ocean Code (SHOC) as the nonlinear model, whose own adjoint (ADJ) and tangent linear (TL) model codes have not been developed. Instead, we use a set of stand-alone ADJ and TL codes, the Advanced Variational Regional Ocean Representer Analyzer (AVRORA; Kurapov et al 2011), which are dynamically consistent with another ocean circulation model, the Regional Ocean Modeling System (ROMS).

The initial and boundary conditions are derived from the 10-km Bluelink ReANalysis (BRAN) version 3p5. The air-sea fluxes are from the ECMWF Interrim Reanalysis (ERA-Interim). The variational representer method is implemented in a series of 2-day time windows, with initial conditions corrected at the beginning of each window. A 4-day forecast is then run with SHOC using the corrected initial conditions. In the next assimilation window, the last two days of the 4-day forecasts are used as the background for the AVRORA linearization and the same assimilation procedure is repeated. The assimilation and forecasts are cycled for the austral summer month of February 2012.

Results

The time-evolving circulation along the Bonney coast in the month of February 2012 simulated by SHOC on a 3km horizontal grid has mostly captured upwelling events during this time period. However, the magnitude of SHOC model cooling was much weaker than that observed SST. The assimilation significantly improved the upwelling signals both in analysis and forecast (Fig. 1). The root-mean-square error (RMSE) of the model free run is much higher than the forecast and analysis, with the analysis having the highest correlation (not shown).

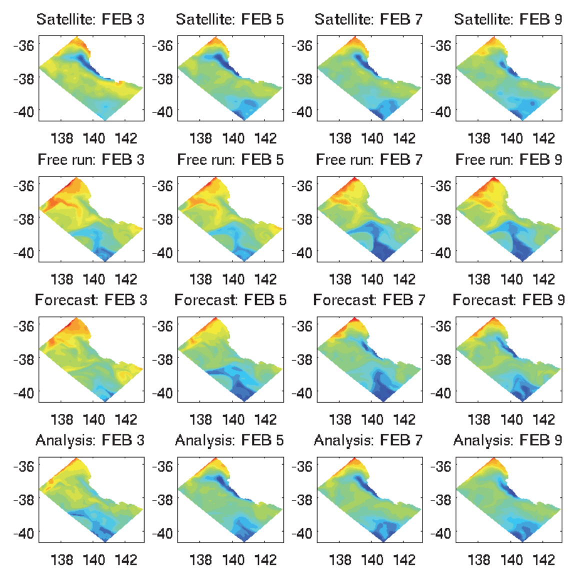


Figure 1: Daily averaged SST from satellite observations (top), SHOC prior solution (second row), SHOC forecast (third row), and SHOC posterior analysis (bottom) for February 3, 5, 7, and 9.

Discussion

The novelty of this study is the use of the ADJ and TL codes developed for ROMS to assimilate SST observation into SHOC, which improves the analysis and forecast of the upwelling signals along the Bonney Coast during February 2012. It is conceivable that the solution will not converge if the assimilation window is too long given the different implementations in the two models (Lorenc A., pers. comm.).

References

Kurapov, A.L., Foley, D., Strub, P.T., Egbert, G.D. and Allen, J.S. 2011: Variational assimilation of satellite observations in a coastal ocean model off Oregon. Journal of Geophysical Research-Oceans, 116, 5006-5006.

Martin, M. and Coauthors. 2012: Group for High Resolution Sea Surface temperature (GHRSST) analysis fields inter-comparisons. Part 1: A GHRSST multi-product ensemble (GMPE). Deep-Sea Res. Part II-Top. Stud. Oceanogr., 77-80, 21-30.

ITERATIVE ENSEMBLE KALMAN FILTER IN PRESENCE OF MODEL ERROR

Pavel Sakov[1], Marc Bocquet[2] , and Jean-Matthieu Haussaire[2]

[1]*Environment and Research Division, Bureau of Meteorology, Melbourne*
[2]*CEREA, Joint laboratory Ecole des Ponts ParisTech and EDF R&D, Universite Paris-Est, Champs-sur-Marne, France*

pavel.sakov@bom.gov.au

The analysis step in Kalman filter (KF) can be seen as a single iteration of the Gauss-Newton minimisation of a nonlinear cost function. It works well in linear or weakly nonlinear cases, but becomes increasingly suboptimal as the system's nonlinearity increases. The same limitation applies to the ensemble Kalman filter (EnKF, Evensen, 1994), which represents a state space formulation of the KF suitable for large-scale applications.

The iterative ensemble Kalman filter in stochastic (EnRML, Gu and Oliver, 2007) or deterministic (IEnKF, Sakov et al., 2012) frameworks represents a generalisation of the EnKF for strongly nonlinear systems. It aims to take advantage of observations assimilated in the current cycle for reducing the uncertainty in the state at the time of the previous analysis, which in turn reduces the nonlinearity of the system. Formally it represents a derivative-less Gauss-Newton solver, with a number of options for approximating the derivatives via ensemble.

The IEnKF algorithm in Sakov et al. (2012) assumes the perfect model case, when evolution of a model state is deterministic. This assumption simplifies updating the system at a time different to the observation time (Evensen and van Leeuwen, 2000; Hunt et al., 2004; Sakov et al., 2010) and makes it possible to apply the IEnKF for smoothing (IEnKS, Bocquet and Sakov, 2014).

In this study we generalise the IEnKF for the case of imperfect model with additive model error.

References

Bocquet, M. and Sakov, P. 2014: An iterative ensemble Kalman smoother. Q. J. R. Meteorol. Soc., 140, 1521–1535.

Evensen, G. 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. J. Geophys. Res., 99, 10143–10162.

Evensen, G. and van Leeuwen, P.J. 2000: An ensemble Kalman smoother for nonlinear dynamics. Mon. Wea. Rev., 128, 1852–1867.

Gu, Y. and Oliver, D.S. 2007: An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. SPE Journal , 12, 438–446.

Hunt, B.R., E. Kalnay, E.J. Kostelich, E. Ott, D.J. Patil, T. Sauer, I. Szunyogh, J.A. Yorke, and A.V. Zimin, 2004: Four-dimensional ensemble Kalman filtering. Tellus, 56A, 273–277.

Sakov, P., Evensen, G. and Bertino, L. 2010: Asynchronous data assimilation with the EnKF. Tellus, 62A, 24–29.

Sakov, P., Oliver, D.S. and Bertino, L. 2012: An iterative EnKF for strongly nonlinear systems. Mon. Wea. Rev., 140, 1988–2004.

THE GIGG-DELTA FILTER: DATA ASSIMILATION FOR EPISODIC VARIABLES WITH SKEWED UNCERTAINTY DISTRIBUTIONS LIKE CLOUD, PRECIPITATION, FIRE AND ICE.

Craig H. Bishop

Naval Research Laboratory, Monterey, CA, USA

Craig.Bishop@nrlmry.navy.mil

The uncertainty distributions of forecasts of episodic variables such as clouds, precipitation, fire and ice often feature a finite probability of non-existence and/or skewness. For such variables, existing data assimilation techniques such as 4DVAR, the Ensemble Kalman Filter (EnKF) and the Particle Filter (PF) fail to yield states that lie within known observational uncertainty bounds when a finite amount of the variable is observed but the variable is uniformly absent (zero) in the prior forecast guess and/or ensemble. Here we extend the previously developed Gamma, Inverse-Gamma and Gaussian (GIGG) variation on the EnKF to accommodate finite probabilities of variable non-existence using a gamma function based variation of Dirac's delta function. The resulting GIGG-Delta filter has the remarkable property that when rain (for example) is observed, the GIGG-Delta filter always produces a posterior ensemble of raining ensemble members that lie within observational uncertainty bounds even when not one of the prior forecast ensemble members contained rain. In addition, the GIGG-Delta filter accurately solves Bayes' theorem when the prior and observational uncertainties are given by gamma and inverse-gamma pdfs, respectively. To improve the multi-variate dynamical balance of the posterior distributions, a new ensemble based iterative balancing procedure is proposed that improves the ability of EnKFs to produce balanced posterior states. The approach is tested, illustrated and compared with existing techniques using a hierarchy of idealized models.

COUPLED DA IN CCFS : A PROTOTYPE MULTI-YEAR TO DECADAL PREDICTION SYSTEM

Terence J. O'Kane[1], Paul A. Sandery[2], Pavel Sakov[2], Matthew A. Chamberlain[1] & Richard Matear[1]

[1]*CSIRO Oceans and Atmosphere, Hobart*
[2]*Bureau of Meteorlogy, Docklands, Melbourne*

terence.okane@csiro.au

Recently CSIRO has committed substantial funds (15 FTE over 10 years) to develop the fundamental science to underpin a national capability in climate forecasting on time scales from 1 to 10 years. In this presentation we outline the initial stages of the development of this capability. The current prototype system comprises 3 components: coupled model, data assimilation and ensemble prediction.

Model: We at present are using the GFDL CM2.1 GCM for development and process studies but in parallel are configuring the ACCESS CM2 model. The GFDL CM2.1 model is using the AusCOM ocean grid, a nominally 1^{o} tripolar grid with horizontal refinement in the tropics and at higher latitudes, a 28 level atmosphere (am2) and sea ice model (SIS). ACCESS CM2 has a $1/4^{o}$ ocean, 74 level atmosphere (UM) and the Los Alamos CICE sea ice model. ACCESS CM2 will be the model of choice in the mature system.

Data assimilation: Surface (satellite sea surface temperature, salinity and height) and subsurface (Argo, XBT, CTD, TAO array) ocean observations are assimilated while the atmospheric prognostic variables (specific humidity, surface pressure, air temperature, meridional and zonal winds) are nudged to the observed large-scale reanalyzed atmosphere. Two assimilation cycles are employed; first the ocean observations are assimilated with typical localization length scales of 400-800m, followed by a second assimilation to determine appropriate covariances between ocean observations and atmospheric observables. This second assimilation employs localization length scales of approximately 4000m and is tuned such that the atmospheric increments are comparable to the tendency. Background covariances are a combination of seasonally varying static covariances from a long control, and flow dependent covariances from a large ensemble of dynamic vectors supplied by the ensemble predictions system.

Ensemble prediction: An ensemble of coupled bred vectors is generated at each assimilation step and evolved forward. The evolved perturbations are rescaled to an appropriate predetermined norm. Two classes of bred vectors are generated: the first are on the timescale of the assimilation windows and are used to update the assimilation background covariances in the data assimilation and for initial conditions for forecasts. The second are generic allowing any particular level, region or even grid point to be perturbed and rescaled in isolation thereby allowing targeted studies of the predictability of particular modes and disturbances on a range of spatio-temporal scales.

Multi-year to decadal prediction is a nascent and challenging field in climate science. Much remains unknown as to what is predictable and on what spatio-temporal scales, while the relative importance of initial conditions versus the forced response is largely unexplored. In addition how to construct coupled data assimilation procedures capable of capturing the long time scale variability of the ocean and the relatively fast atmospheric response is a largely open question. This work is an attempt to try to understand such issues.

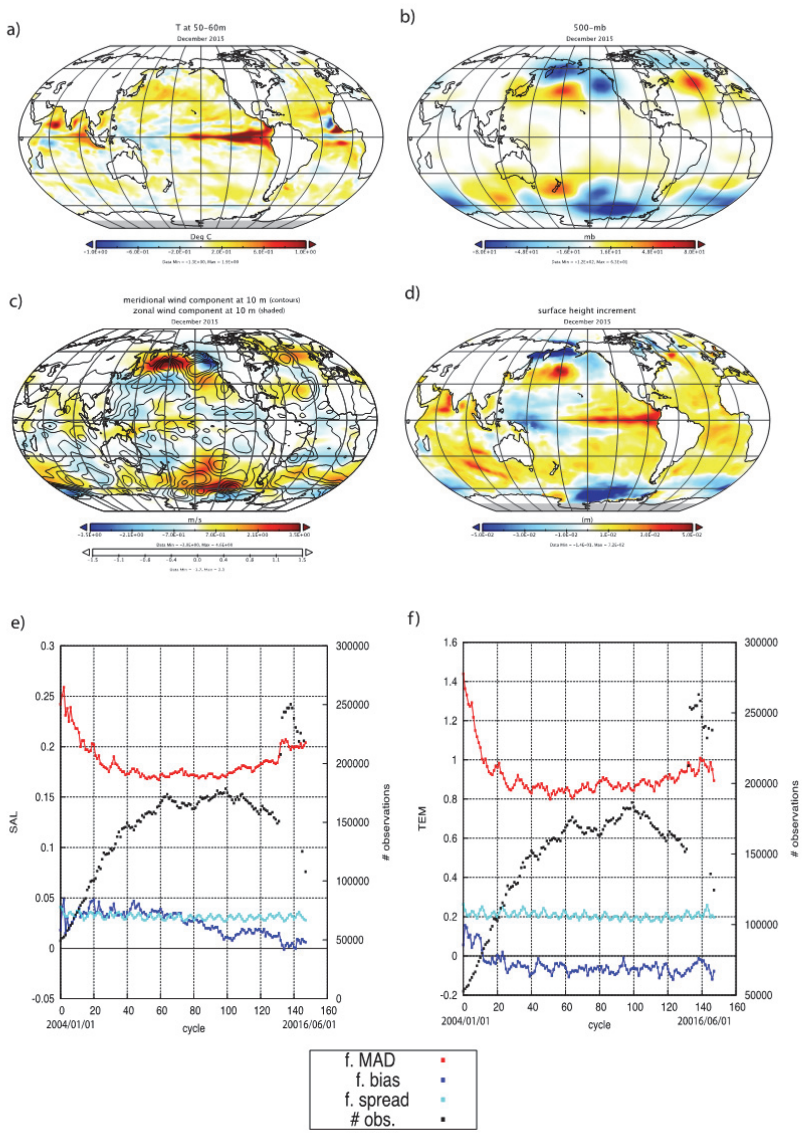


Figure 1: a) 50-60 ocean temperature increment, b) 500mb geopotential height increment, c) meridional and zonal 10m wind increments, d) sea surface height (SSH) increment. In panels a) & d) we show typical increments for ocean temperature and SSH from the ocean data assimilation. In panels b) & c) we show atmospheric increments based on covariances between ocean observations and the atmosphere. In e) salinity and f) temperature, we show globally integrated (full depth) statistics for the ocean assimilation from 2004 to present.

COUPLED DATA ASSIMILATION IN ACCESS-S

Angus Gray-Weale, Yonghong Yin, Oscar Alves, Pavel Sakov, Debra Hudson, Xiaobing Zhou, Hailin Yan, Mei Zhao

Environment and Research Division, Bureau of Meteorology, Melbourne

angus.gray-weale@bom.gov.au

We report a combination of data assimilation methods configured for seasonal prediction in the ACCESS-S system. We discuss in detail a prototype for ACCESS-S2, which contains features planned for inclusion in ACCESS-S3. Distinct weakly coupled algorithms are used to assimilate data on the atmosphere, on sea surface salinity and temperature, and on the bulk ocean into one trajectory, the central member of an ensemble. The system is compared to its predecessor, ACCESS-S1, which uses initial conditions supplied by the UK Met Office.

Observational data on ocean temperature and salinity are assimilated daily using the ensemble optimal interpolation method as implemented in the EnKF-C code.[1] Full details of this part of the calculation will be provided in a separate talk.

The sea-surface salinity and temperature are controlled using differential equations as implemented in NEMO 3.4.[2] In the example of sea-surface temperature, the surface heat flux is modified in proportion to the deviation of model sea surface temperature from a relevant observational field. The constant of proportionality is calculated from a relaxation time and the mixed layer depth. Here the relaxation time is, roughly speaking, the timescale of fluctuations of model sea surface temperature around the observational field. A closely analogous method is used for the sea-surface salinity. We choose temperature and salinity relaxation times of 1 day and 1 year respectively.

The atmosphere's zonal and meridional velocity fields, its potential temperature, and humidity are 'nudged' towards ERA Interim reanalyzed fields once per day. Each 'nudge' shifts each of these four fields according to,

$$X_{n+1} = \alpha\, Y_{n+1} + (1 - \alpha)\, x_n,$$

where x_n is the output from the $(n - 1)$th day of simulation, X_{n+1} is the input to the nth day of simulation, and Y_{n+1} is the relevant ERA Interim field. We typically choose α to be equal to 1, i.e. the model field is replaced by ERA Interim daily.

The ensemble typically consists of ten trajectories in addition to the central member. These are constrained daily so that the means of all assimilated variables follow the central member, and so that their spread around the central member is controlled. This algorithm is compared to an earlier, similar approach.[3]

A first set of hindcasts using this system is described, and possible improvements to coupling in the assimilation methods for ACCESS-S3 are discussed.

---

[1] https://github.com/sakov/enkf-c.git

[2] Madec, G., and the NEMO team, 2008: NEMO ocean engine. Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No 27, ISSN 1288-1619, p134.

[3] Hudson, D., Marshall, A. G., Yin, Y., Alves, O., Hendon, H., 2013: Improving Seasonal Prediction with a New Ensemble Generation Strategy, Mon. Wea. Rev., **141**, 4429-49.

OCEAN DATA ASSIMILATION IN ACCESS-S2

Yonghong Yin, Angus Gray-Weale, Oscar Alves, Pavel Sakov, Debra Hudson, Xiaobing Zhou, Hailing Yan and Mei Zhao

Environment and Research Division, Bureau of Meteorology, Melbourne

y.yin@bom.gov.au

ACCESS-S, the seasonal prediction version of the Australian Community Climate and Earth-System Simulator (ACCESS), is the next generation sub-seasonal to seasonal forecasting system in the Bureau of Meteorology. The coupled model in ACCESS-S is the global coupled model developed by the UK Met Office (UKMO GC2 or GC3 etc.). This model combines the atmospheric model UM (Unified Model) at N216 (~60 km in the mid-latitudes) horizontal resolution on 85 vertical levels, with the land surface model JULES which has 4 soil levels, the ocean model NEMO (Nucleus for European Modelling of the Ocean) at 25 km (at the equator) horizontal resolution on 75 vertical levels, and the sea-ice model CICE at the same resolution as NEMO. The atmosphere and the ocean/sea-ice are coupled every 3-hours using the OASIS coupler.

The development of ACCESS-S will undergo three stages which are named ACCESS-S1, ACCESS-S2 and ACCESS-S3 respectively. We are currently developing a coupled data assimilation (CDA) system for ACCESS-S which is based on an Ensemble Kalman Filter (EnKF) code called EnKF-C developed by Pavel Sakov (Sakov 2015), written in C and efficient on massively parallel systems. A preliminary version of the CDA system for ACCESS-S2 has been designed to be weakly coupled, to have seasonally varying static background error covariance fields and to only assimilate ocean observations into the coupled model (the atmospheric component is nudged towards a pre-exiting atmospheric analysis). The ultimate goal of this development is fully coupled EnKF for data assimilation and ensemble generation in ACCESS-S3.

The ensemble optimal interpolation (EnOI, Evensen, 2003) we will be implementing in the ACCESS-S2 ocean data assimilation is closer to 3D-Var and can be defined as the EnKF with a static or pre-defined ensemble. In contrast to the EnKF, due to the use of a static ensemble, the EnOI avoids potential problems related to the ensemble spread such as underestimates of the ensemble variance which can lead to poor performance. The seasonally varying static ensemble anomalies are generated by removing the 3-month running mean (keeping the intra-seasonal anomalies) from the 23-year (1990-2012) initial conditions of the 1st, 9th, 17th and 25th day of every month for GC2 from the UK Met Office. The standard ensemble size is 92 (23×4) for each month, and an augmented ensemble size of 184 can be obtained by including half of the ensemble members from the closest ensembles from the previous and the next month. Sea surface temperatures (SST) in the ocean model are strongly relaxed to a daily SST analysis (Reynolds et al 2007). The in situ ocean observations, including temperature and salinity profiles from Argo, XBT, CTD and Moorings (e.g., Fig. 1) sourced from EN4 (Good et al 2013, EN4.2.0, Gouretski and Reseghetti (2010) corrections), are assimilated and the analyses for temperature, salinity, u and v currents are computed utilizing the ensemble-based covariances.

This presentation will describe the ocean data assimilation system and some preliminary validation and results obtained from experiment runs.

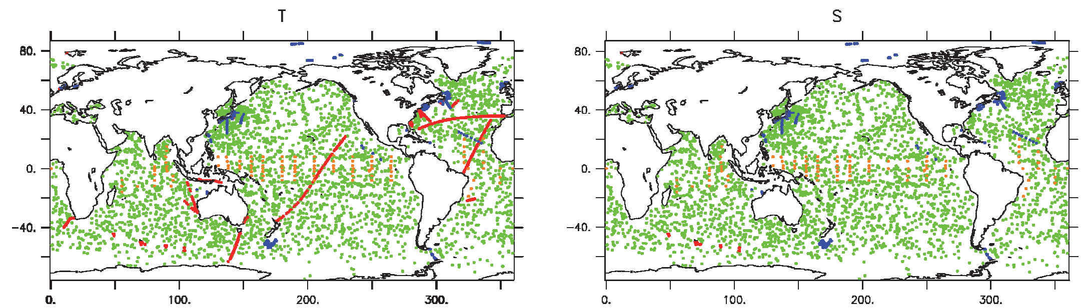


Fig.1 Spatial distribution of the assimilated ocean observations for temperature (left) and salinity (right) during the period of 1st-10th December 2011 (Argo in green, XBT in red, CTD in blue and Moorings in orange.

References

Evensen, G. 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean Dyn., 53, 343–367

Good, S.A., M.J. Martin and N.A. Rayner, 2013. EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, J. Geophys. Res., 118C, 6704-6716, doi:10.1002/2013JC009067

Gouretski, V. and F. Reseghetti, 2010: On depth and temperature biases in bathythermograph data: development of a new correction scheme based on analysis of a global ocean database, *Deep-Sea Research I, **57**, 6. doi: http://dx.doi.org/10.1016/j.dsr.2010.03.011*

Reynolds, R.W. et al 2007: Daily high-resolution blended analyses for sea surface temperature, J. Climate, 20, 5473-5496

Sakov, P. 2015: EnKF-C user guide, arXiv:1410.1233

DEVELOPMENT OF A 4DENVAR-BASED ENSEMBLE AT THE MET OFFICE, AND EXPERIMENTS WITH THE NEW ENSEMBLE COVARIANCES IN HYBRID DA

Neill Bowler[1], Adam Clayton[1], Mohamed Jardak[1], Peter Jermey[1], Eunjoo Lee[2], Andrew Lorenc[1], Chiara Piccolo[1], Stephen Pring[1], Marek Wlasak[1], Dale Barker[1], Gordon Inverarity[1] and Richard Swinbank[1]

[1]*Met Office, Exeter, UK*
[2]*Korea Meteorological Administration, Seoul, Republic of Korea*

adam.clayton@metoffice.gov.uk

In July 2011, the Met Office upgraded its operational global 4DVar data assimilation scheme to use hybrid climatological/ensemble covariances, with ensemble data provided by “MOGREPS-G” - the Met Office’s ETKF-based global ensemble prediction system (EPS).

Concerns about the future scalability and maintainability of the linear “Perturbation Forecast” (PF) model and its adjoint used in 4DVar have since led us to develop a hybrid-4DEnVar system, which uses 4D rather than 3D ensemble covariances, eliminating the need for the PF model. Similar systems are now operational at ECCC and NCEP, but at the Met Office performance relative to the existing hybrid-4DVar system is around 2% worse, so there are currently no plans to implement it operationally. However, the hybrid-4DEnVar code has now also been generalised to perform ensemble updates - a so-called “En-4DEnVar” scheme. A new EPS based on this scheme is now giving promising results, particularly as an improved source of ensemble error covariances for hybrid-4DVar. In the first part of this talk, I will summarise the design of the new ensemble, and present results from low-resolution trials.

Another strand of recent work - led by Andrew Lorenc - has focussed on improving the way ensemble forecasts are processed to specify the ensemble covariances used by hybrid DA. Significant improvements have been found from

1. Use of time-lagged and time-shifted ensemble perturbations.
2. Splitting the ensemble perturbations into spectral wavebands, and using band-specific horizontal localisation scales.

In the second part of the talk, I will describe these two processing methods, and show their impact in low-resolution hybrid DA trials.

RECENT EXPERIENCES WITH OPERATIONAL INITIALIZATION, PREDICTION AND VERIFICATION OF TROPICAL CYCLONES

Noel Davidson, Yi Xiao, Zifeng Yu[#], Hui Yu[#], Ying Jun Chen, Difei Deng, Xun Li, Beth Ebert, Yimin Ma, Xudong Sun and Numerous Colleagues

Research and Development Branch, Bureau of Meteorology, Melbourne

[#]Shanghai Typhoon Institute, China Meteorological Administration

n.davidson@bom.gov.au

In a recent discussion paper on Tropical Cyclone (TC) prediction, some of the key issues identified were;

1. To represent the inner-core structure, where the extreme winds and heavy rain often occur, higher resolution is required for both the data assimilation and the model. This is necessary if we are to initialize and forecast changes in the radius of maximum winds and TC size, as well as the distribution of rainfall, particularly for strong storms where the radius of maximum winds can be a few tens of kilometres.
2. Detailed verification of TC structure and rainfall, as well as track and intensity (central pressure, maximum wind), should be continued, broadened and enhanced.
3. Analysis of forecast busts would provide background on systematic deficiencies in ACCESS-TC.
4. Ensemble initialization and prediction, particularly for structure and intensity, need to be addressed to provide "guidance on the guidance", and will be highly desirable or even essential for data assimilation in the future.
5. The sensitivity of prediction of track, intensity, structure and rainfall to initial vortex structure is required to improve vortex initialization.
6. With ever-improving data coverage and assimilation techniques, the need for revised vortex specification should be evaluated at regular intervals.
7. Processes that influence TC structure change and rainfall should continue to be studied with a view to improving understanding, initialization and prediction of landfall.
8. Systematic testing and evaluation of ACCESS-TC on tropical depressions and TC genesis should receive high priority. These are very important for operational forecasting and for understanding of processes.

We will discuss progress on some of these projects at the workshop.

(a) ACCESS-TC has run operationally on named storms over the western Pacific and eastern Indian Oceans in both hemispheres since 2011. The base system runs at a resolution of 0.11º and 70 levels. The domain is re-locatable and nested in coarser-resolution ACCESS forecasts. Initialization consists of 5 cycles of 4D-VAR assimilation over 24 hours prior to the initial time, and forecasts to 72 hours are made. Without vortex specification, initial conditions usually contain a weak and misplaced circulation. Based on estimates of central pressure and storm size, vortex specification is used to filter the analysed circulation from the original analysis, construct the inner-core of the storm, re-locate it to the observed position and merge it with the large scale analysis at outer radii. From this idealized structure, synthetic MSLP observations are extracted and given to the 4D-VAR, with the objective of defining the intensity (central pressure, maximum wind), and structure (Rmax and R34).

(b) Verification indicates a competitive level of performance for both track and intensity, with mean 24-hour errors of approximately 100 km and 15 hPa. Major forecast failures, defined by errors exceeding approximately three standard deviations from the mean, occur in about 4% of forecasts. These failures damage verification statistics and have been studied in further detail. Analysis of forecast failures indicates a systematic issue in some situations with separation of low-level and mid-level circulations. We will discuss these forecast busts and other initialization aspects of vortex structure.

(c) In collaboration with the Shanghai Typhoon Institute, we have commenced systematic verification of forecast rainfall from ACCESS-TC for landfalling TCs, using the Contiguous Rain Area, CRA method. The aim is to provide benchmark verification statistics for these important, occasionally heavy rain events. Results are encouraging, but errors associated with (a) excessive rainfall over inner-radii (compared to TRMM estimates), and (b) displacement and pattern errors of rain features, are evident. We will discuss initialization aspects for TC rainfall prediction.

(d) Detailed analysis of case studies of: (a) extreme rain during the landfall of TC Bilis, and (b) rapid intensification of TC Rammasun from an ensemble of high-resolution WRF forecasts, has suggested critical processes during these events. Time-permitting, we will discuss our view on the critical importance of vortex structure and interaction with the environment for initialization and prediction of these high impact weather events. ***But how well can we initialize these critical circulation features?*** It seems clear that ensemble TC initialization and prediction is needed for such events.

THE IMPACT OF BACKGROUND FIELD ON THE TC BOGUS DATA ASSIMILATION

Xingbao Wang, Jeff Kepert, Noel Davidson, Yi Xiao and Jim Fraser

Research and Development Division, Bureau of Meteorology, Melbourne

Xingbao.Wang@bom.gov.au

Precise prediction of tropical cyclone (TC) track, structure, and intensity requires realistic representation of the vortex in the initial conditions except accurate large scale background field. However, most of the TCs originate over tropical oceans where meteorological observations are generally insufficient to accurately define vortex structure, particularly the inner core structure. Using the position and intensity information of the initial vortex, which can be well determined from satellite imagery, the TC structure can be indirectly recovered through the technique of 4DVAR TC bogus data assimilation (Davidson et al 2014). Our experience with a 4-km grid spacing mesoscale ACCESS for TC forecasting (ACCESS-TCX) during the past two years indicates that the effect of the TC bogus data assimilation (DA) is dependent on many factors, e.g., the quality of large scale background (first-guess) field, the bogus data in single central pressure and its position or more complicated vortex specification forms. In this presentation the impact of background field on the effect of bogus data assimilation in ACCESS-TCX is investigated:

1) In the genesis stage, the TC is not well-organized and the TC position and structure in the background field are not well resolved or even misplaced. In addition, the central sea surface pressure and position deduced from satellite imaginary are not accurate. The TC bogus DA may be of large error.
2) In the developing and mature stage, the TC structure is well-defined. The TC position and intensity can be accurately deduced from satellite imagery. The TC center in the first-guess field with respect to the observed TC center is crucial for TC bogus DA. A discrepancy distance could be defined as the distance from the TC center in first-guess field to the TC center in the observation. When the discrepancy distance is less than a few times of radius of maximum wind (RMW), the bogus DA can effectively influence the TC forecasting through the incremental analysis fields. Particularly, the bogus DA improves the TC intensity significantly. When the discrepancy distance is longer than several RMW, the bogus DA can modestly influence the TC structure and intensity. The bogus TC data with some discrepancy distance intensify the TC through axisymmetriztion of the asymmetric disturbances generated by the TC bogus data. The effect of the TC bogus DA is also dependent on the TC intensity and size. For the same discrepancy distance the impact of the same TC bogus may be larger in a strong and large TC than in a weak and small TC.
3) ACCESS-TCX can be run cyclically with just one cold-start. However, our experiments indicate that cyclic warm running of ACCESS-TCX may not be the best way to get the best TC forecasting skill in terms of track and intensity. In some TC cases the discrepancy distance does not decrease significantly in the cyclic run. Sometimes the skill of TC forecasting is better with new cold start though there is certain sensitivity of TCX to cold-started. These results indicate that there are still some biases in the ACCESS-TCX's DA system. These results may give us some clues and direction for the improvement and development of ACCESS-TCX forecast systems in the future.

References

Davidson, N. E., and Co-authors, 2014: ACCESS-TC: Vortex specification, 4DVAR initialization, verification, and structure diagnostics. *Mon. Wea. Rev.*, **142**, 1265–1289, doi:10.1175/MWR-D-13-00062.1.

RECENT RESEARCH ON IMPROVING THE USE OF ENSEMBLES IN ENVAR FOR DETERMINISTIC WEATHER PREDICTION

Mark Buehner[1], Jean-Francois Caron[1] and Ping Du[2]

[1]*Data Assimilation and Satellite Meteorology Research Section, Environment and Climate Change Canada, Dorval, Canada*
[2]*Data Assimilation and Quality Control Development Section, Environment and Climate Change Canada, Dorval, Canada*

mark.buehner@canada.ca

Introduction

Two projects related to the use of ensemble-variational data assimilation (EnVar) to deterministic weather prediction are described. The first is scale-dependent localization and the second is an approach for computing the forecast sensitivity observation impact (FSOI) that is appropriate for EnVar.

Scale-dependent covariance localization

A new approach is presented and evaluated for efficiently applying scale-dependent spatial localization to ensemble background-error covariances within an EnVar system. The approach is primarily motivated by the requirements of future data assimilation systems for global numerical weather prediction that will be capable of resolving the convective scale. Such systems must estimate the global and synoptic scales at least as well as current global systems while also effectively making use of information from frequent and spatially dense observation networks to constrain convective-scale features. Scale-dependent covariance localization (described in detail by Buehner and Shlyaeva, 2015) allows a wider range of scales to be efficiently estimated while simultaneously assimilating all available observations. Unlike other methods, scale-dependent localization also maintains, to the extent possible, the between-scale covariances. These cross-covariances are related to the heterogeneity of the covariances such that their elimination is equivalent to applying a spatial averaging to the covariances.

Single observation experiments (Figure 1) show how scale-dependent localization has an effect similar to other adaptive localization approaches in that the apparent amount of localization varies spatially. This spatial variation is simply a result of spatial variations in the relative distribution of error as a function of spatial scale. Consequently, in areas with mostly large scale errors (upper panels of Figure 1), less localization is applied than in areas with mostly small scale errors (lower panels of Figure 1). Preliminary full data assimilation experiments show a small overall improvement to forecast scores when using scale-dependent localization as compared with using a single fixed localization function for all scales.

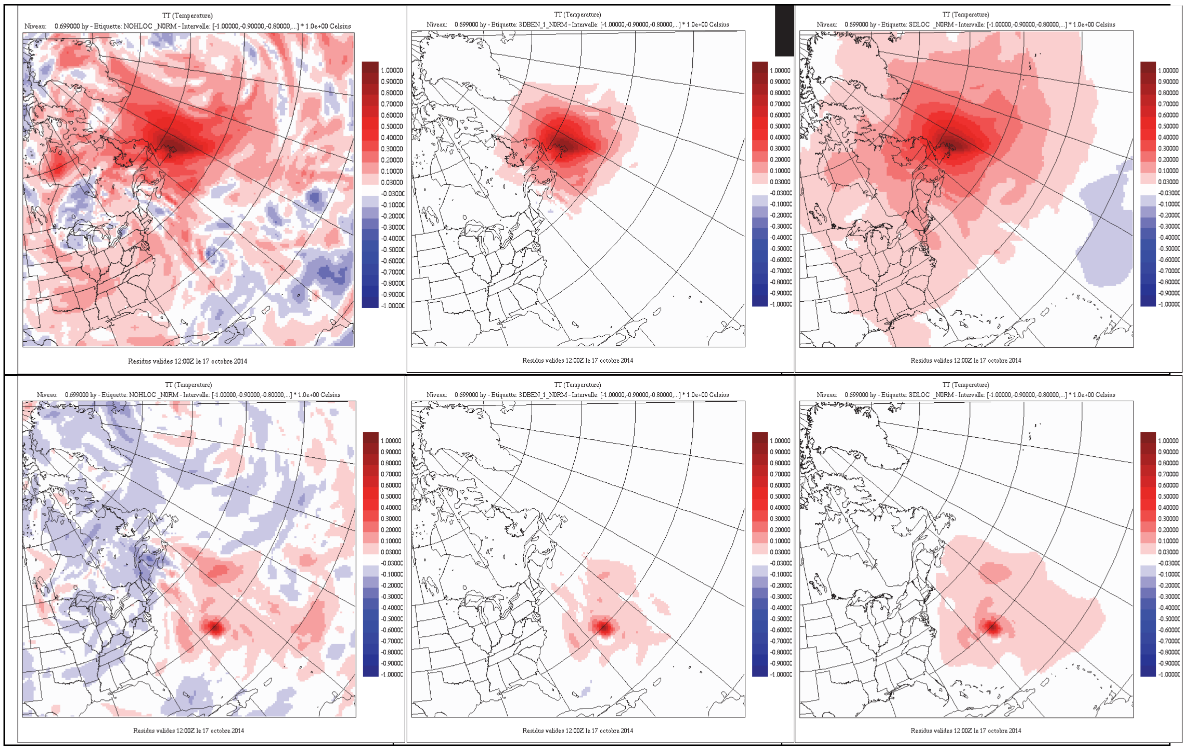


Figure 1:Analysis increments of temperature after assimilating a single temperature observation at 700 hPa either at the center of high pressure system (upper panels) or near the center of hurricane Gonzalo (October 2014). The result is shown when applying no localization (left panels), the standard fixed amount of localization (center panels) or when using scale-dependent localization (right panels).

Ensemble-variational approach to FSOI

A new approach to FSOI that is appropriate for EnVar data assimilation is currently being tested. This approach is similar to standard adjoint approaches to FSOI in that the adjoint of the data assimilation procedure is implemented through the iterative minimization of a cost function. However, like purely ensemble approaches to FSOI, the adjoint of the forecast is obtained by using an ensemble of forecasts instead of the actual adjoint version of the forecast model code. Therefore, the new approach shares properties of both the standard adjoint and purely ensemble approaches to FSOI.

Preliminary results have been obtained and compared with using the adjoint of the forecast model code instead of the ensemble of forecasts. Since the ensemble approach requires the application of spatial localization, this limits the spatial propagation of information in the FSOI calculation between analysis and forecast times. To ameliorate this, horizontal advection of the localization function is used. However, comparisons with using the adjoint model show that vertical propagation is also restricted. Possible approaches for allowing greater vertical propagation of the information will be explored.

References

Buehner, M. and Shlyaeva, A. 2015: Scale-dependent background-error covariance localization, *Tellus A*, **67**, 28027, doi: http://dx.doi.org/10.3402/tellusa.v67.28027

AN ENSEMBLE KALMAN FILTER FOR NUMERICAL WEATHER PREDICTION BASED ON VARIATIONAL DATA ASSIMILATION: VARENKF

Mark Buehner[1], Ron McTaggart-Cowan[2] and Sylvain Heilliette[1]

[1]*Data Assimilation and Satellite Meteorology Research Section, Environment and Climate Change Canada, Dorval, Canada*
[2]*Numerical Weather Prediction Research Section, Environment and Climate Change Canada, Dorval, Canada*

mark.buehner@canada.ca

Introduction

Several NWP centers currently employ a variational data assimilation approach for their deterministic forecasts and a separate EnKF system that is used for both initializing the ensemble forecasts and for providing ensemble covariances for the deterministic system. The data assimilation procedures used within current EnKF systems differ fundamentally from the variational approach. Several practical benefits may be realized by using the same data assimilation approach for both deterministic and ensemble prediction, including a reduction in the effort required to develop and maintain the computer code and an improved consistency of the impacts from major changes made to the two systems. To that end, the goal of this presentation is to evaluate a new approach for performing the data assimilation step within a perturbed-observation EnKF by adapting the variational algorithm used for deterministic data assimilation. The analysis increment is computed with a variational assimilation approach separately for the ensemble mean and for all of the ensemble perturbations (i.e. the deviations of each member from the ensemble mean). To obtain a computationally efficient algorithm, a much simpler configuration is used for the ensemble perturbations, whereas the configuration used for the ensemble mean is similar to that used for the deterministic system. Since the new approach is essentially an EnKF implemented with a variational approach, it is called VarEnKF.

Comparison between EnKF and VarEnKF

Several experiments were conducted (more details in Buehner et al. 2016) to evaluate the impact of first using a variational approach for computing only the ensemble mean analysis increment. This is equivalent to recentering the EnKF analysis ensemble on the variational analysis. When using the same set of observations as the EnKF itself, using EnVar for the ensemble mean produced some improvement in the forecast quality. Further improvement was seen when assimilating the full set of observations as used in the deterministic prediction system. The observation types assimilated in the deterministic system, but not currently in the EnKF include all hyper-spectral infrared radiances, geostationary satellite radiances, SSMIS microwave radiances and ground-based GPS. An additional experiment that uses a simplified variational approach for computing the ensemble perturbation analysis increments was also conducted. The simplifications include using only static background-error covariances at a lower spectral truncation, assimilating only the same observation types as in the EnKF and using a diagonal observation-error covariance matrix. These simplifications were made only for the ensemble perturbations and not the ensemble mean. The impact on the resulting forecast quality was much smaller than the improvements from using the full 4DEnVar configuration for the ensemble mean. Figure 1 shows the evaluation of the fully variational approach, called VarEnKF, versus the standard configuration of the EnKF.

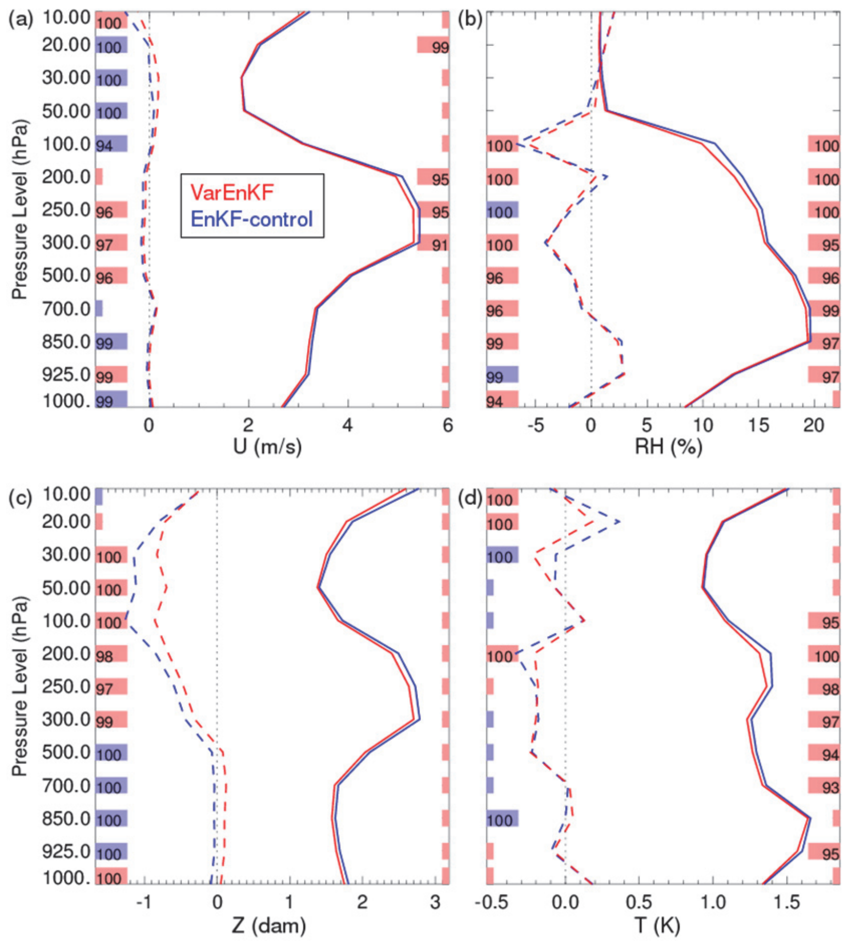


Figure 1: The error in the 72h global control member forecasts relative to ERA-interim analyses from the VarEnKF (red) and EnKF-control (blue) experiments computed over the period 0000 UTC 3 January 2015 to 1200 UTC 28 January 2015. Both the standard deviation (solid curves) and average (dashed curves) of the error are shown for (a) zonal wind, (b) relative humidity, (c) geopotential height, and (d) temperature. Boxes containing numbers on the right and left side of each panel indicate the level of statistical significance that, respectively, the error standard deviation and the mean error are different for the two experiments.

References

Buehner, M., McTaggart-Cowan, R. and Heilliette, S. 2016: An Ensemble Kalman Filter for Numerical Weather Prediction based on Variational Data Assimilation: VarEnKF, *Monthly Weather Review*, Early Online Release, doi: http://dx.doi.org/10.1175/MWR-D-16-0106.1

DATA ASSIMILATION BACKGROUND COVARIANCE AND GAIN MATRIX ANALYSIS

Xudong Sun

R&D Division, Bureau of Meteorology, Melbourne

x.sun@bom.gov.au

Background covariance and its calculation plays an important role in global and regional variational and ensemble data assimilation. Covariance statistical calculation can come from a variety of NWP models and ensemble sources with different spatial resolutions and domains. Furthermore, different season data could be used with different forecast error characteristics. As for ensemble data assimilation, the calculation of such statistics in real-time provides flow-dependent information.

Therefore, understanding the ensemble statistics for different data assimilation systems in relation to the covariance is highly desirable.

In this research, the majorization theory is used to compare different covariance matrices. Subsequently, the Kalman gain and variational gain matrix is also evaluated based on the different background covariance used.

a. Vector and Matrix Majorization

The concept of majorization describes how two vectors or two matrices are related to each other according to randomness, smoothness etc.. Let $x=(x_1,x_2,...,x_n)$ and $y=(y_1,y_2,...,y_n)$ be real numbers in decreasing order. Then, x majorizes y indicates that for each k=1, 2, ..., n, the following is true (i. e: Marshall et al., 2011) :

$$\sum_{i=1}^{k} x_i \geq \sum_{i=1}^{k} y_i ,$$

Majorization is under the framework of total sets of sum of x and y is equal when k = n. If this condition is not fulfilled, it is called weak majorization.

The description of matrix majorization is made using its eigenvalues.

b. Covariance and correlation majorization studies

Considering two different background covariance matrices, A and B, both of them are positive semi-definite. If A ≥B, it is said that A-B is positive semi-definite. If this is the case for two covariance (correlation) matrices, much of the matrix theory and inequality can be readily applied. However, A-B may not be positive semi-definite. For instance, this is caused by the matrix diagonal of A-B being smaller than its off diagonal elements. The use of matrix majorization provides an alternative way to compare A and B properties. A is defined to be better "correlated" than B if eigenvalue of A majorizes eigenvalue of B for the same matrix dimensions (Serpedin et al., 2012). Several theorems and corollaries are applied and derived specifically in our background covariance analysis. Matrix majorization on two important data assimilation issues: covariance localization and inflation is discussed.

c. Matrix inequality for weighting and Kalman gain

Further studies on two sets of weighting gain matrix: $(A+O)^{-1}A$ and $(B+O)^{-1}B$, where A and B is defined in Section b and O is observational error covariance (correlation) matrix. This enables us identify the mechanism of weight matrix impact on 3D variational and ensemble data assimilation (EnKF). If $A>0$, $B>0$ and $A \geq B$, then first gain matrix is greater than second gain matrix (Wang et al. 1999), and its trace of first gain matrix is greater than second gain matrix. Similar to Section b, when $A \geq B$ is not satisfied, majorization properties of A and B can be used to evaluate gain matrix properties, such as its maximum and minimum eigenvalues. Notice that the inverse of A+O caused the majorization under weak matrix majorization conditions. Finally, the role of observational error covariance matrix to the gain matrix can also be evaluated.

d. Conclusion

The procedure for comparisons of different data assimilation background covariances and their gain matrices is presented. It is expected that by applying it on actual covariance statistics obtained at different scales in ACCESS data assimilation, their diagnostics can be found and used to evaluate the current global and regional Hybrid-4dVAR forecast scheme.

References

Marshall A.W., Olkin, I. and Arnold, B. 2011. Inequalities: Theory of Majorization and Its Applications, Second edition. Springer Series in Statistics. Springer, New York. 909 p.

Wang B., Xi, B. and Zhang, F. 1999, Some inequalities for sum and product of positive semidefinite matrices. Linear Algebra and its Applications. Vol. 293, Issues 1–3, 39-49 p.

Serpedin, E., Chen, T.M. and Rajan D. 2012. Mathematical foundations for signal processing, communications, and networking. CRC Press, 797 p.

SOME THOUGHTS ON HYBRID APPROACH TO DATA ASSIMILATION

Monika Krysta

Environment and Research Division, Bureau of Meteorology, Melbourne

monika.krysta@bom.gov.au

Variational data assimilation strongly relies on an appropriate formulation of the background error covariance matrix. The background error covariance matrix is responsible for spatial spreading of the information provided by observations as well as for filtering of this information.

Initially, variational data assimilation methods employed a static background error covariance matrix for its cycled simulations. In parallel, a class of data assimilation methods issued from Kalman filtering developed a broad experience in defining and exploiting time dependent error covariance matrices. In recent years, variational data assimilation has been increasingly using this long standing experience of the sequential methods. In fact, a new class of methods, referred to as hybrid (e.g. Clayton et al, 2013), in which the static climatological background error covariance matrix is supplemented with an ensemble-based flow-dependent component, emerged.

In this presentation, we propose to focus on the characteristics of the background error covariance matrix, which are crucial to the performance of the hybrid ensemble/4D-Var assimilation scheme. We will look at structural and filtering properties of the background error covariance matrix and investigate how these properties change with the introduction of the time-evolving component (e.g. Brousseau et al, 2012). We will also discuss their impact on the spatial structure of the analysis increments in hybrid data assimilation experiments.

References

Clayton, A.M., Lorenc, A.C. and Barker, D.M. 2013: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office, Q. J. R. Meteorol. Soc., 139, 1445-1461

Brousseau, P., Berre, L., Bouttier, F. and Desroziers, G. 2012: Flow-dependent background-error covariances for a convective-scale data assimilation system, Q. J. R. Meteorol. Soc., 138, 310-322

APPROACHES TO CONVECTIVE SCALE DATA ASSIMILATION

Tijana Janjic[1], Heiner Lange[2], Yvonne Ruckstuhl[2], Yuefei Zeng[2]

[1] *Hans-Ertel-Centre for Weather Research, Deutscher Wetterdienst, Munich, Germany*
[2] *Hans-Ertel-Centre for Weather Research, LMU, Munich, Germany*

tijana.janjic-pfander@dwd.de

The applications of data assimilation on convective scales require a numerical model of the atmosphere with single digit horizontal resolution in km and time evolving error covariances. Past studies have shown that ensemble Kalman filter (EnKF) algorithm can be applied to the convective scales since it is capable of handling complex and highly nonlinear processes. However, some challenges for the convective scale applications still remain. These include a need to estimate on convective scale, fields that are nonnegative (such as rain, graupel, snow) and to use data sets such as radar reflectivity or cloud products that have the same property. What underlines these examples are errors that are non-Gaussian in nature causing a problem with the EnKF that uses Gaussian error assumptions to produce the estimates from the previous forecast and the incoming data. Since the proper estimates of hydrometeors are crucial for prediction on convective scales, the question arises whether the EnKF method can be modified to improve these estimates and whether there is a way of optimizing the use of high-resolution data such as radar observations or Mode-S wind data (with rapid updates) to initialize numerical weather prediction models.

In this talk, we first review the challenges that convective scale data assimilation methods are facing when assimilating radar data. This is done using the non-hydrostatic convection permitting COSMO model (Baldauf et al. 2011), and the local ensemble transform Kalman filter (LETKF, Hunt et al. 2007) as implemented in KENDA (Km-scale Ensemble Data Assimilation) system of German Weather Service (Schraff et al. 2016). Currently KENDA uses latent heat nudging for radar data assimilation (see Schraff et al. 2016), but the transition to localized EnKF for radar data has been tested as well (Bick et al. 2016). We will show that due to the Gaussian assumptions that underline the LETKF algorithm, the analyses of water species will become negative in some grid points of the COSMO model. These values are set to zero after the LETKF analysis step, in order not to give the numerical model unphysical values. The tests done within this setup show that such a procedure introduces a bias in the analysis ensemble with respect to the truth, that increases in time due to the cycled data assimilation. Further, the noise produced by assimilation of radar reflectivity with LETKF is larger than if only conventional data (including the high resolution Mode-S data) are assimilated. If the horizontal resolution of numerical model is changed from 2.8 km (COSMO-DE operational resolution) to 1.4 km, the noise even increases further and surpasses the noise levels of experiments that used latent heat nudging for assimilation. In all EnKF experiments performed, assimilation of the radial wind measurements in addition reduces the noise. In order to better understand some of the above challenges, the localized EnKF has been tested in an idealized framework.

We show the behavior of the localized EnKF with respect to preservation of positivity, conservation of mass, energy and enstrophy in toy models that conserve these properties. In order to preserve physical properties in the analysis as well as to deal with the non-Gaussianity in an EnKF framework, Janjic et al. 2014 proposed the use of physically based constraints in the analysis step to constrain the solution and therefore change the analysis error statistics. This approach led to the new algorithm based on the EnKF and the quadratic programming (QPEns: Quadratic Programming Ensemble filter). In Janjic et al. 2014 it was shown on a very simple example that for state estimation, the inclusion of the constrained estimation can improve the ensemble Kalman filter results in case of strongly non-Gaussian error distributions. Importantly, only methods, which preserve positivity and mass together,

produce accurate analysis. The QPEns algorithm was further tested on the modified shallow water model (Würsch and Craig, 2014), which was designed to mimic the most important characteristics of convective motion. Perfect model experiments were performed with observations that are taken every 5 minutes at the grid points where there is rain for all three variables of the modified shallow water model (wind, height of the fluid and rain). It was shown that the mass conservation- and positivity-constrained rain significantly suppresses noise seen in localized EnKF results. This is highly desirable in order to avoid spurious storms from appearing in the forecast starting from this initial condition (Lange and Craig 2014). In addition, the root mean square error (RMSE) is reduced for all fields and total mass of the rain is correctly simulated.

Further, using perfect model experiments with mass, total energy and momentum conserving 2D shallow water model that also conserves enstrophy for non-divergent flow, we illustrate the importance of assimilation of the wind data (Zeng and Janjic, 2016). The data assimilation uses LETKF with varying localization radius, thinning interval, observed variable and inflation. During assimilation, the total mass remains consistent with that of the nature run and the total energy of the analysis mean converges towards the nature run value. However, enstrophy, divergence, as well as energy spectra are strongly affected by localization radius, thinning interval, and inflation and depend on the variable that is observed. Only observations of wind are able to (at least) control unrealistic enstrophy increase in one-hour data assimilation updates. In this idealized setup, we tested the effects on prediction depending on the type of errors in the initial condition. By measuring the nonlinear energy cascade through a scalar, domain averaged, noise term, we show that the accumulated noise during assimilation and the analysis RMSE are good indicators of quality of the prediction.

Noise produced by assimilation of radar reflectivity data seems to be partially reduced by adding constraints of mass and positivity at least in the idealized setup. The noise can be even further reduced in the EnKF framework if additional velocity data are assimilated. This is shown for both COSMO-KENDA and idealized experiments. It remains to be seen whether this noise reduction can be related to better control of enstrophy or divergence.

References

Baldauf, M., A. Seifert, J. Förstner, D. Majewski, and M. Raschendorfer, 2011: Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. Mon. Wea. Rev., 139, 3887–3905, doi:10.1175/MWR-D-10-05013.1.

Bick, T., Simmer, C., Trömel, S., Wapler, K., Hendricks Franssen, H.-J., Stephan, K., Blahak, U., Schraff, C., Reich, H., Zeng, Y. and Potthast, R. (2016), Assimilation of 3D radar reflectivities with an ensemble Kalman filter on the convective scale. Q.J.R. Meteorol. Soc., 142: 1490–1504. doi:10.1002/qj.2751

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. Physica D, 230, 112–126, doi:10.1016/ j.physd.2006.11.008.

Janjic, T., D. McLaughlin, S. E. Cohn, and M. Verlaan, 2014: Conservation of mass and preservation of positivity with ensemble-type Kalman filter algorithms. Monthly Weather Review, 142 (2), 755–773.

Lange, H. and G.C. Craig, 2014: The Impact of Data Assimilation Length Scales on Analysis and Prediction of Convective Storms, Monthly Weather Review, 142, 3781–3808.

Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Periáñez, A. and Potthast, R. (2016), Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). Q.J.R. Meteorol. Soc., 142: 1453–1472. doi:10.1002/qj.2748

Würsch, M and Craig, G.C., 2014: A simple dynamical model of cumulus convection for data assimilation research. Meteorol. Z., 23, 483-490.

Zeng, Y. and Janjić, T. (2016), Study of conservation laws with the Local Ensemble Transform Kalman Filter. Q.J.R. Meteorol. Soc., 142: 2359–2372. doi:10.1002/qj.2829

DOPPLER RADAR WIND OBSERVATIONS FOR HIGH RESOLUTION DATA ASSIMILATION

Susan Rennie and Peter Steinle

Environment and Research Division, Bureau of Meteorology, Melbourne

susan.rennie@bom.gov.au

Australia has a growing network of Doppler radars within the Weather Watch radar network (Figure 1). These measure radial wind velocity at high temporal and spatial resolution. However, such observations are normally only available when there is e.g. precipitation to reflect the radar beam.

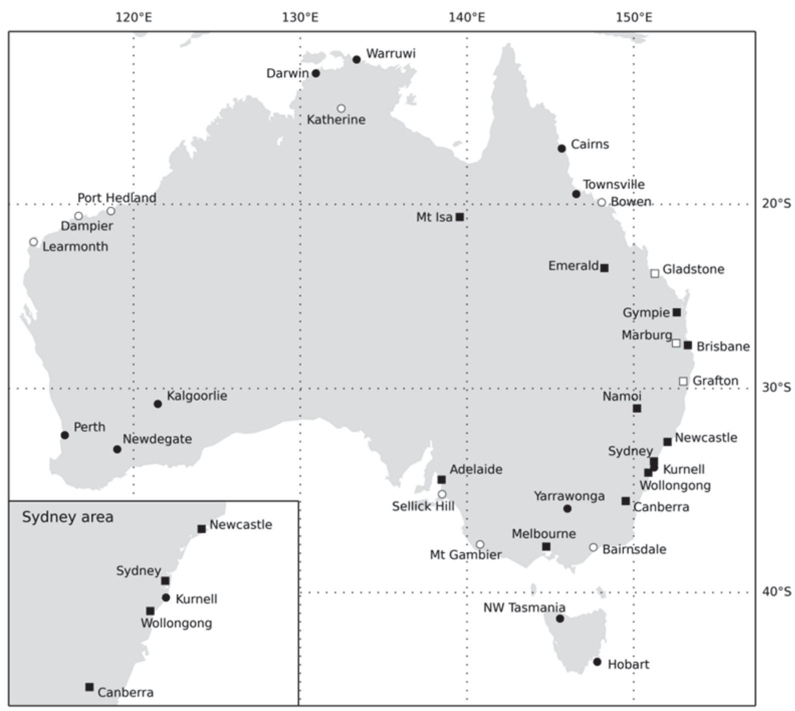


Figure 1. Map of Australian Weather Watch radars. Doppler radars are black markers, non-Doppler are white markers. C-band radars are circles and S-band radars are squares. Part time radars are not shown.

High resolution radar observations can be assimilated in the new 1.5 km ACCESS-City system, as at that resolution the model starts to resolve features visible with radar observations. The 'convection-permitting' City NWP systems are the first to assimilate radar velocity observations.

A major hurdle for the assimilation has been the quality control (QC). Radars echoes from the Earth's surface (ground and sea clutter) yield zero-velocity measurements rather than the velocity of the atmosphere above the surface. Additionally, there are radar echoes from atmospheric-borne stuff apart from precipitation, including insects and birds, smoke and chaff. A quality control system was implemented that uses thresholding techniques and Bayesian classification (Rennie et al. 2015) to extract the radar echoes that provide valid wind estimates. These include precipitation echoes and, experimentally, insect echoes. For single-polarisation radar QC methods, there remains a degree of

clutter contamination. Additional QC as part of the pre-assimilation observation processing helped reduce this, such that the assimilated observations appeared to have reasonable quality.

Radar data assimilation was trialled during the Forecast Demonstration Project, which ran ACCESS-Sydney for 10 weeks in spring 2014. During this period, radar wind observations from both precipitation and insect echoes were assimilated. Observation statistics were collected, which were used to assess the quality of the raw and assimilated radar observations. These showed that clutter still affects some radar scans, but after all stages of quality control the observation error was substantially reduced. Overall the quality of observations from precipitation and insects was comparable, with clutter likely to dominate other error sources, e.g. from insect flight. Figure 2 shows the mean whole-scan speed bias based on the method of Salonen et al. (2007), calculated by model minus observed speed. High positive values indicate clutter contamination contributing to a reduced net speed. Data for both precipitation and insects (clear air) are shown.

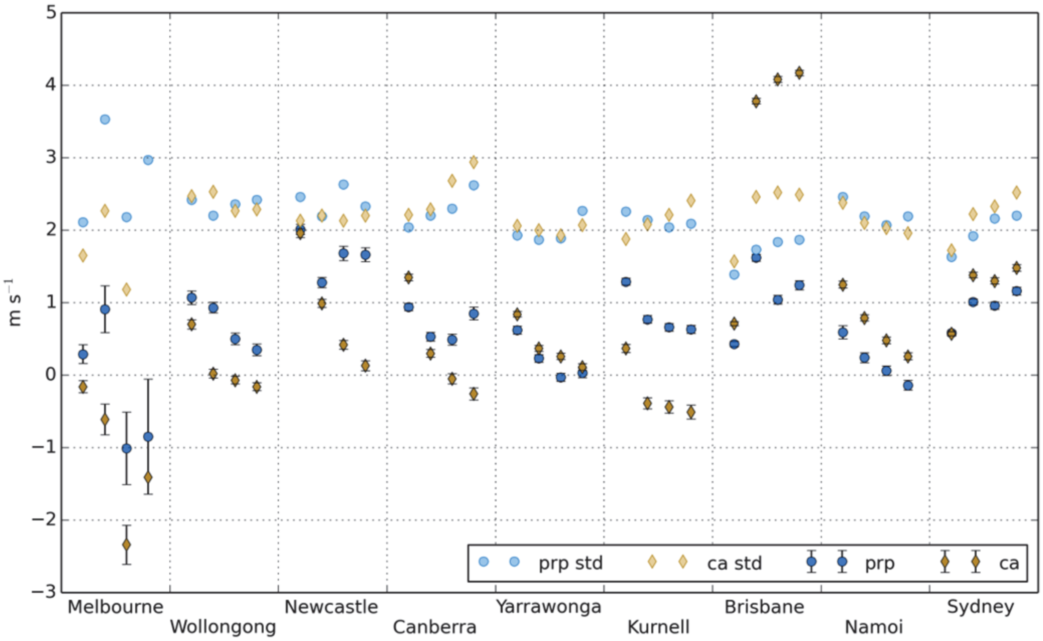


Figure 2. The speed bias mean and std. dev. for each radar and elevation, for precipitation (prp) and clear air (ca, i.e. insects), from the duration of the FDP. For each radar, the data points from left to right represent the 0.9°, 2.4°, 4.2° and 5.6° elevation scans. A large positive mean speed bias can result from cluttered observations.

It is intended that radar observations will be assimilated in future operational ACCESS-City versions. Development will involve evaluation of the impact of radar wind observations on the forecast. It is anticipated that high-resolution near-surface observations will benefit convective development forecasts. There are several aspects to be addressed to best utilise radar observations. Firstly, the radars only observe near-surface winds close to the radar. Since the lowest elevation scan has the most clutter, it was not used during the FDP. Improvements to clutter removal, such as by upgrading to dual-polarisation, will increase the quality and utility of low-elevation scans. Secondly, assimilation involves spreading observation information horizontally and vertically, based on the estimated covariances of observations and of the model. The covariances need to capture, not smear out, the high-resolution features observed by the radar. Conversely, assimilating erroneous observations may introduce false small-scale wind features.

Ongoing effort at the Bureau of Meteorology to improve the radar network, quality control methods, assimilation methods and covariance estimates will all contribute to the benefit to be gained from Doppler radar wind observations.

References

Rennie, S. J., M. Curtis, J. Peter, A. W. Seed, P. J. Steinle, and G. Wen, 2015: Bayesian echo classification for Australian single-polarization weather radar with application to assimilation of radial velocity observations. *J. Atmos. Ocean. Technol.*, **32**, 1341–1355, doi:10.1175/JTECH-D-14-00206.1.

Salonen, K., H. Järvinen, R. Eresmaa, and S. Niemelä, 2007: Bias estimation of Doppler-radar radial-wind observations. *Q. J. R. Meteorol. Soc.*, **133**, 1501–1507, doi:10.1002/qj.114.

CURRENT STATUS AND FUTURE PLANS FOR THE KMA DATA ASSIMILATION SYSTEM

Hyun-Cheol Shin, Eun Joo Lee, Eun Hee Lee, Jung-Rim Lee, Eun-Hee Kim, Hyeyoung Kim, Heejung Kang, Adam Clayton, Sangwon Joo

Korea Meteorological Administration, Seoul, Korea

sinhyo@korea.kr

The performance of KMA's NWP system has significantly improved since KMA introduced the Met Office Unified Model in 2010. Currently, KMA is operating a 17km global model with hybrid-4DVar, a 12km regional model with (non-hybrid) 4DVar, and a 1.5km local model with 3DVar. The 17km global model was launched in this June this year, replacing the previous 25 km model, and including an upgrade to the "ENDGame" dynamical core. The data assimilation system was also upgraded to include assimilation of GOES Clear, ATMS, CrIS, MVIRIClear and Ground GPS observations.

The 1.5km local model was also improved, by upgrading to "ENDGame" dynamics and extending the model domain (but with a lower resolution in the extension region). The previous 1.5km model had a domain covering the Korea peninsula and surrounding sea. The extended domain now covers Japan and the eastern part of China, bringing an increase in the number of observations used by the 3-hourly 3DVar data assimilation algorithm. Next year we plan to introduce assimilation of Ground-GNSS, Himawari-8 AMV & CSR, AMSU-B and IASI satellite observations, all of which have given a positive impact in research trials.

In the near future, KMA will implement a nowcasting model, again with 1.5km horizontal resolution, but using hourly rather than 3-hourly 3DVar, and for a smaller domain covering the Korean peninsula. We will shortly begin experiments with a 4DVar version which we hope to implement next year.

KMA is currently running a 12km regional model for East Asia, but this will become obsolete when the global model is upgraded to 10km in 2018. We are currently evaluating the possibility of replacing this with a 4km model for the same domain, or with a further-extended version of the 1.5km model.

DATA ASSIMILATION FOR TERRESTRIAL BIOGEOCHEMISTRY, WHY IS IT DIFFERENT?

Peter Rayner

School of Earth Sciences, University of Melbourne, Melbourne

prayner@unimelb.edu.au

We can classify data assimilation problems according to the information we seek, the growth rates of perturbations in the model state and our confidence in the underlying dynamics. Numerical Weather Prediction usually seeks knowledge of the initial condition, contains unstable modes and has dynamics, in large part, based on sound physical principles. The case of terrestrial data assimilation, especially of terrestrial biogeochemistry, is very different. In this talk I will walk briefly through the dominant dynamics of the terrestrial carbon cycle and explain their impact on data assimilation in this domain. We will see that, most of the time, perturbations are evanescent but that there is little theory and less agreement guiding the choice of parameterisations. This drives a different set of data assimilation approaches. We will demonstrate these with a very simple model.

INCORPORATING LAND SURFACE OBSERVATIONS INTO REANALYSES: NASA, GMAO'S MERRA-2 AND BEYOND

Clara Draper[1,2], Rolf Reichle[1], Qing Liu[1,3], Randy Koster[1], Sarith Mahanama[1,3], Manuela Girotto[1,2]

[1] *Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt , MD, USA.*
[2.] *Universities Space Research Association / GESTAR*
[3] *Science Systems and Applications Inc*

clara.draper@nasa.gov

This presentation provides an overview of progress at the NASA Global Modeling and Assimilation Office (GMAO) to incorporate observations relevant to the land surface into current and future atmospheric/Earth system reanalyses. First, the land component of the recently released Modern-Era Retrospective Analysis for Research and Applications 2 (MERRA-2) will be briefly reviewed. While the land surface states in MERRA-2 are not directly updated through data assimilation, errors in the land surface states due to errors in the AGCM-generated precipitation are corrected through direct insertion of observed precipitation at the land surface. The methodology for using the observed precipitation was refined from that used in the (land-only) MERRA-Land reanalysis. In both MERRA-Land and MERRA-2, the use of observed precipitation clearly improves the land surface hydrology, compared to the previous MERRA system. Additionally, by correcting the precipitation within the coupled land-atmosphere system, MERRA-2 has an advantage over MERRA-Land in that the near-surface air temperature and humidity can respond to the improved precipitation. MERRA-2 thus provides more self-consistent surface meteorological data than were available from MERRA-Land. Second, the GMAO is moving towards an integrated Earth (atmosphere-ocean-land-ice) system analysis, which will include a weakly-coupled land data assimilation component. The land data assimilation will be performed every assimilation update cycle, using the Ensemble Kalman Filter-based GMAO Land Data Assimilation System (LDAS). The first priorities are to assimilate remotely sensed snow cover and near-surface soil moisture, and in situ snow depth observations, as well as to implement recent land modeling advances. Finally, early work towards integrating the GMAO LDAS into the GEOS-5 hybrid 3D-Var system demonstrates the potential benefits of a more tightly coupled land and atmospheric assimilation.

A NEW HIGH RESOLUTION LAND DRYNESS ANALYSIS SYSTEM FOR AUSTRALIA

Imtiaz Dharssi and Vinodkumar

Bureau of Meteorology, Melbourne

i.dharssi@bom.gov.au

Accurate land surface information is critical for many applications such as weather and seasonal forecasting, fire behaviour modelling and fire danger prediction, flood forecasting, prediction of droughts and heat waves, soil water forecasting for agriculture, evaporation forecasts for water companies, fog forecasting for aviation and climate prediction. For example, recent work improving the initialisation of soil moisture in the ACCESS seasonal forecasting system has resulted in significantly better forecasts. However, providing accurate land surface information is very challenging since the land surface has a very high spatial variability and there are very few ground based observations.

Currently, landscape dryness for fire danger prediction is estimated using very simple water balance models developed in the 1960s that ignore many important factors such as incident solar radiation, soil types and vegetation properties. The most prominent of those used in Australia are the Keetch-Byram Drought Index (KBDI) developed by the US Forest Service, and the related Soil Dryness Index (MSDI) developed by Forestry Tasmania. Previous work by Dharssi and Vinodkumar (2015) show that these old dryness indices are less accurate than soil moisture analyses from the ACCESS numerical weather prediction system.

This work presents a prototype high resolution soil moisture analysis system that can include data from many sources; such as surface observations of rainfall, temperature, dew-point temperature, wind speed, surface pressure as well as satellite derived measurements of rainfall, surface soil moisture, downward surface shortwave radiation, skin temperature, leaf area index and tree heights. The analysis system estimates soil moisture on four soil layers over the top 3 meters of soil, the surface layer has a thickness of 10 cm. The system takes into account the effect of different vegetation types, root depth, stomatal resistance and spatially varying soil texture. The analysis system has a one hour time-step with daily updating. Verification against ground based soil moisture observations from the OzNet, CosmOz and OzFlux networks shows that the new system is significantly more accurate than other soil moisture analysis systems.

References

Dharssi, I. and Vinodkumar 2015: Inter-Comparison of Land Surface Model Soil Moisture Data with Traditional Soil Dryness Indices. 21st International Congress on Modelling and Simulation, Gold Coast, Australia. http://www.mssanz.org.au/modsim2015/A4/dharssi.pdf

USING REMOTE SENSING DATA FOR HYDROLOGICAL AND HYDRAULIC FLOOD FORECASTING

Yuan Li, Stefania Grimaldi, Valentijn Pauwels, Jeffrey Walker, Ashley Wright

Department of Civil Engineering, Monash Universit, Melbourne

Yuan.Li2@monash.edu

Flooding is one of the most destructive and frequent natural disasters in Australia. It threatens people's lives (e.g., 1859 recorded deaths from 1900 to 2015 in Australia) and possessions (e.g., approximately $377 million per year during the last 40 years in Australia). An accurate and reliable flood forecast with sufficient lead-time can provide vital information for flood preparedness, warning delivery, and emergency response. An operational flood forecasting system usually consists of a hydrologic model, which simulates runoff generation and concentration, and a hydraulic model, which models riverine flood routing and floodplain inundation. Nevertheless, these two types of models are subject to uncertainties from meteorological forcing, catchment initial conditions, model physics and parameters.

Operational flood forecasting systems are conventionally calibrated and/or updated using streamflow measurements, and such applications are limited in well-gauged areas. The recent development of remote sensing (RS) techniques provide new approaches to monitor the water cycle from space, which thereby offers new opportunities for improved flood investigation and forecasting. Based on an Australian case study, this presentation will discuss the use of 1) RS soil moisture data to constrain a hydrologic model, and 2) RS-derived flood extent and level to constrain a hydraulic model.

The hydrological model is based on a semi-distributed system coupled with a two-soil-layer rainfall-runoff model GRKAL and a linear Muskingum routing model. Model calibration was performed using either 1) streamflow data only or 2) both streamflow and RS soil moisture data. The model was then further constrained through the integration of real-time soil moisture data.

The hydraulic model is based on LISFLOOD-FP which solves the 2D inertial approximation of the shallow water equations. Streamflow data and RS-derived flood extent and levels were used to calibrate and validate the model. The effectiveness of each data source or their combinations on flood propagation and inundation forecasting will be investigated and discussed.

LAND SURFACE DATA ASSIMILATION AT THE MET OFFICE

Breogan Gomez, Brett Candy, Samantha Pullen, Keir Bovis, Richard Renshaw

Met Office, Exeter, UK

richard.renshaw@metoffice.gov.uk

Operational NWP at the Met Office uses data assimilation to initialise land surface conditions (snow cover, soil moisture and temperature) for global forecast modelling. This workshop presentation will outline the methods and data used, and will describe recent and planned developments and results.

Snow

A daily Northern Hemisphere snow analysis is produced using the NESDIS IMS snow cover product to update model snow amount in a simple update scheme. Snow cover in the southern hemisphere, and in regional models, is free-running in the model. In particular, there is no snow analysis for the convection-permitting 1.5km UK-area model. Work is underway on a new snow analysis scheme that will make use of high-resolution data over the UK to update model snow fields.
The new scheme will use OI (optimal interpolation) to analyse both snow cover and snow depth. It will make use of ground-based obs of snow depth, and state of ground (snow or no snow) from the synoptic observation network. These will be combined with satellite-derived snow cover from the H-SAF (MSG-SEVIRI) daily product.

Soil moisture

An extended Kalman filter (EKF) scheme is used every six hours to update soil moisture in the global NWP model. The EKF uses both the ASCAT soil wetness product (from satellites Metop-A and Metop-B) and screen-level observations of temperature and humidity from the synoptic observation network. The screen-level data is processed through a 3Dvar analysis which gives gridded fields from observation minus model background differences. These differences can be related to soil moisture and temperature through multiple runs of the JULES land surface model which provide estimates of the sensitivity of screen-level variables to the soil fields.
Regional NWP models then take their values from the global model, reconfigured to the regional model grid. A recent development has been to adapt the EKF scheme such that it can be used directly for regional models. Early results are shown using this scheme in the 1.5km UK model.

Land surface temperature

It is known that the UM global NWP model on occasion shows bias in land surface temperature, for instance over some desert surfaces during the day. The availability of several sources of good quality land surface temperature (LST) measurements motivates to try to evaluate the benefits of assimilation these observations. In particular, the Met Office has been involved in assessing the ESA GlobTemperature product of land surface temperature. The EKF is sufficiently flexible to allow use of a range of observation types. Results are presented showing the impact of assimilating LST from GlobTemperature for two trial periods. Overall impact on NWP forecast skill is small but in some areas a positive impact on screen-level temperature can be seen.

MULTIVARIATE ASSIMILATION OF LAND SURFACE REMOTE SENSING DATASETS: ADVANCES, GAPS AND CHALLENGES

Sujay V. Kumar[1], Michael Jasinski[1] , David Mocko[1] , Jordan Borak[1], Matthew Rodell[1] and Christa Peters-Lidard[1]

[1]Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Maryland, USA

Sujay.V.Kumar@nasa.gov

Remote sensing advancements in recent years have enabled monitoring of the Earth's land surface with unprecedented scale and frequency. In the past two decades, remote sensing observations of the land surface have become available from a number of satellite instruments and platforms including soil moisture (AMSR-E, ASCAT, AMSR2, SMOS, SMAP), snow depth (AMSR-E, AMSR2), snow cover (MODIS, VIIRS) and terrestrial water storage (GRACE), among others. The need for a comprehensive modeling infrastructure that can effectively synthesize these observations has motivated the development of the NASA Land Information System (LIS; lis.gsfc.nasa.gov; Kumar et al. 2006). LIS provides the capabilities to integrate terrestrial hydrologic observations using advanced land surface modeling techniques to produce improved estimates of land surface conditions such as soil moisture, evaporation, snow pack, and runoff at multiple spatial and temporal resolutions. Using LIS, an integrated terrestrial water analysis system has been created by assimilating various land surface data products over the US, during the satellite era (1979-present), using the model configurations used in the North American Land Data Assimilation System (NLDAS), using two different land surface models; Noah (version 3.3) and Catchment (version Fortuna 2.5).

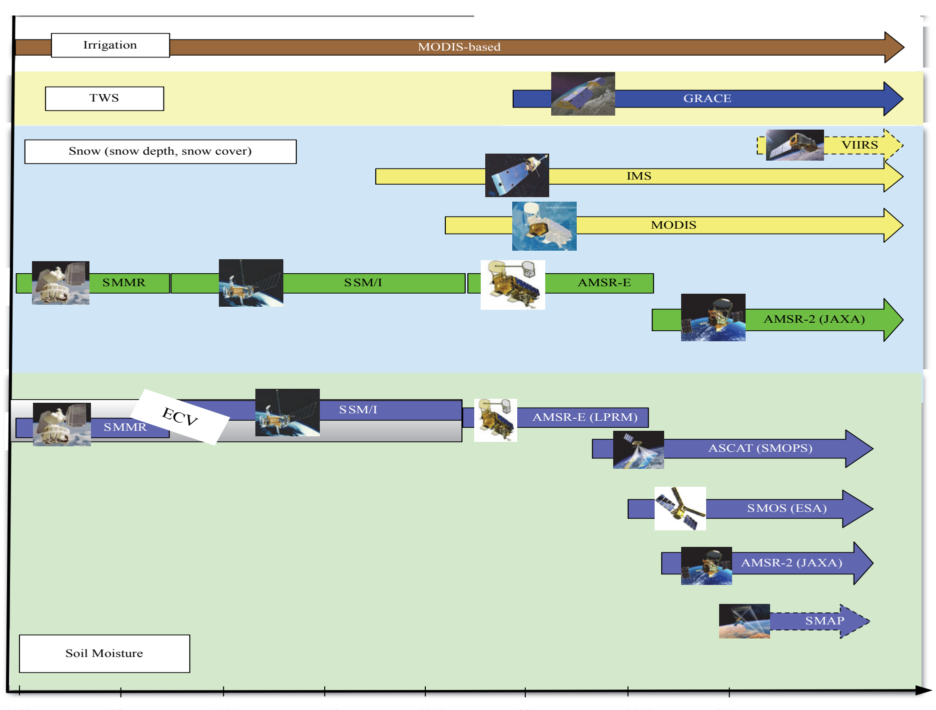


Figure 2: Chronological schematic of the satellite datasets assimilated in the next phase of NLDAS

The soil moisture and snow depth retrievals are assimilated using a one-dimensional ensemble Kalman Filter (EnKF), whereas the time-averaged (monthly) GRACE datasets are assimilated using an ensemble smoother algorithm. Small perturbations applied to the meteorological forcing and land surface states are used to generate the model ensemble spread. The snow cover retrievals are not directly assimilated, but used as constraint in the assimilation of the passive microwave retrievals. The assimilation of soil moisture and snow remote sensing datasets is found to provide marginal improvements in the soil moisture/snow fields. Further, these enhancements also provided improvements in the modelled streamflow (Figure 2; Kumar et al. 2014).

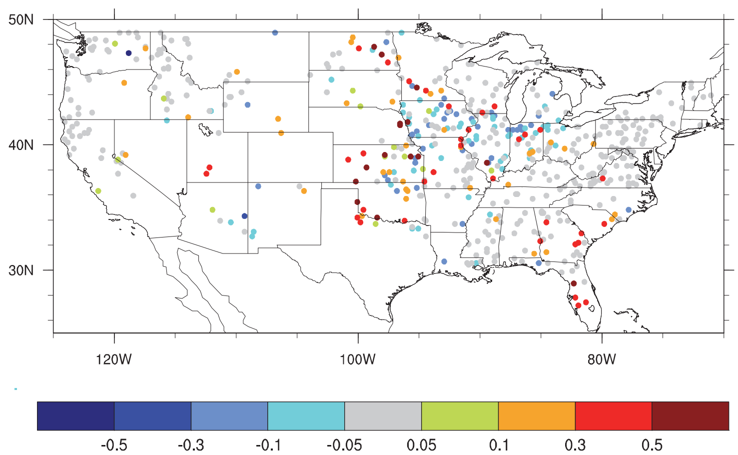


Figure 3: Normalized improvements in streamflow as a result of soil moisture data assimilation (warm colors show improvement, cool colors show degradation).

The assimilation of GRACE terrestrial water storage observations were found to have a positive impact on the simulation of unconfined groundwater variability across the majority of the US (Kumar et al. 2016). GRACE DA was also found beneficial in improving the representation of droughts (Figure 3).

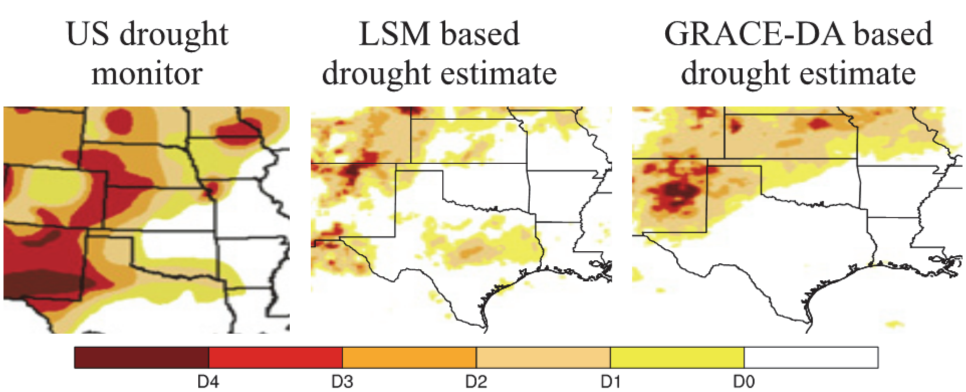


Figure 4: Beneficial impact of GRACE-DA on drought estimates in the southern US (Sep 2012)

There are significant challenges in land data assimilation systems. The accurate specification of the model error covariances is a difficult task. Most uncoupled land DA systems use a single forcing dataset to develop representations of model error covariances. The underestimation of the model error background can significantly limit the success of data assimilation, even if high quality observations

are available. The use of hybrid forcing initialization strategies is helpful in improving the performance of land DA in such cases. Figure 4 shows an example of this issue for assimilating snow depth retrievals from AMSR2. The single forcing based (OL1) simulation underestimates the snow evolution significantly and the lack of adequate model background significantly limits the skill of data assimilation (DA1). In contrast, when the ensemble is forced with a combination of different forcing products (FENS) or climatology (FCLIM), the assimilation performs better in incorporating AMSR2 observations.

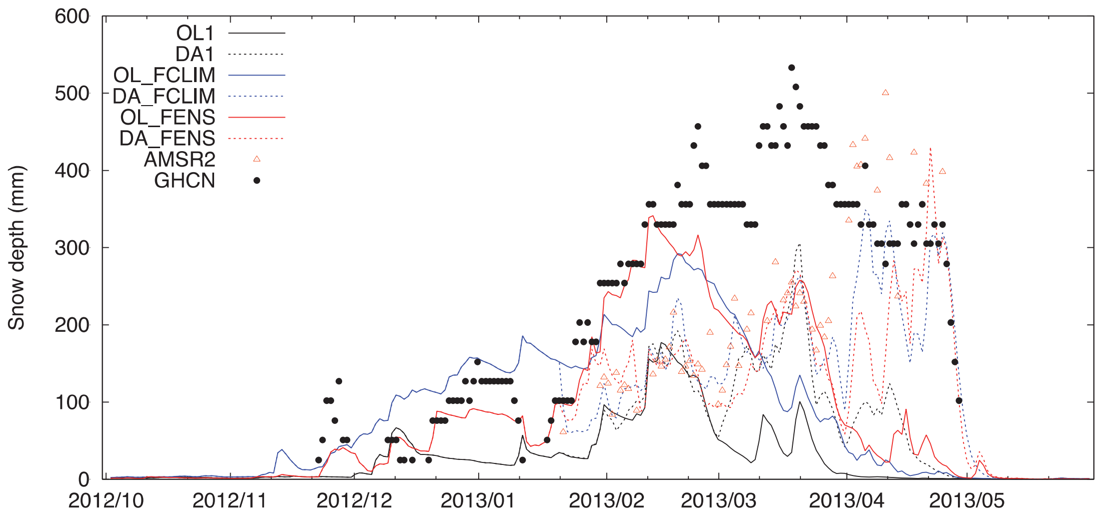


Figure 5: Time series of snow depth fields at a location near the Great Lakes region in the US from different model open loop and DA integrations.

Though satellite remote sensing is the most practical and effective way to observe and represent the complexity and heterogeneity of land surface, there are significant challenges in representing unmodeled processes through data assimilation. The recent study (Kumar et al. 2015) indicated that the current bias correction practices in land data assimilation are largely deficient when unmodeled processes dominate the bias between the model and observations. In such instances, strategies such as quantile mapping and trained forward modelling lead to the exclusion of signals from unmodeled processes. New methods are needed to preserve the observational information about unmodeled processes during data assimilation.

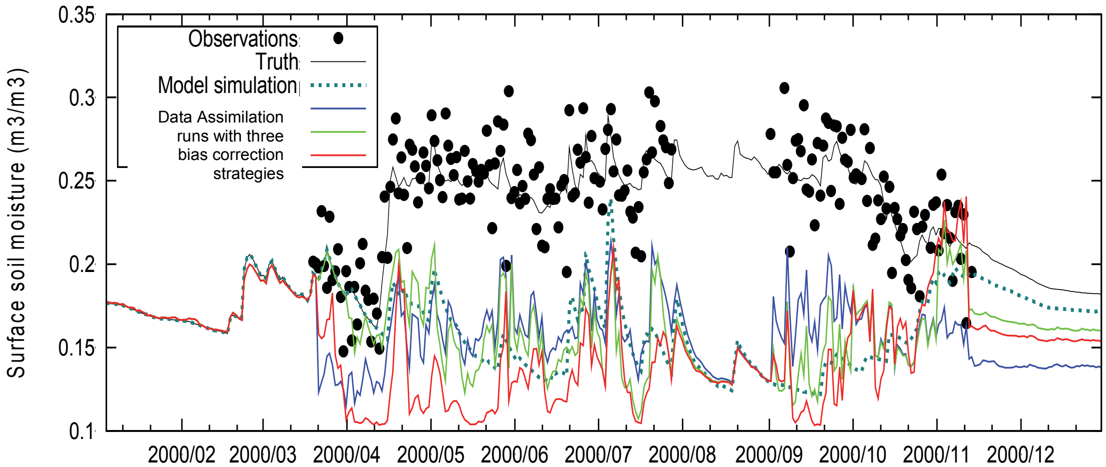


Figure 6: A comparison of different bias correction strategies in an idealized DA integration when the significant source of biases are from an unmodeled process such as irrigation. The DA integrations with different bias correction strategies (blue, green, red lines) lead to the exclusion of the true signal (black line).

References

Kumar, S.V., C.D. Peters-Lidard, Y. Tian, P.R. Houser, J. Geiger, S. Olden, L. Lighty, J.L. Eastman, B. Doty, P. Dirmeyer, J. Adams, K. Mitchell, E. F. Wood and J. Sheffield, 2006:Land Information System - An Interoperable Framework for High Resolution Land Surface Modeling. *Environmental Modelling & Software*, Vol.21, 1402-1415.2015.

Kumar, S.V., C.D. Peters-Lidard, D. Mocko, R. Reichle, Y. Liu, K. Arsenault, Y. Xia, M. Ek, G. Riggs, B. Livneh, M. Cosh, 2014: Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. *Journal of Hydrometeorology*, 15, 2446-2469.

Kumar, S.V., C.D. Peters-Lidard, J.A. Santanello, R.H. Reichle, C.S. Draper, R.D. Koster, G. Nearing, M.F. Jasinski, 2015: Evaluating the utility of satellite soil moisture retrievals over irrigated areas and the ability of land data assimilation methods to correct for unmodeled processes, Hydrol. Earth Syst. Sci., 19, 4463—4478.

Kumar, S.V., B.F. Zaitchik, C.D. Peters-Lidard, M. Rodell, R.H. Reichle, B. Li, D. Mocko, A. Getirana, G. Lannoy, M. Cosh, C. Hain, M. Anderson, K. Arsenault, Y. Xia, M. Ek, 2016: Assimilation of gridded GRACE terrestrial water storage estimates in the North American Land Data Assimilation System, J. Hydrometeor., in print, doi: 10.1175/JHM-D-15-0157.1.

ASSIMILATION OF EVAPORATIVE FRACTION INTO A SOIL VEGETATION ATMOSPHERE TRANSFER MODEL TO IMPROVE ROOT-ZONE SOIL MOISTURE

Dongryeol Ryu[1], Minoo Hashemian[1], Wade T. Crow[2], William P. Kustas[2], and Fuqin Li[3]

[1] *Department of Infrastructure Engineering, The University of Melbourne, Melbourne*
[2] *USDA-ARS Hydrology and Remote Sensing Lab, Bldg. 007, BARC-West, Beltsville, MD 20705, USA*
[3] *National Earth and Marine Observations Branch, Geoscience Australia, GPO Box 378, ACT 2601*

minooh@student.unimelb.edu.au

There is a growing interest in using remotely sensed thermal infrared observations to improve root-zone soil water content. Efficacy of the approach however relies on correctly reproducing the biophysical processes linking the surface thermal states to the root-zone water content. We present how variations in the modelled connection between surface fluxes and root-zone soil moisture affect the efficiency and veracity of transferring surface information into the root-zone through sequential data assimilation. In particular, the evaporative fraction (EF) derived from thermal infrared (TIR) measurements at the OPE3 site via the two-source energy balance model (TSEB) is assimilated into two different land surface models: 1) the simple water and energy balance-soil vegetation atmosphere transfer (SWEB-SVAT) model and 2) the multi-layer WEB-SVAT with dynamic root distribution (MWSDR). Results show that the assimilation of TIR-based EF products improves root-zone soil moisture only when the SVAT component correctly reproduces the biophysical links between surface-root zone using MWSDR.

CONSTRAINTS ON THE GLOBAL MARINE IRON CYCLE FROM A SIMPLE INVERSE MODEL

Frants, M.[1], Holzer, M.[1] , DeVries, T.[2] and Matear, R.J.[3]

[1]*Department of Applied Mathematics, School of Mathematics and Statistics, University of New South Wales, NSW*
[2]*Department of Geography/Earth Research Institute, University of California, Santa Barbara, California, USA*
[3]*CSIRO Ocean and Atmosphere, Hobart, Tasmania*

richard.matear@csiro.au

Iron plays an important role in the ocean's uptake and cycling of carbon but iron's behavior in the ocean is complex and poorly constrained. Hence, global biogeochemical models use many different ways to represent iron cycling, none of which are well grounded in the observations. Here, I present a simple model of the global marine iron cycle is used to constrain the sources, sinks, and biological cycling of iron. The iron model is embedded in a data-assimilated steady state circulation, with biological cycling driven by a prescribed, data-constrained phosphate cycle. Biogeochemical parameters are determined by minimizing a suitably weighted quadratic mismatch with available dissolved iron (dFe) observations, including GEOTRACES transects. Because the effective iron sources and sinks overlap, current dFe observations cannot constrain sources and sinks independently. We therefore determine a family of optimal solutions for a range of the aeolian source strength σ_A from 0.3 to 6.1 Gmol/yr. We find that the dFe observations constrain the maximum Fe:P uptake ratio R to be proportional to σ_A, with a range that spans most available measurements. Thus, with either R or σ_A specified, a unique solution is determined. Global inventories of total and free iron are well constrained at $(7.4 \pm 0.2) \times 10^{11}$and $(1.39 \pm 0.05) \times 10^{10}$ mol Fe, respectively. The dFe distributions are very similar across the family of solutions, with iron limitation in the known high-nutrient low-chlorophyll regions. Hydrothermal source strength ranges from 0.57 to 0.73 Gmol/yr and does not vary systematically with σ_A suggesting that the hydrothermal and aeolian parts of the iron cycle are largely decoupled. The hydrothermal dFe anomaly in the euphotic zone is ∼10% and concentrated in subpolar regions of iron limitation. Enhanced ligand concentrations in old waters and in hydrothermal plumes are necessary to capture key features of the dFe observations.

WHAT DO YOU DO WHEN YOU DON'T KNOW WHAT YOU'RE DOING: PARAMETER ESTIMATION WITH AN ENSEMBLE OF MODELS

Peter Rayner

School of Earth Sciences, University of Melbourne, Melbourne

prayner@unimelb.edu.au

One aspect of biogeochemical modelling is low confidence about the basic equations. This leads to a higher divergence of modelling approaches than are possible in better understood systems. Naturally any statements we make about quantities in such models depend on the choice of model; in technical terms our inferences are conditioned on the model. Traditionally this is handled by an intercomparison in which we make inferences with an ensemble of models. The problem is naturally handled by hierarchical Bayesian approaches. In this talk we will briefly sketch the basic theory and present an example from the TRANSCOM intercomparison of atmospheric transport models and large-scale carbon fluxes. We see that the available data provides an unrealistically strong discriminant among the models and that this is due to incorrect specification of the data covariance. Ensemble statistics are surprisingly insensitive to the method we use to construct the model ensemble. Cross-validation, in which the models are rated against some withheld data, does produce strikingly different statistics than the conventional ensemble mean and variance.

RECENT ADVANCES IN DA AT NCEP

John C Derber

Environmental Modeling Center
NCEP/NWS/NOAA, USA

john.derber@noaa.gov

The evolution of data assimilation techniques at all operational NWP centers over the last 30 years has been very rapid. The DA system at NCEP is no exception. Recently, the operational global system has evolved to a 4d-Hybrid EnVar system with the direct use of satellite radiances (all-sky for the microwave) and bending angles for the GPS-RO data. While the changes to the operational DA systems has been rapid and resulted in very significant improvement in the analyses, there are still many opportunities to make significant enhancements to the system. Of course, all modifications to the operational systems must adhere to operational timeliness and reliability requirements. In this presentation, the shorter term developments and long term development path at NCEP will be presented.

RECENT DEVELOPMENTS IN SATELLITE DATA ASSIMILATION AT THE MET OFFICE

Bill Bell[1], Nigel Atkinson[1], Anna Booton[1], Chris Burrows[1], James Cameron[1], Fabien Carminati[1], James Cotton[1], Amy Doherty[1], John Eyre[1], Mary Forsythe[1], Pete Francis[1], Katie Lean[2], Dingmin Li[1], Andrew Lorenc[1], Stefano Migliorini[1], Stu Newman[1], Ed Pavelin[1], Indira Rani[3], Rick Rawlins[1], Roger Saunders[1], Andrew Smith[1], Fiona Smith[1], Warren Tennant[1], Mike Thurlow[1], Francis Warwick[1], Peter Weston[2], Marek Wlasek[1]

[1] *Weather Science, Met Office, Exeter, UK*
[2] *Research Department, ECMWF, Reading , UK*
[3] *NCMRWF, Noida, India*

william.bell@metoffice.gov.uk

Introduction

Improvements in the assimilation of satellite observations have led to significant performance gains in the Met Office global model over the last two years (Figure 1). Upgrades in the treatment of satellite data have formed the core of the global forecasting system updates at Parallel Suite 35 (February 2015), Parallel Suite 37 (March 2016) and Parallel Suite 38 (November 2016) and have reduced RMS errors for Day 3 forecasts by 4% (10%) in northern (southern) hemisphere sea level pressure and 500 hPa geopotential height, for example.

These upgrades have included:

- A significant expansion of the satellite observing system
- Improvements in the treatment of existing data
- Novel developments in the treatment of sounding data over land

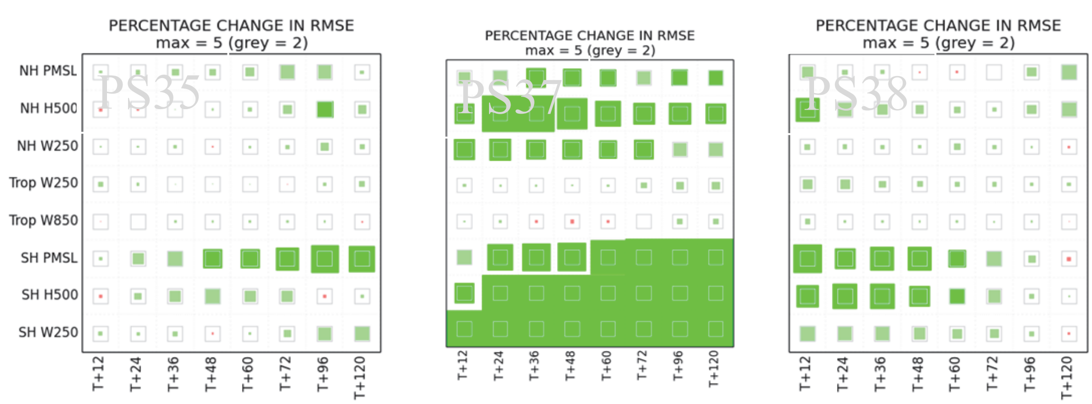


Figure 1: Change in RMS errors for mean sea level pressure, 500 hPa heights, and 250 hPa winds in the Extratropics and 250 / 850 hPa winds in the Tropics for forecast ranges to Day 5 for 2015-16 Parallel Suites relative to the operational global model for PS35 (left), PS37 (centre) and PS38 (right). Verification is relative to analysis and includes more than 150 forecasts for each upgrade.

New Data

The operational upgrade at PS37 included observations from two microwave imagers: AMSR-2 (pm-orbit) and F-17 SSMIS (am-orbit) over ocean. Observations from the Megha-Tropiques humidity sounder, SAPHIR, were also introduced. All six 183 GHz channels are assimilated over land and sea. PS37 also saw the first use of data from China's polar orbiting FY3 satellites, with humidity sounding channels from MWHS-2. Observations from the FY3-B humidity sounder, MWHS-1, were introduced at PS38. Wind observations from RapidScat aboard the International Space Station were introduced into operations in early 2016. Himawari-8 clear sky radiances were introduced at PS38 as well as VIIRS winds.

Improved treatment of existing data

Changes have been made in the treatment of biases, observation errors, and in the quality control of observations. Most significantly, variational bias correction (VarBC) was introduced at PS37. The implementation is similar to that at other centres, with some exceptions: the introduction of a bias halving time parameter, which determines the rate of convergence of the scheme; a two step correction of cross scan biases, using a spot dependent offset, coupled with a set of Legendre polynomials for the correction of residual, time dependent biases; and a Fourier series correction for orbital biases. VarBC accounts for ~70% of the benefit realised in PS37. VarBC reduces a longstanding warm bias in lower tropospheric temperatures relative to other NWP centres, and reduces the spin down of low level temperatures.

Correlated inter-channel errors for AIRS (PS35) and CrIS (PS37) were introduced following the treatment of IASI (implemented at PS31, Weston *et al* (2014)). The observation errors for humidity sounding radiances were also retuned at PS38, typically resulting in smaller observation errors. Background fits to independent humidity observations were significantly improved. The background error matrix in the Met Office 1D-Var pre-processor was updated to reflect recent changes in the 4D-Var B-matrix, resulting in more stringent quality control of humidity sensitive radiances.

Developments in the assimilation of sounder data over land

Low peaking sounding channels from AIRS and CrIS observations over land were introduced at PS38 following the 1D-Var dynamic emissivity estimation method developed by Pavelin and Candy (2014). Low peaking AMSU-A channels (4 and 5, at 52.8 and 53.6 GHz respectively) were introduced at PS38, using dynamically estimated emissivity (using AMSU-A channel 3 at 50.3 GHz) from 1D-Var.

References

P. P. Weston, W. Bell and J. R. Eyre, Accounting for correlated error in the assimilation of high resolution sounder data, Quarterly Journal of the Royal Meteorological Society, Volume 140, Issue 685, October 2014 Part B, Pages: 2420–2429, DOI: 10.1002/qj.2306

E. G. Pavelin and B. Candy, Assimilation of surface-sensitive infrared radiances over land: Estimation of land surface temperature and emissivity, Quarterly Journal of the Royal Meteorological Society, Volume 140, Issue 681, April 2014 Part B, Pages: 1198–1208, DOI: 10.1002/qj.2218.

ON THE USE OF ATMOSPHERIC MOTION VECTORS AT NCEP GFS

Iliana Genkova[1], Sharon Nebuda[2], Dave Santek[2], Andrew Collard[1]

[1]*I. M. Systems Group for Environmental Modeling Center, NOAA, Maryland, USA*
[2]*University of Wisconsin, Madison, WI, USA*

iliana.genkova@noaa.gov

Satellite winds are changing for multiple reasons – new imaging instruments are launched, novel retrieval algorithms are implemented, more Atmospheric Motion Vectors (AMVs) producers are joining the Global Observing System. Simultaneously NWP models evolve as well.

NCEP's GSI is now updated and ready to assimilate the following Atmospheric Motion Vector products: i) winds from HIMAWARI-8; ii) polar VIIRS winds; iii) GOES Clear Air Water Vapor winds.

The approaching launch of GOES-R led to improving the AMVs observation operator in order to accommodate a new winds retrieval algorithm and utilize some new AMV characteristics.

Leo-Geo winds are a new data set which would help with reconciling the coverage discontinuity between polar winds and winds from geostationary imagers. Preliminary results from an evaluation will be shown.

We'll also present a study focusing on the complimentary nature of AMVs and Doppler lidar winds, in anticipation of the JMA and ESA's Doppler winds lidar space instruments.

BENEFITS FROM ADVANCES IN THE ASSIMILATION OF EARTH OBSERVATIONS FROM SPACE

John Le Marshall[1], David Howard[1], Robert Norman[2], Yi Xiao[1], Jim Jung[3], Chris Tingwell[1], Kefei Zhang[2], Paul Lehmann[1], Tim Morrow[1], Jaime Daniels[4] , Steve Wanzong[3] and Wayne Bresky[4]

[1] *Bureau of Meteorology, Melbourne*
[2] *RMIT University, Melbourne*
[3] *UW/CIMSS, Madison, USA*
[4] *NOAA/NESDIS/STAR, College Park, USA*

Earth Observation from Space (EOS) make a considerable impact on the accuracy of numerical weather prediction (NWP) in the southern hemisphere. The quality of the one-day forecast over the southern hemisphere based on conventional observations is the same as the quality of a four-day forecast based on satellite (EOS) and conventional data, when verification is based on analyses using all data (EOS and conventional). In effect EOS data extend the length of a high quality NWP forecast by a factor of four. In the northern hemisphere they extend the length of a high quality forecast by a factor of 1.6. This is summarised in Le Marshall et al, 2013.

Currently a number of comparatively new technologies and instruments have been placed in space for use in NWP and a number are soon to follow. These include the advanced sounders AIRS, IASI and CrIS, the Advanced Himawari Imager (AHI) on Himawari-8, the Advanced Baseline Imager (ABI) and the Geostationary Lightning Mapper (GLM) on GOES-R, the wind lidar ADM Aeolus, JPSS and the COSMIC-1 and 2 constellation of satellites. The important contribution these new technologies and instruments have and will make, particularly over Australia and in the southern hemisphere are described in some detail.

In relation to the mass field advanced sounders AIRS, IASI and CrIS provide very large impact per instrument in current Forecast Sensitivity to Observations (FSO) studies while still having the potential to contribute far more to temperature and humidity analysis through use of the information content of their observations. See for example Le Marshall et al., 2008, 2014 where the benefits of using additional channels, more moisture information and partially cloudy observations in the assimilation process are documented.

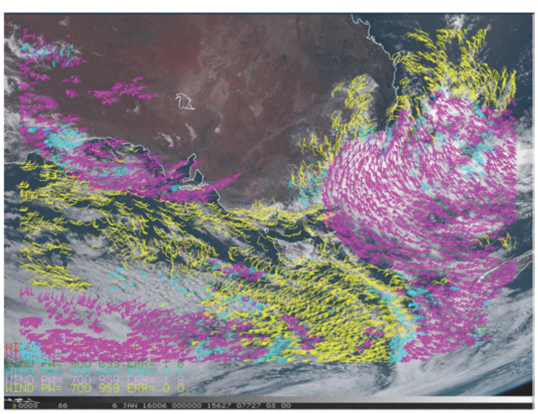


Fig1 Himawari-8 AMVs using IR (11 µm) channel 14 tracers at 00 UTC 16 January 2016 using the current operational system

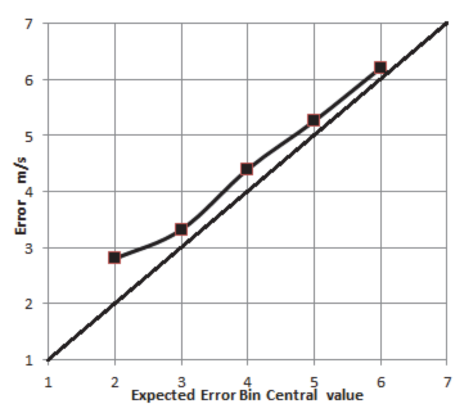


Fig. 2. Measured error (m/ s) vs Expected Error (m/s) for upper-level Himawari-8 IR winds (1 August –31August 2016).

In relation to the wind field, high-resolution near continuous (hourly) atmospheric motion vectors (AMVs), derived locally from Himawari-6 and 7 observations, have provided significant benefit to operational numerical weather prediction in the southern hemisphere when assimilated using 4D Var. (Le Marshall et al., 2013a). The new generation of satellites such as Himawari-8 and GOES-R will provide far greater spatial, temporal and spectral resolution observations. To date high spatial and temporal resolution AMV's have been locally generated from Himawari-6 and 8 images separated by 10 min. These AMV's have been error characterised and used in a series of data assimilation experiments, for example Le Marshall et al., 2016. An example of such winds, generated in real time every 10 minutes for operational use from Himawari-8 data, is seen in Figure 1 while Figure 2 shows the accuracy of the error characterisation through use of the Expected Error. The full impact these locally generated ten minute winds from Himawari-8, on operational analysis and forecasting is still being determined however assimilation results so far have led to the introduction of these continuous 10 min winds into the Bureau's operational numerical weather prediction system.

Other benefits from advances in the use of EOS have also been recorded. These include benefits arriving from the use of radio occultation data in the operation NWP system (see for example Le Marshall et al., 2012). The beneficial impact on the basic forecast system is quite large, while impact related to the observations being near bias free, assists in the generation of climate quality analyses and in monitoring the temperature trends over the southern hemisphere. Related work has also been undertaken in terms of detailing the modelling of the occultation process. This has resulted in improved understanding of the position of GNSS rays in relation to horizontal refractive index gradients and in understanding the impact of changing ionospheric conditions on the occultation process.

In summary, the considerable benefits from using EOS in NWP have been noted. Improved observational capacity will continue to be expanded with the launch of instruments such as the Advanced Baseline Imager (ABI) and the Geostationary Lightning Mapper (GLM) on GOES-R, the wind lidar ADM Aeolus, JPSS and the COSMIC-2 constellation of satellites. The important contribution the new technologies and instruments will make, particularly over Australia points to the next decade being one with potentially significant improvement in observational capacity and hence a significant benefits from its exploitation.

References

Le Marshall, J., Jung,J., , Goldberg, M.., Barnet, C., Wolf, W., Derber, J., Treadon, R. and S. Lord. 2008. Using Cloudy AIRS Fields of View in Numerical Weather Prediction . *Aust. Meteor. Mag*.57, 249 – 254.

Le Marshall John, Yi Xiao, Robert Norman, Kefei Zhang, Anthony Rea, Lidia Cucurull, Rolf Seecamp, Peter Steinle, K. Puri and Tan Le. 2012, , The Application of Radio Occultation for Climate Monitoring and Numerical Weather Prediction in Australian Region . *Aust. Meteor. and Ocean. Jnl.* **62**, 323 – 334.

Le Marshall, J. F., Lee, J., Jung, J. A., Gregory P. and Roux, B. (2013). The Considerable Impact of Earth Observations from Space on Numerical Weather Prediction. *Aust. Meteor. and Ocean. Jnl*., **63**(4), 497-500.

John Le Marshall , Rolf Seecamp, Yi Xiao, Paul Gregory, Jim Jung, Peter Steinle, Terry Skinner, Chris Tingwell and Tan Le. 2013a. The Operational Generation of Continuous Winds in the Australian region and their Assimilation with 4DVar. 2013. *Wea. Forecasting,***28,** 504 – 514.

Le Marshall J., Jung J,A., Lee J., Barnet C. and E. M. 2014. Improving Tropospheric and Stratospheric Moisture Analysis with Hyperspectral Infrared Radiances, 2014, . *Aust. Meteor. and Ocean. Jnl*., **64**, 283-288.

Le Marshall John., Yi Xiao, David Howard, Chris Tingwell, Jeff Freeman, Paul Gregory, Tan Le, Denis Margetic, Jamie Daniels, Jim Jung, Wayne Bresky, Andrew Bailey and Steve Wanzong. 2016. First Results from the Generation and Assimilation of Ten Minute Atmospheric Motion Vectors in the Australian Region.2016, JSHESS, **66**, 12-19.

ESTIMATION OF DIRECTIONAL TROPOSPHERIC HORIZONTAL GRADIENTS AND ITS IMPACT ON GPS-DERIVED TROPOSPHERIC ZENITH DELAY PRODUCTS

Salim Masoumi, Simon McClusky, Achraf Koulali and Paul Tregoning

Research School of Earth Sciences, Australian National University, Canberra

salim.masoumi@anu.edu.au

Improper modelling of horizontal tropospheric gradients in Global Positioning System (GPS) analysis induces errors in the estimation of tropospheric path delays. Under certain conditions where there are isolated and rapid spatial changes of tropospheric refractivity at discrete azimuths around a GPS station, the conventional two-axis (North-South and East-West; NS/EW) tilted plane model of horizontal tropospheric gradients in GPS analysis fails to provide an accurate representation of tropospheric gradients. These directionally asymmetric weather scenarios are caused by phenomena such as topography-induced gravity waves in the atmosphere or the passage of isolated severe thunder storms. The GPS-estimated tropospheric delays are systematically biased in these situations, if the gradients are not properly modelled. In this study we have developed a new parametrisation of tropospheric gradients whereby an arbitrary number of gradients are estimated as discrete directional wedges around a GPS station, with a piecewise formulation connecting the wedges. We show both via simulations and two case studies that the new directional model is able to significantly increase accuracy of derived parameters in complicated weather scenarios where tropospheric delays are non-planar and highly direction dependent.

Simulations

We simulated several different tropospheric gradient scenarios, under a range of conditions, and attempted to recover tropospheric properties using both conventional planar gradient model and the new directional gradient parametrisation. We show that the proposed directional model of gradients is able to significantly improve parameter estimates, particularly tropospheric zenith total delays and station heights, for specific simulation cases when the tropospheric delay changes do not follow a simple planar model. Under these particular scenarios, the errors in the estimates of zenith total delays are reduced by ~88% when using the new directional gradient model compared to when using the conventional planar assumption for the gradients (Figure 1).

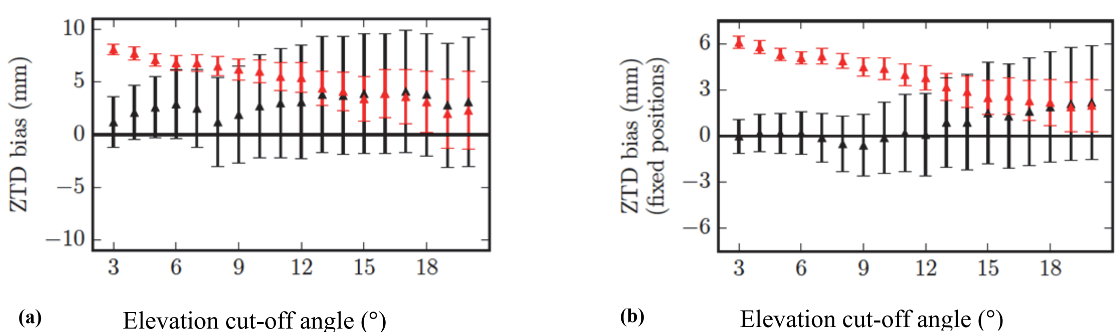


Figure 1: Errors in the estimation of zenith total delay (ZTD) at a simulation scenario with different choices of elevation cut-off angle for GPS observations where there are asymmetric gradients to the north-east of a GPS station and no gradients to the south-west, using conventional planar gradient model (red) and new directional

gradient model (black): (a) when the position parameters are estimated along with tropospheric parameters, and (b) when the position components are kept fixed to their true values during the GPS processing

Case study of 8-9 September 2002 in southern France

For an extreme rain event that occurred in September 2002 in southern France, we show that the new tropospheric gradient parametrisation provides a more accurate image of the amount of water vapour accumulation, isolating the strong gradients in particular azimuths consistent with the V shape pattern of precipitation. Figure 2 shows the hourly precipitation data at one epoch on 8 September 2002 together with the GPS tropospheric delays due to the horizontal gradients estimated by the directional gradient model at GPS station MTPL.

(a) **(b)**

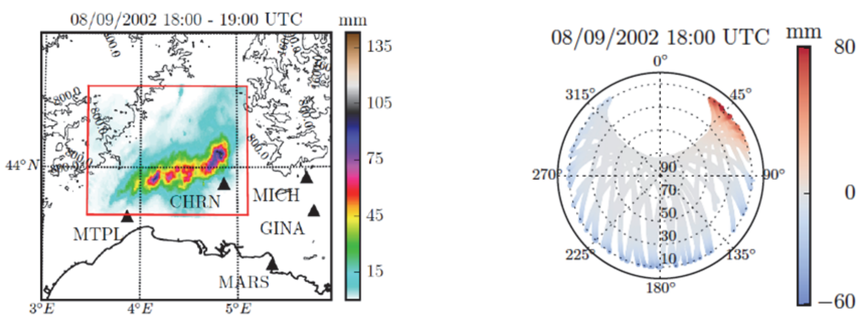
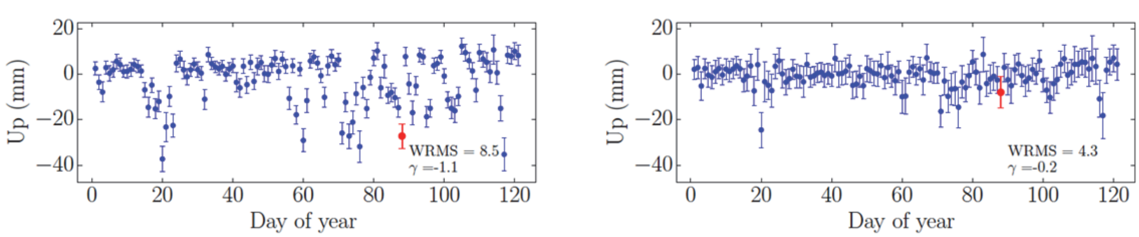


Figure 2: (a) Hourly precipitation on 8 September 2002 from 18:00 to 19:00 UTC time in the Gard region in southern France (Data courtesy of http://www.ohmcv.fr); (b) GPS delays due to gradients estimated at site MTPL using the directional gradient model.

Plate Boundary Observatory GPS stations in Sierra Nevada region

We show that the directional model of gradients removes the majority of outliers in the highly scattered height time series of a set of GPS stations around Mammoth Lakes region in California, improving the repeatability of the average height time series of the stations by about 31% when a 3^{o} elevation cut-off is chosen for Gfig3bPS observations. The improvement is from 8.5 mm to 4.3 mm in terms of WRMS for one of the worst affected sites P631. For a particular day at the same station, there are biases of larger than 15 mm in the estimation of zenith delays when comparing the solutions with conventional gradient model and the directional model of gradients. These biases are due to mismodeling of the tropospheric gradients by the conventional model, and are important for meteorological applications.



(a) NS/EW model **(b)** directional model

(c) Tropospheric zenith total delay adjustments

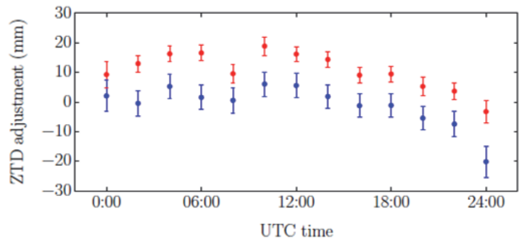


Figure 2: Time series of the vertical positions for site P631 and for the period of 1 January 2012 to 30 April 2012 when using (a) the conventional NS/EW gradient model of gradients, compared to (b) when using a directional model for the gradients. The date 28 March 2012 is signified by red colour. The 2-hourly adjustments made to the zenith total delays for the same site on 28 March 2012 are displayed in (c) when using the NS/EW model (red) compared to when using the directional model (blue) for the gradients.

ANOMALOUS GNSS RADIO OCCULTATION DATA

Robert Norman[1,2], Brett A. Carter[1,2], John Le Marshall[3,2], Ara Cate[2] and K. Zhang[2]

[1]SERC Limited, Canberra

[2]SPACE Research Centre, RMIT University, Melbourne

[3]Research and Development Branch, Bureau of Meteorology, Melbourne

robert.noman@rmit.edu.au

The Global Navigation Satellite Systems (GNSS) Radio Occultation (RO) is a space-based technique for sounding the Earth's atmosphere with global coverage and is becoming more widely used in operational applications. The fundamentals of the atmospheric parameter retrieval techniques from the GNSS RO data rely on simplifying assumptions, such as spherical stratification of the refractive medium and disregard the downrange and transverse refractive gradients in the atmosphere.

EUMETSAT, in cooperation with the United Kingdom Meteorological Office (UKMO) revealed regions in the Earth's atmosphere where up to 60 % of GRAS (GNSS Receiver for Atmospheric Sounding) RO measurements exhibit anomalous bending angles. GRAS forms part of the payload on the LEO Metop series of meteorological satellites. In these regions the GPS L1 signal bending angle is greater than that of the corresponding bending angle of the GPS L2 signal paths at Tangent point altitudes of 50 km. They also found that the associated L1, L2 bending angles are smaller than the corresponding calculated neutral (ignoring the ionosphere) bending angles. This has also been shown in COSMIC data with very similar characteristics.

The electron density gradients in the ionosphere cause the GNSS signals to bend/refract. The ionosphere is a dispersive medium and this means that the refractive gradients in the ionosphere will cause the GPS L2 signal paths to bend more than the corresponding GPS L1 signal paths.

The global percentage occurrences of anomalous GRAS GNSS RO results, at the 50 km tangent height, during the 2011 December (a) and June (c) solstices and March (b) and September (d) equinoxes using operational data are shown in Figure 1. Each season includes 3 months of data, including the month of the solstice or equinox and one month before and after. The greatest fraction of anomalous GRAS GNSS RO events occur poleward of geomagnetic latitude 30º in both hemispheres. Figures (1a & c) reveal that during the December and June solstices the anomalous occurrences are located predominately in the wintertime hemispheres. Figures (1a - d) show that the anomalous RO events occur predominately along the geomagnetic latitudes of ±30º and ±60º and near the geomagnetic poles. In fact, the occurrence of anomalous RO events reaches 60% in these regions. At these locations there exist major features in the ionosphere.

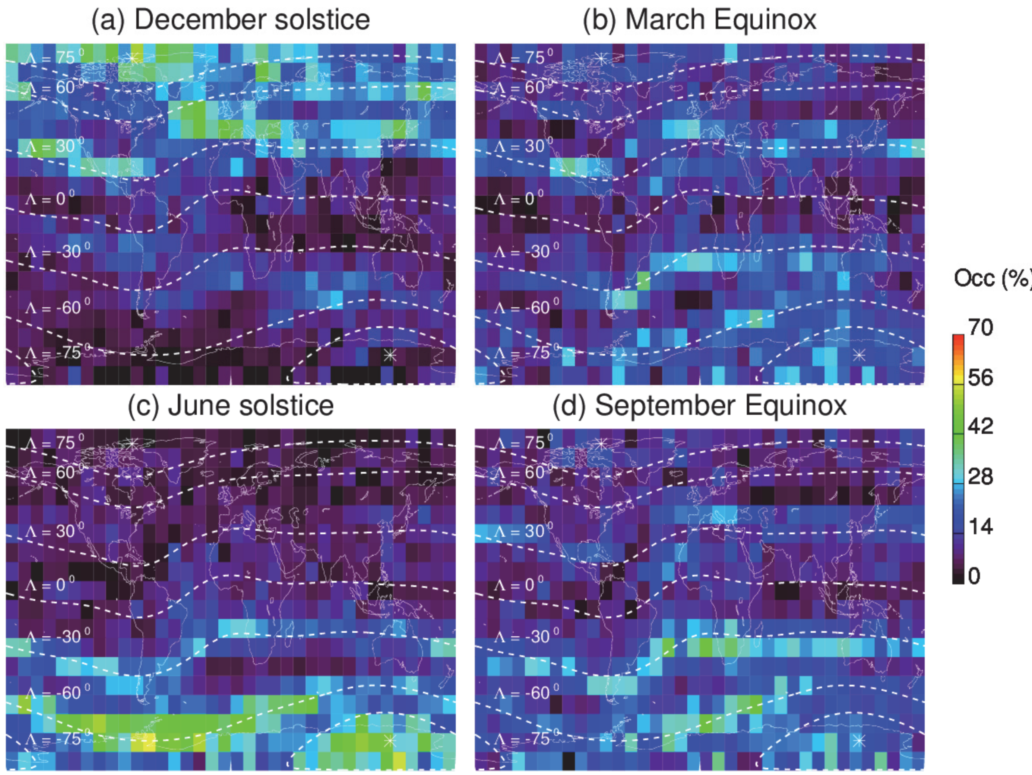


Figure 1: The occurrence of anomalous GPS RO events during the December (a) and June (c) solstices and March (b) and September (d) equinoxes during 2011. Each season includes 3 months of data, including the month of the solstice or equinoxes and one month before and after. The dashed lines show the magnetic latitudes 0°, ±30°, ±0° and ±75°, and the asterisks show the locations of the magnetic poles. The figure is from Norman et al (2016).

In this contribution a three dimensional numerical ray tracing techniques based on geometrical optics is used together with models of the ionosphere, lower atmosphere and magnetic field to simulate the GNSS RO signal paths. Identifying the regions of the ionosphere causing increased anomalous GNSS RO results and understanding the ionospheric characteristics producing these anomalous results will be discussed.

The research findings on the anomalous GNSS RO events will enhance current understanding of GNSS RO signal paths and how they may affect atmospheric retrieval techniques. The results presented will provide greater confidence in RO data quality control techniques which are important for agencies around the world that utilize GNSS RO data for operational meteorology, short term climate trend analysis, ionospheric modeling and monitoring of space weather.

Acknowledgements

The authors would like to acknowledge the support of the Cooperative Research Centre for Space Environment Management (SERC Limited) through the Australian Government's Cooperative Research Centre Programme.

References

Norman, R., Carter, B.A., Healy, S.B., Culverwell, I.D., von Engeln, A., Le Marshall, J., Younger, J.P., Cate, A. and K. Zhang, 2016: Ionospheric regions producing anomalous radio occultation results, IEEE TGRS (submitted Oct 2016).

ASSIMILATING OBSERVATIONS WITH SPATIALLY AND TEMPORALLY CORRELATED ERRORS IN A GLOBAL ATMOSPHERIC MODEL

Jeffrey Anderson

Institute for Mathematics Applied to Geosciences
National Center for Atmospheric Research, USA

jla@ucar.edu

Ensemble data assimilation systems for large geophysical problems like the atmosphere and ocean often ignore the possibility of correlated errors between observations at different spatio-temporal locations. However, most instruments are known to have correlated errors. The correlated errors can range from simple time-averaged bias to complicated functions of both the state of the geophysical system and the observation geometry for instruments like satellite radiometers. One possible solution is to construct a statistical model that predicts the correlated part of the error for a given instrument and remove the estimated error before assimilation. Here, a complementary approach is studied in which differences between correlated observations are assimilated rather than the raw observations. Low-order model results comparing assimilation of raw observations with correlated errors to assimilations of various types of differences are presented. Results from OSSEs using an atmospheric general circulation model with simulated observations with correlated errors are presented to illustrate the impacts of assimilating differences for numerical weather prediction. The relative performance of assimilating observations that are temporal differences versus spatial differences provides insight into the strength of correlations between various differences and the model's nonlinear dynamics.

NCUM DATA ASSIMILATION SYSTEM: PRESENT STATUS

John P. George

National Centre for Medium Range Weather Forecasting (NCMRWF)
Ministry of Earth Sciences, A-50, Sector-62, NOIDA, India

john@ncmrwf.gov.in

Unified Model implemented at NCMRWF has been used for Numerical Weather Prediction (NWP) since 2012 (Rajagopal et al., 2012). The NCMRWF UM (NCUM) system is being upgraded periodically to adapt the new developments. Recently, Met Office Parallel Suite 37 (PS37) based Hybrid 4D-Var data assimilation system has been implemented at NCMRWF for global NWP. The NCUM assimilation-forecast system uses user friendly “Rose/cylc” software environment. This paper describes details of the NCUM data assimilation system and the performance evaluation of the assimilation-forecast system.

The major components of the NCUM data assimilation (NCUM-DA) system includes (i) Observation Pre-processing system developed at NCMRWF which packs the various observations received at NCMRWF into “obstore” format, (ii) Observation Processing System (OPS) which prepares the observations for assimilation (iii) four dimensional variational data assimilation system (4D-Var) which produce the initial conditions for the model (iv) the Unified Model used for Global Numerical Weather Prediction. UM Surface data preparation system (SURF) implemented at NCMRWF is used for the preparation of SST, Sea Ice and Soil Moisture analysis. Soil moisture analysis uses Extended Kalman Filter (EKF) algorithm.

Horizontal resolution of new NCUM system is 17 km and has 70 levels in the vertical (N768L70) extends from surface to 80 km height. Hybrid 4D-Var at 40 km horizontal resolution (N320L70) produces the analysis increments for the model. NCUM data assimilation system runs at 6 hourly cycle. 10-day forecasts based on 00 UTC analysis and 5-day forecasts based 12 UTC analysis are generated routinely. NCMRWF Global Ensemble Prediction system (NGEPS) based on UK Met Office MOGREPS, provides the ensemble forecasts to the Hybrid 4D-Var system. The N400L70 resolution (~33km horizontal resolution) NGEPS system has 44 members.

Assimilation of more observations is of utmost importance and various efforts are being made to ingest more observational data into the data assimilation system. NCMRWF was actively involved in the development of Indian radiance assimilation capability in the Met Office 4D-Var system. INSAT-3D Sounder radiances and Megha-Tropiques SAPHIR radiances are operationally used in the NCUM-DA system in addition to other observations. Immediate future plans for satellite data assimilation includes the efforts to ingest new Indian satellites, INSAT-3DR (INSAT-3D Repeat) and SCATSAT-1 observations into the NCUM-DA system. Table-1 provides the list of various satellite observations used in the NCUM Hybrid 4D-Var data assimilation system. Studies are being carryout to understand the impact of various satellite observations in the NCUM system. Aerosol optical depth (AOD) from MODIS is included in the DA system for dust assimilation. Additionally, data assimilation experiments are being carryout with INSAT-3D AOD.

A high resolution DA system (3D-Var) for the 4-km resolution regional NCUM model for the Indian region is being tested. Radial wind from Indian DWR is also included in this high resolution regional data assimilation system.

To improve the initial position of the tropical cyclones (TC) in NCUM system, a Vortex Specification (VS) method of Davidson et al., 2014 was tested in the Global NCUM-DA system. Synthetic vortex structure is generated based on observed location, central pressure and radius of outer closed isobar. The synthetic observations of Mean Sea Level Pressure (MSLP) produced by VS scheme over the vicinity of cyclone is used in 4D-Var data assimilation to improve the TC forecast over the Indian seas experimentally.

IMDAA regional reanalysis system developed by UK Met Office is implemented at NCMRWF and the production runs are started for the initial 10-year period (1979-1988). This satellite-era atmospheric regional reanalysis at 12 km horizontal resolution aims to improve the understanding of the Asian monsoon.

Performance evaluations of NCUM forecast is being carried out systematically and evaluation scores are routinely presented in the NCMRWF web site (www.ncmrwf.gov.in). The details of the performance evaluation of the NCUM system will be presented in the meeting.

Table-1: Satellite observations used in the NCUM Hybrid 4D-Var DA system

Observation Type	Satellite
AMSU/MHS radiances	NOAA , MetOp Satellites
HIRS clear radiances	NOAA , MetOp Satellites
IASI and AIRS radiances	MetOp satellites (IASI), Aqua (AIRS)
ATMS & CrIS radiances	Suomi NPP
SAPHIR radiances	Megha-Tropiques
Geo Imager – cloud clear IR radiance	Meteosat , GOES Satellites
Geo Sounder – cloud clear IR radiance	INSAT-3D
GPS RO bending angles	COSMIC, MetOp/GRAS, GRACE
GPS ZTDs	Global observations (various locations)
Satellite Atmospheric Motion Vectors	INSAT-3D, Meteosat-7& 10, GOES-E& W, Himavari , AQUA, NOAA & MetOp Satellites
Scatterometer: Sea-surface winds	MetOp Satellites (ASCAT)
Surface soil moisture/wetness	MetOp Satellites (ASCAT)
Aerosol Optical Depth (for dust)	AQUA (MODIS)

References:

Davidson, N. E., and Co-authors 2014: ACCESS-TC: Vortex Specification, 4D-VAR Initialization, Verification and Structure Diagnostics, Monthly Weather Review, 142(3): 1265-1289.

Rajagopal E.N., G.R. Iyengar, John P. George, Munmun Das Gupta, Saji Mohandas, Renu Siddharth, Anjari Gupta, Manjusha Chourasia, V.S. Prasad, Aditi, Kuldeep Sharma and Amit Ashish, 2012: Implementation of Unified Model based Analysis-Forecast System at NCMRWF, Technical report, NMRF/TR/2/2012.

FORECAST SENSITIVITY TO OBSERVATIONS IN ACCESS

Chris Tingwell, Sergei Soldatenko and Peter Steinle

Environment and Research Division, Bureau of Meteorology, Melbourne

chris.tingwell@bom.gov.au

Traditionally, the impact of an observation system, type, or individual instrument on the forecast skill of an NWP system is assessed by performing an Observing System Experiment (OSE). In an OSE, two parallel runs of the NWP system, only one of which assimilates the observation set of interest, are compared, and the change in forecast skill is attributed to the extra observations. The technique can be used to assess the impact both of newly available observations and of routinely assimilated observation sets. OSEs are informative but have disadvantages: they are computationally expensive and time-consuming to carry out, and their results are not especially fine-grained. They are not suited to evaluate the impact of, say, a single station in an observing network because of the very long duration of the experiment that would be required to produce statistically significant results.

Relatively new techniques, which make use of the adjoint models used in variational assimilation systems, are able to calculate the individual impact that each assimilated observation has in a cycling NWP system - the impact is typically measured by the reduction in the 24-hour forecast error (typically expressed as a total dry or moist energy norm). Such techniques were pioneered by Baker and Daley (2000) and Langland and Baker (2004) and then developed in a number of NWP centres. They have the advantage of being able to continually generate and aggregate forecast impacts for all observations, and allow much more fine-grained impact statistics to be generated than is feasible in OSEs. At the UK Met Office, Lorenc and Marriott (2014) developed a method of calculating Forecast Sensitivity to Observations (FSO) by means of the technology already used in the Met Office 4D-Var system: specifically the adjoint perturbation forecast model which is used in the minimisation of the 4D-Var cost function. The development of this capability in the Met Office gave rise to the opportunity to implement it in the ACCESS global NWP system.

The significant investment by the Bureau of Meteorology in the national observing network, and the constant evolution of observational technologies, require an ongoing assessment of the value of the network components. Numerical Weather Prediction (NWP) is one of the major mechanisms for converting observed data into information and services, so an objective measure of the impact of each observation on the quality of short-term ACCESS forecasts – provided by adjoint-based FSO – can potentially guide decisions related to network efficiency and effectiveness. FSO calculations can also inform the development and assessment of the ACCESS data assimilation system. These considerations have motivated a project, jointly sponsored by the Bureau's Environment & Research and Observations & Infrastructure divisions, whose goal is the implementation of a real-time ACCESS global FSO suite and the development of a set of tools to process, analyse and visualise the forecast sensitivity data produced by the suite in ways that are meaningful and informative to network planners and other stakeholders.

The ACCESS FSO suite is an adaptation of the Met Office global FSO suite, modified to run as a post-processor on the NCI *Raijin* machine from output generated by the Bureau's operational APS2 ACCESS-G suite. For each ACCESS-G assimilation cycle, the FSO suite produces a listing of the amount each assimilated observation changed the 24 hour forecast error as measured by a moist energy norm calculated over the entire global domain of ACCESS-G, from the surface up to a pressure level of 150 hPa. (A second suite employs an energy norm which is restricted to the Australian region.) These forecast impacts are processed into a set of JavaScript Object Notation (JSON) files,

from which a set of Python based tools can aggregate the individual forecast sensitivities on the basis of observation type or station and statistically analyse and visualise the results. A schematic of the FSO calculation is shown in Figure 1.

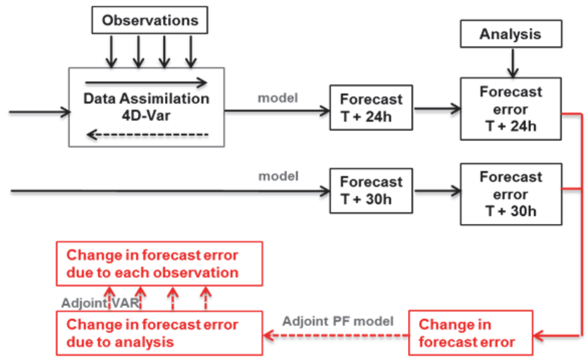


Figure 1: FSO calculation schematic. The change in T+24h forecast error caused by assimilating observations at T is integrated back in time via the VAR adjoint perturbation forecast model. The resulting forecast sensitivity to the analysis is then mapped back on to individual observations via the adjoint of the VAR assimilation.

The ACCESS-G FSO suite has now been running for over a year and we have completed several preliminary assessments of the forecast sensitivity of ACCESS-G to the various components of the Global Observing System, and to particular components of the Australian observing network. For example, Figure 2 shows the distribution of forecast impacts of radiosonde stations in the Australian network aggregated over the last 4 months of 2015. The most negative impacts (representing the largest *reductions* in forecast error) come from the most remote stations, whilst stations in the eastern states (where the observation network is densest) have comparatively small impacts. A number of other such observation impact comparisons will be presented in the talk. We will also discuss plans for further work.

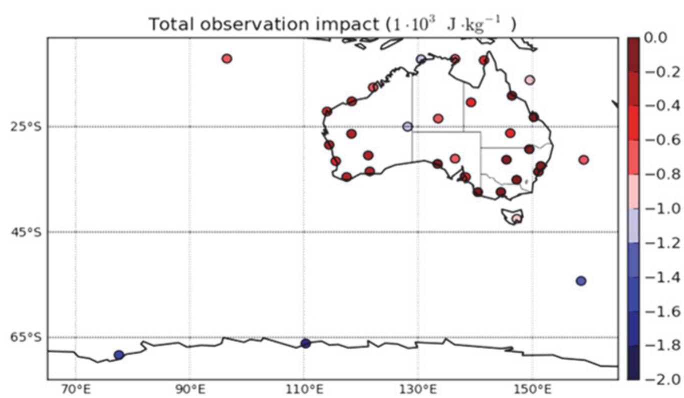


Figure 2: The distribution of Australian network radiosonde forecast impacts/day for the period September – December 2015. More negative impacts represent greater forecast error reduction.

References

Baker N.L. and R. Daley 2000: Observation and background adjoint sensitivity in the adaptive observation targeting problem. Quarterly Journal of the Royal Meteorological Society, 126, 1431-1454.

Langland R.H. and N.L. Baker 2004: Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. Tellus, 56A, 189-201.

Lorenc A.C. and R.T. Marriott 2014: Forecast sensitivity to observations in the Met Office global NWP system. Quarterly Journal of the Royal Meteorological Society, 140, 209-240.

CONVECTIVE-SCALE REANALYSIS FOR NEW ZEALAND

Stuart Moore

Meteorology and Remote Sensing, National Institute of Water and Atmospheric Research Ltd, Wellington, NZ

stuart.moore@niwa.co.nz

Due to the traditionally global domains of climate and earth simulation models and the coarse horizontal resolutions that their use necessitated, previous reanalysis work has focussed on representing and understanding the large-scale state of the past atmosphere and studying key climate variables. This translated also to projections of the future climate too, with coarse resolution models used to run through various climate change scenarios.

It is, however, the extreme weather events, such as convective cells and thunderstorms, frontal systems, cut off lows and downslope wind storms and their related impacts, such as flooding, that are perhaps more important when it comes to characterising the past climate and providing a baseline for understanding what might be different in the future under the auspices of climate change. Recent advances in both the science of Numerical Weather Prediction (NWP) and in the computational hardware available, has meant reanalysis projects have been initiated at higher horizontal resolutions. Some recent examples being the EURO4M (http://www.euro4m.eu/index.html) and German-led COSMO-REA6 and COSMO-RE2 (https://www.herz-tb4.uni-bonn.de/) projects. The ability to run NWP models at resolutions that explicitly resolve many of the atmospheric processes that occur at or near the scale of the landscape is thus an exciting opportunity to go beyond the large-scale atmospheric state and begin to put together a robust and physically self-consistent estimate of past, current and future weather extremes.

The New Zealand landscape, which varies from high snow-covered alpine mountains to lush temperate rain forests at sea-level and everything in between, often over very short distances, provides a stern challenge for NWP. NIWA has, since late 2013, been running a NWP model, called the New Zealand Convective-Scale Model (NZCSM) with a horizontal resolution of 1.5 km for the last three years generating forecasts four times daily. This model is a local configuration of the Met Office Unified Model (Davies et al. 2005). In this time, New Zealand has been subject to a number of extreme weather events, for example ex-TC Pam and ex TC Ita, and shown itself to be more than capable of forecasting well such extreme events. Indeed, while some biases still exist, it has proven a valuable improvement over NIWA's mesoscale 12 km resolution New Zealand Limited Area Model (NZLAM). Figure 7 compares the total annual rainfall as predicted by NZLAM and NZCSM against NIWA's observation-based Virtual Climate Station Network (Tait et al. 2006) product. From Figure 7 it is clear how NZCSM, compared to NZLAM, has captured the higher rainfall totals around the high altitude areas of the North Island and has increased spatial accuracy along the Southern Alps of the South Island.

Confident that NZCSM is capable of simulating the weather over New Zealand's complex landscape, NIWA plans to conduct a number of experiments with the aim of characterising the historical extreme weather New Zealand experienced in the years since 1979 through to the present day and into the future. This information will enable New Zealand users of the climate system to make informed decisions as to how best to prepare for and mitigate the impacts of climate change and the nature of future extreme weather hazards. Making use of the latest version of the Unified Model, featuring the ENDGame dynamical core (Wood et al. 2014), the proposed experiments are;

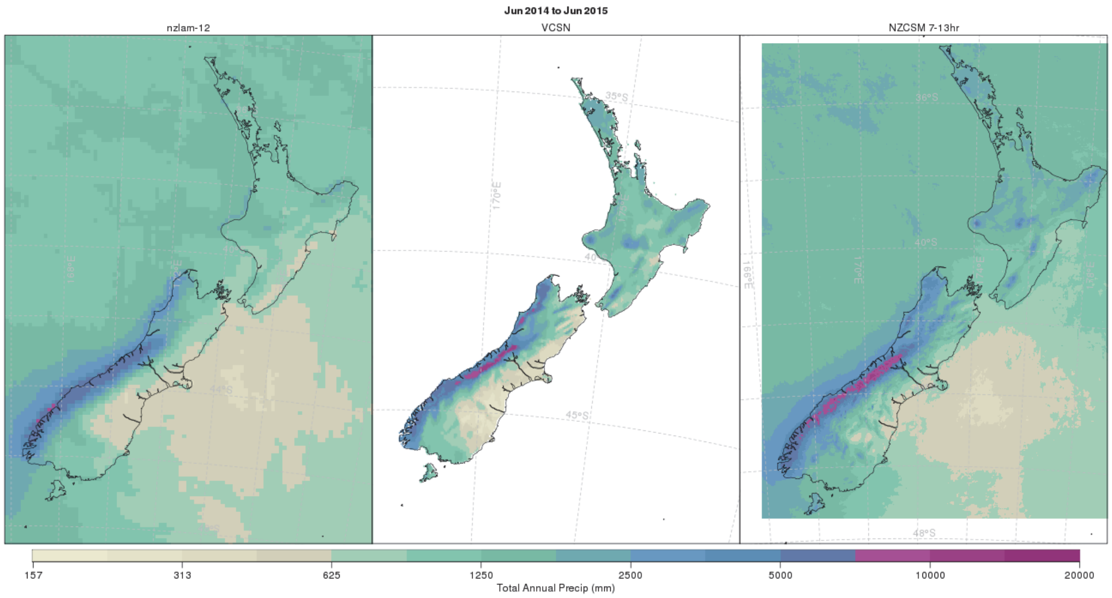


Figure 7: Comparison of NZLAM (left panel), VCSN (centre panel) and NZCSM (right panel) total annual rainfall for the period June 2014 to June 2015.

Past Climate

Perform a full data-assimilating reanalysis using ERA-Interim (Dee et al. 2011) data, in collaboration with the Bureau of Meteorology, to 12 km horizontal resolution and then downscale to 1.5 km over the full NZCSM domain for the period January 1979 to the present day.

Current Climate

Starting from ERA-Interim analyses, perform a numerical downscaling for the period January 2014 to December 2016 down to the horizontal resolution of NZCSM to simulate the current climate and evaluate the performance of this model against the current NZCSM forecast archive.

Future Climate

Downscale, using the same updated version of NZCSM, output from the Deep South Earth System Model (http://www.deepsouthchallenge.co.nz/programmes/earth-system-modelling-and-prediction) simulations of future climate to the convective-scale to generate a multi-decadal dataset of late 21st century weather and compare it against the "past climate" dataset to understand how weather extremes may change in New Zealand's future.

For each climatic period, output from the model runs will be used as input to downstream hydrological, inundation and sea state models to simulate and characterise the nature of weather-related impacts of all extreme weather events.

In this talk, the challenges of convective-scale modelling over New Zealand will be discussed and the proposed experiments further detailed with some initial results discussed.

References

Davies T, Cullen M, Malcolm A, Mawson M, Staniforth A, White A, Wood N. 2005. A new dynamical core for the Met Office's global and regional modelling of the atmosphere. Q.J.R. Meteorol. Soc. 131, 1759–1782.

D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut and F. Vitart, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteorol. Soc. 137, 553-597.

Tait A, Henderson R, Turner R, Zheng XG, 2006: Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. International Journal of Climatology 26(14), 2097–2115.

Wood N, Staniforth A, White A, Allen T, Diamantakis M, Gross M, Melvin T, Smith C, Vosper S, Zerroukat M, Thuburn J. 2014. An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. Q.J.R. Meteorol. Soc. 140, 1505-1520.

TOWARDS A HIGH-RESOLUTION ATMOSPHERIC REANALYSIS FOR AUSTRALIA

Chun-Hsu Su, Greg Kociuba, Nathan Eizenberg, Peter Steinle, Dörte Jakob and Paul Fox-Hughes

Bureau of Meteorology, Melbourne

chunhsu.su@bom.gov.au

A consistent, long-term atmospheric reanalysis can support high quality assessments of environmental risk and likelihood of extreme events. At present, most reanalyses are based on global systems that are generally too coarse (~100 km) to be suitable for regional-scale environmental assessments. The Bureau of Meteorology is therefore undertaking the early tranches of a higher resolution reanalysis (1.5 km to 12 km) over Australia and New Zealand to produce a sequence of credible three-dimensional atmospheric conditions at hourly intervals for a period covering approximately the last 25 years. This project is enabled by funding from three emergency management agencies in Australia, namely the New South Wales Rural Fire Service (NSW RFS), the Tasmanian Department of Police and Emergency Management, administered through University of Tasmania (UTAS), and the Department of Fire and Emergency Service in Western Australia. This provides a focus on producing value-added products to build a climatology of bushfire-related parameters such as soil dryness, temperature and atmospheric stability. Other applications of the reanalysis are expected for water and land use management, primary industries, and the health, energy and mineral sectors. The reanalysis will also benefit weather research by providing initial conditions and verification datasets for case studies, and defining climatology and ranks of anomalies.

The reanalysis is being run with a fixed numerical weather modelling suite at a sub-daily temporal resolution and assimilates available observations using a consistent data assimilation scheme over a long period in the past. The modelling suite is based on the European FP7 reanalysis project UERRA and is closely related to the numerical weather prediction (NWP) model implemented in the Bureau's operational NWP (i.e., ACCESS) systems. This work is carried out within the UM Partnership framework, in close collaboration and in concert with UKMO, NCMRWF, NIWA and KMA in their own regional reanalyses over Europe, India, New Zealand and East Asia, respectively.

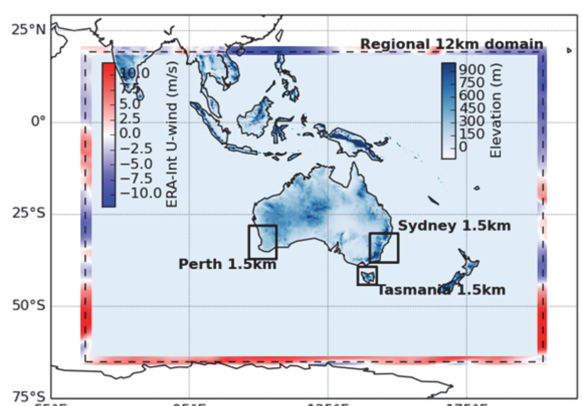


Figure 1: Nested reanalysis setup, consisting of a regional 12 km reanalysis area and 1.5 km sub-domains. The former is bounded by ERA-Interim reanalysis data (1 February 2015 depicted).

The Bureau's reanalysis suite uses a four-dimensional variational (4DVAR) data assimilation scheme to optimally combine observations (surface observations, sondes, buoys, satellites, tropical cyclone track) and short model forecasts to provide a 12 km reanalysis over the Australia-New Zealand region

(Figure 1). The 12 km model is bounded by the coarse-scale ERA-Interim global reanalysis (Dee et al. 2011) that provides the required boundary and initial conditions. This analysis is in turn used to drive convective-scale (1.5 km) downscaling models over smaller selected areas. In particular, NSW RFS funds the project to downscale the 12 km analysis over the eastern NSW domain, while UTAS supports a 1.5 km resolution sub-domain analysis over Tasmania. This combination of models has been shown to improve estimates of sub-daily rainfall in the ACCESS systems (Bureau of Meteorology 2013a; 2013b). The temporal resolution of the gridded analysis fields for both the regional and higher resolution domains will generally be one hour, with some fields such as 10 m winds and 2 m temperatures and dew points available every 10 minutes. The reanalysis also produces a large set of variables that include moisture, cloud cover, precipitation, evaporation, soil water, and energy fluxes, which represents a superset of the standard ACCESS NWP model outputs. Importantly, the project will generate required data for nesting other higher-scale NWP models anywhere within the regional 12 km domain.

The reanalysis suites, both regional and sub-domain, have been set up to use the recent versions of the NWP components for observation processing, 4DVAR and atmospheric model runs and land-surface analysis. First results are expected in late 2016.

Initial benchmarking results

In the first stage of this project, an initial, benchmarking 1.5 km data set, referred to as Initial Analysis (IA), has been constructed over the NSW and Tasmania domains, using the 2011-2015 archived operational ACCESS city model data sets. This involves harmonising two data sets with different native resolutions (6 and 4 km) and generated by different model versions and ancillaries. The IA is a prototype for the future reanalysis product. It allows developing suitable quality assurance processes and downstream products, namely climatology, soil dryness and fire danger indices and return period statistics, which immediately benefit funding agencies.

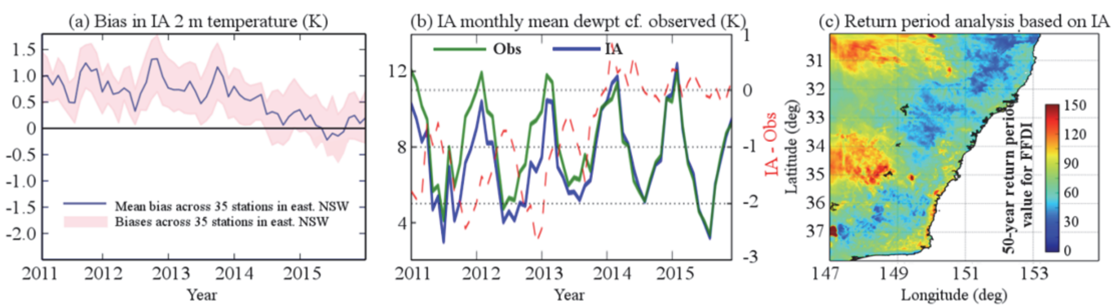


Figure 2: (a) Evaluation of IA 2 m temperature against surface observations over eastern NSW. (b) A comparison between IA and observed dewpoint climatology at a Tasmanian station. (c) illustrates the use of multi-year data to map return period estimates of FFDI across eastern NSW.

Our evaluation of the IA against surface observations from over 60 automated weather stations in eastern NSW and Tasmania identified biases in modelled screen-level temperature (Figure 2a for NSW evaluation), dewpoint and wind speed, which change over time, particularly at the cutover between two operational ACCESS-C systems. There are strong correlations between IA and AWS data. The limited skill in topographically complex areas is attributed to interpolation from coarse-scale model data to the 1.5 km scale and AWS locations. The model does not fully capture long dry spell periods, and has a tendency to show fewer dry days and more wet days. These factors all have noticeable impacts on the skill of the IA-derived McArthur forest fire danger index (FFDI). Additional variability presented in the IA-derived climatology (Figure 2b), which is defined loosely here as monthly statistics, is due to the NWP system changeover. Nevertheless, several of the modelled climatologies – based on 2 m temperature, 2 m dewpoint, 10 m wind, precipitation, and FFDI - show high statistical similarity, small overall bias, and/or good temporal correlation with the observed

climatology. These problems are expected to be partially addressed by the reanalysis, which uses a 1.5 km grid and consistent model setups across the analysis period. Figure 2c shows return periods of FFDI derived via extreme value modelling based on the IA.

References

Bureau of Meteorology (BOM), 2013a: APS1 upgrade of the ACCESS-R Numerical Weather Prediction system Op. Bull. 98, http://www.bom.gov.au/australia/charts/bulletins/nmoc_bulletin.shtml

BOM, 2013b: APS1 upgrade of the ACCESS-C Numerical Weather Prediction system (Aug 2013) Op. Bull. 99, http://www.bom.gov.au/australia/charts/bulletins/nmoc_bulletin.shtml

Dee, D.P., et al., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quart. J. Royal Meteorol. Soc., 137 (656 – Part A), 553-597.

Jermey, P., 2015: UERRA Ensemble Data Assimilation Configuration, Forecasting research technical report 597, www.metoffice.gov.uk/media/pdf/8/a/FRTR597b.pdf

ENSEMBLE REGIONAL REANALYSIS OVER EUROPE

Peter Jermey, Richard Renshaw

Met Office, Exeter, UK

Peter.Jermey@metoffice.gov.uk

Model-based reanalyses attempt to reconstruct past climate and specific historical weather events through an 'optimal' merging of past observations and modern NWP via the process of data assimilation (DA). They provide complete gridded analyses of observed variables (e.g. winds, temperatures, precipitation, etc.) as well as those not observed (fluxes, 3D cloud, etc.) through the complex, multivariate, dynamically and physically consistent processes represented within modern NWP and DA systems.

Global reanalyses have found use in research, for climate and modelling studies, and in the commercial world. Regional reanalysis has similar potential. Data is though of little use unless it is known how reliable that data is. This requirement shapes one of the goals of the EU FP7-funded project UERRA – Uncertainties in Ensembles of Regional Re-Analyses (http://uerra.eu). UERRA aims to produce high-quality climate information over Europe, using ensembles of reanalyses to quantify the uncertainties and errors in the products. It is expected that estimates of uncertainy will give added value in showing the impact of changes in the observing network over time.

The Met Office contribution to UERRA is to produce both an ensemble and a deterministic reanalysis to cover the satellite-era (1979-present). Both use the EU-CORDEX domain (Figure 1), the deterministic at 12km resolution and the 20-member ensemble at 36km resolution. Lateral boundary conditions are taken from the ERA-Interim global reanalysis.

The deterministic reanalysis aims to provide a high-quality reanalysis dataset. The ensemble reanalysis provides information on the reanalysis uncertainty (through the ensemble spread) and is also used directly as input to the data assimilation for the deterministic reanalysis. This workshop presentation describes the data assimilation system used for both, and shows early results.

UERRA Ensemble

This is a 20-member ensemble. The model (UM) runs at 36-km resolution, and the data assimilation system is 4DVAR with a 6-hour cycle, using a linearised forecast model at 72km resolution. Other Met Office ensemble systems such as MOGREPS use calibration to maintain ensemble spread, for instance applying additive inflation at every cycle. These calibrations are themselves indirectly dependent on the density of the observation network. One aim of UERRA is to show how reanalysis uncertainty changes as the observation network changes. The approach taken here is to apply realistic perturbations to those inputs where possible, and to make no further intervention to the ensemble spread. Each member runs independently of the others, with no recalibration of ensemble spread in the data assimilation process.

Ensemble spread is maintained by using perturbated observations, SST and sea-ice analyses and, importantly, through adding perturbations to mimic model error.

Piccolo and Cullen (2016) show that, given certain assumptions, analysis increments from a 4DVAR system will have a similar distribution to the true model error. Therefore different realisations of the model can be produced by perturbing the model with a randomly selected analysis increment. One of

the assumptions is that the analysis increments derive from an extensive observation network. For UERRA these analysis increments are taken from a reanalysis of 2015 specifically for this purpose, using the full range of observations that are used in Met Office operational NWP.

UERRA Deterministic

The deterministic will use hybrid 4DVAR, which couples the deterministic system to the ensemble system. In hybrid 4DVAR the background error covariance is effectively a linear combination of the specified climatological covarianes and covariances from the ensemble of reanalyses from the same cycle. In this way the ensemble will provide the deterministic system with error characteristics dependent on the synoptic situation. Hybrid 4DVAR has been used operationally for global data assimilation at the Met Office since July 2011 (Clayton et al., 2013). The regional variant will be used for the first time with the deterministic reanalysis and is still under development.

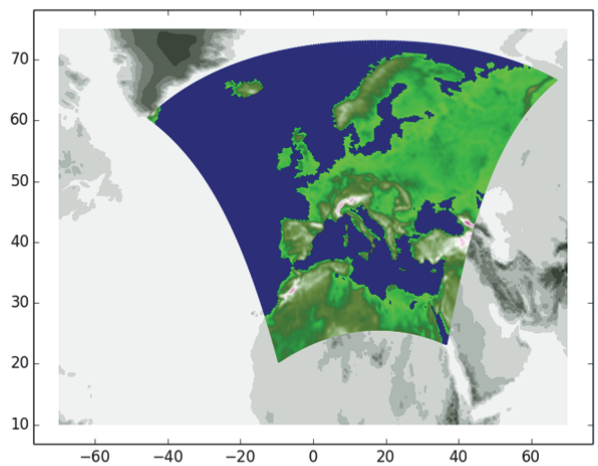


Figure 8UERRA (EU-CORDEX) domain highlighted. Colours are model orography.

The research leading to these results has received funding from the European Union, Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 607193

References

Clayton, A. M., Lorenc, A. C., and Barker, D. M., 2013: Operational implementation of a hybrid ensemble/4d-var global data assimilation system at the Met Office. Q.J.R. Meteorol. Soc, 139(675):1445–1461

Piccolo, C. and Cullen, M., 2016: Ensemble data assimilation using a unified representation of model error. Mon. Wea. Rev. 144(1), 213-224

4DVAR OPTIMIZATION & USE-CASES FOR DEEP LEARNING IN EARTH SCIENCES

Phil Brown

Segment Leader, Earth Sciences, Cray, United Kingdom

philipb@cray.com

This talk will present a update on Cray's activities in the Earth sciences and summarize some of the ongoing optimisation of the Unified Model 4DVAR we have been undertaking in partnership with the Met Office. I will then briefly introduce an emerging data analysis technology and discuss some potential use-cases in remote sensing, data assimilation and numerical weather prediction fields.

A PERFORMANCE EXPLORATION OF 4D-VAR AT HIGH RESOLUTION

Dale Roberts and Ben Evans

National Computational Infrastructure, Australian National University, Canberra

ds.roberts@anu.edu.au

An overview of the performance of 4D-VAR is presented at the N320 resolution, using a variety of software stacks and profiling applications. We first assess the effect of different MPI distributions on the performance, and find that Intel MPI outperforms OpenMPI at small decompositions. However, at larger decompositions, the situation is reversed and OpenMPI outperforms Intel MPI by as much almost 20% at 960 cores. Profiling with IPM and Score-P indicates that, at these large decompositions, OpenMPI is significantly outperforming Intel MPI in the repeated small-buffer collective communication operations required during the elliptical solver. Profiling suites reveals that there are few ‘hotspots’ in the application, that is, small sections of code responsible for a significant amount of runtime. Two general areas identified by profiling that could be improved within 4D-VAR are cache tiling and OpenMP coverage. Given the complex nature of the application, this would require a significant development effort, with overall gains in application performance coming slowly as cache access and threading is improved. Beyond this, however, are two areas that could benefit from new approaches to MPI communication, namely the metadata exchange during field interpolation, and the row-wise reductions over the polar extremities.

In performing the field interpolation in the SISL integration scheme, an MPI process may have data in memory that will be required for a different region of the grid, and must be sent to a different MPI process. This is complicated by the fact that the MPI process that requires the data does not know in advance how much it will need or which process it will come from. In the current version of the application, this is resolved by having each process determine how much data it needs to send to every other process, effectively creating one row of a send-receive count matrix. Once this is determined, this matrix is transposed among all MPI ranks using an expensive ‘All to all’ MPI operation, thus giving each rank the metadata required to initiate the actual data transfer.

At the N320 resolution on 960 cores, this one step accounts for almost half of the cost of the interpolation. We believe that it is possible to replace this ‘All to all’ operation with the ‘Non-blocking consensus’ algorithm[1]. Using asynchronous features of the MPI standard, every MPI process is able to signal that data they have sent has been received whilst simultaneously receiving data from other ranks. Once all ranks reach a consensus that their data has been received by their intended target, program execution can continue. With this algorithm, the metadata is encoded with the message in the form of the message size and message sender, thus eliminating the need for a separate, and costly, metadata exchange step.

Load imbalance in polar regions of 4D-VAR has also been identified as a key performance issue, with the majority of MPI processes in any given domain decomposition needing to wait for additional row-wise reductions to be performed at the polar extremities throughout the execution time of the model. In some cases, it may be possible to alleviate the cost of this operation by first precomputing the quantities that need to be reduced at the polar extremities before they are calculated for the remainder of the domain, then performing the reductions asynchronously whilst calculations are performed on the remainder of the domain. The key area where application of this approach may prove most beneficial is in the evaluation of the elliptic operator, in which the reductions at the polar extremities account for over 5% of the total model runtime. Once applied, this may be able to reduce the penalty

when running the model at more ‘square’ domain decompositions, which can reduce the size of the subgrid halos, and thus the data required to be transmitted between MPI processes in halo exchange operations.

References

Hoefler, T., Siebert, C. and Lumsdaine, A., 2010. Scalable communication protocols for dynamic sparse data exchange. *ACM Sigplan Notices*, *45*(5), pp.159-168.