School of Mathematical
and Physical Sciences

University of Reading

Department of Mathematics and Statistics

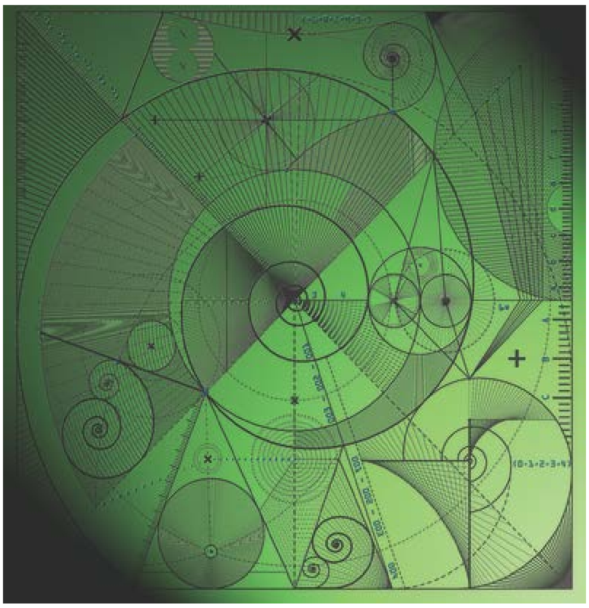
Preprint MPS-2015-10

3 July 2015

Application of data assimilation to ocean and climate prediction

by

Michael J. Bell, Matthew J. Martin
and **Nancy K. Nichols**



Application of data assimilation to ocean and climate prediction

Michael J Bell, Matthew J Martin and Nancy K Nichols

Abstract Ocean prediction systems are now able to analyse and predict temperature, salinity and velocity structures within the ocean by assimilating measurements of the ocean's temperature and salinity into physically based ocean models. Data assimilation combines current estimates of state variables, such as temperature and salinity, from a computational model with measurements of the ocean and atmosphere in order to improve forecasts and reduce uncertainty in the forecast accuracy. Data assimilation generally works well with ocean models away from the equator but has been found to induce vigorous and unrealistic overturning circulations near the equator. A pressure correction method was developed at the University of Reading and the Met Office to control these circulations using ideas from control theory and an understanding of equatorial dynamics. The method has been used for the last 10 years in seasonal forecasting and ocean prediction systems at the Met Office and European Center for Medium-range Weather Forecasting (ECMWF). It has been an important element in recent re-analyses of the ocean heat uptake that mitigates climate change.

Michael J Bell
Met Office, Fitzroy Road, Exeter, EX1 3PB UK e-mail: mike.bell@metoffice.gov.uk

Matthew J Martin
Met Office, Fitzroy Road, Exeter, EX1 3PB UK e-mail: matthew.martin@metoffice.gov.uk

Nancy K Nichols
University of Reading, School of Mathematical and Physical Sciences, PO Box 220, Whiteknights, Reading, RG6 6AX UK e-mail: n.k.nichols@reading.ac.uk

[1] **To be published in:** *UK Success Stories in Industrial Mathematics*, (eds P. Aston, A. Mulholland, and K. Tant), Springer International Publishing AG, 2016.

1 Introduction

Variations in ocean heat content account for much of the energy that drives weather and climate systems; errors in the representation of the ocean dynamics and thermodynamics in computational models affect the validity of forecasts of the ocean and atmosphere on daily, seasonal and decadal time scales. To improve the forecasts, a state estimation technique from control theory, known as 'data assimilation' [4], is used to combine current estimates of state variables such as temperature, salinity and velocity from a numerical model with observations of the ocean. The data assimilation process generally works well with ocean models away from the equator, but spurious ocean circulation occurs when thermal data are assimilated near the equator into models with systematic errors. Research undertaken by the University of Reading and the Met Office investigated these errors and developed a novel pressure correction technique that improves the analyses and forecasts by restoring dynamical balances, eliminating spurious deep overturning circulations [2]. The technique has been implemented by the Met Office and by the European Centre for Medium Range Weather Forecasting (ECMWF) in their forecasting systems, resulting in major improvements to ocean forecasting and seasonal prediction using coupled ocean-atmosphere models.

2 Data Assimilation

A computational model can never completely describe the complex physical processes involved in the behaviour of a real world dynamical system. Data assimilation techniques are used to improve predictions from the numerical models. Mathematically, the assimilation problem is an inverse problem, matching model data to observations [4]. To illustrate the problem we assume we have a model of the evolution of the height $h(x,y,t)$ of the ocean surface and that at a given time t we have observations of the surface height at N locations (x_n, y_n), $n = 1,2,\ldots,N,$ which we write in a vector $\mathbf{h^o}$ of dimension N. We suppose also that we have a prior forecast (or background) vector of surface heights at these same points, $\mathbf{h^b}$, generated by the model dynamics. An improved estimate of the heights is then obtained by minimizing the differences between the measurements and the prior model forecast states, weighted by forecast and measurement error statistics. Statistically, the Best Linear Unbiased Estimate (BLUE) of $\mathbf{h}$, is obtained by minimizing the objective function

$$J(\mathbf{h}) = (\mathbf{h}-\mathbf{h^o})^{\mathbf{T}}\mathbf{R^{-1}}(\mathbf{h}-\mathbf{h^o}) + (\mathbf{h}-\mathbf{h^b})^{\mathbf{T}}\mathbf{B^{-1}}(\mathbf{h}-\mathbf{h^b}). \tag{1}$$

with respect to $\mathbf{h}$, where $\mathbf{R}$ and $\mathbf{B}$ are square, symmetric positive-definite matrices representing the covariances of the errors expected in the observations and in the forecast field, respectively. The errors in the observations arise from instrument error, but also from the fact that the observation may measure small scale phenomena (such as a wave crest of small wavelength) that the model cannot represent or predict. It is because of nonlinear transfers between scales within fluids that ocean and weather predictions have an inherently stochastic nature.

The updated vector of heights from the minimization is used to initialize and evolve the numerical model forward to obtain a forecast at the next time that observations are available; the minimization process is then repeated. In practice, if only the surface heights are updated, poor forecasts may be obtained. Ideally we would also update the horizontal velocities u and v that describe the evolution of the flow. The geostrophic balance relationship between the forces per unit mass due to gradients in the surface height fields and the Coriolis acceleration can be used for example, away from the equator, to update the velocities and provide better initial conditions for predictions. Alternatively a dynamical model, such as the shallow water equations, can be used to forecast and assimilate the heights and the horizontal velocities simultaneously.

In practice we do not have as many observations as we have state variables in our model equations and the locations of the observations are not necessarily at the same locations in space as our forecast model states. In this case we map the model states to variables that can be compared directly to the observations, using an 'observation operator'; this could be an interpolation operator, for example. We then minimize the differences between the measurements and the observation operator applied to the vector of model states, together with the differences between the prior forecast (background) and the model states, weighted by the measurement and forecast error statistics respectively. This leads to a nonlinear least-squares problem with an objective function similar to (1); the minimum of the objective function provides an improved estimate of all the forecast states. The problem is well-posed even if the number of observations is less than the number of state variables. Uniqueness of the solution is ensured because the covariance matrix of the errors in the prior forecast (background) field is nonsingular (see [4]).

3 Unrealistic motions generated by data assimilation near the equator

In addition to improving forecasts from the model, data assimilation can be used to identify systematic errors within the model, highlighting where the model forecasts are consistently incorrect in relation to observations. An analysis of early results obtained from a system assimilating measurements of thermal profiles into a global ocean model revealed that in the eastern equatorial Pacific, where there was a particularly good source of data from the Tropical Atmosphere-Ocean (TAO) mooring array, the data assimilation was continuously making large temperature updates of the same sign. Heating by these assimilation updates was generating "equatorial bonfires" involving large unrealistic vertical velocities and overturning circulations extending to considerable depths (see Figure 1a).

To understand this behaviour we first note that near the equator the Coriolis acceleration is small and flows along the equator are not in geostrophic balance. In the ocean to a first approximation the momentum balance along the equator in the near surface layers is between momentum input by the surface wind stress, which is mixed downward by turbulent motions, and the pressure gradient along the equator. The surface wind stresses and turbulent mixing at the equator are not accurately known and measurements of salinity (which also affects the density and hence the pressures within the ocean) are also limited (although they are much more abundant since the Argo array of profiling floats reached maturity in about 2004). As an ocean model is integrated forward in time its density field adjusts until the forces due to the winds and the pressure gradients come into "balance".

Because the surface wind stresses and vertical mixing may not be accurate, the mass fields are likely to be inaccurate. Hence measurements of the ocean temperatures will differ from those of the model and the assimilation system will make updates to the temperatures proportional to the difference. Heating (or cooling) below the surface near the equator will set off incorrect convective circulations, with the convection tending to oppose and reduce the heating or cooling and to return the model towards the state in which the wind stress and pressure gradients are in balance. Consequently the temperature of the model at a given location retains a bias relative to the measurements and the assimilation scheme can make very large mean temperature updates over time (see Figure 2a).

Given this physical picture of the origin of the problem, a number of solutions to it could have been explored. The approach we chose was to use the data assimilation updates to calculate an additional pressure field. This pressure field accumulates with time so that it reflects the mean updates over time made by the assimilation scheme and its gradients contribute to the momentum equations in the opposite direction to the pressure gradients generated by the temperature updates themselves.

This approach appeared to be ad hoc and we were concerned about whether it would be unstable (blow-up over time) or fail to converge to an acceptable mean state. It was very helpful to formulate and analyse the problem mathematically using control theory and a shallow water model in a way that gave reassurance that the method would be stable and converge to a good estimate of the mean state.

4 Analysis using the shallow water equations

We use a simplified model (compared to the full ocean general circulation model used in practice) in order to assess the proposed pressure correction method analytically. The response near the equator to thermal forcing was analysed in [3] and the vertical structure of the solutions was found to be separable from the horizontal and temporal structure. The latter is determined by the linearised shallow water equations (SWEs). This allows the assimilation of thermal profile data to be analysed in terms of the assimilation of surface height data within the SWEs.

We let the true fields be denoted by a superscript "t" and suppose that they evolve according to the SWEs when driven by the true surface winds τ_x^t and τ_y^t so that

$$\frac{\partial u^t}{\partial t} - fv^t = -g\frac{\partial h^t}{\partial x} + \tau^t_x, \tag{2}$$

$$\frac{\partial v^t}{\partial t} + fu^t = -g\frac{\partial h^t}{\partial y} + \tau^t_y, \tag{3}$$

$$\frac{\partial h^t}{\partial t} + H_e\left(\frac{\partial u^t}{\partial x} + \frac{\partial v^t}{\partial y}\right) = 0, \tag{4}$$

in which h is the ocean surface height, u, v are depth averaged horizontal velocities, g is gravitational acceleration, H_e is the equivalent depth of the fluid (which is the constant of separation) and f is the Coriolis parameter (which is proportional to the distance from the equator).

A simple representation of the assimilation scheme proposed in Section 3 is given by

$$\frac{\partial u^b}{\partial t} - fv^b = -g\frac{\partial}{\partial x}(h^b + \eta^b) + \tau^b_x, \tag{5}$$

$$\frac{\partial v^b}{\partial t} + fu^b = -g\frac{\partial}{\partial y}(h^b + \eta^b) + \tau^b_y, \tag{6}$$

$$\frac{\partial h^b}{\partial t} + \varepsilon(h^b - h^t) + H_e\left(\frac{\partial u^b}{\partial x} + \frac{\partial v^b}{\partial y}\right) = 0, \tag{7}$$

$$\frac{\partial \eta^b}{\partial t} - \gamma(h^b - h^t) = 0. \tag{8}$$

In these equations the superscript b represents the model forecast fields and:

1. the momentum equations of the model are the same as the true momentum equations except that the model is driven by model winds rather than the true winds and the pressure field gh^b has been augmented by the pressure correction field $g\eta^b$;
2. data assimilation has been represented by the second term on the left-hand side of (7), which acts as a feedback control mechanism and forces the model field towards the true field; and
3. the assimilation updates are also added to the pressure correction field in (8). Here γ satisfies $0 < \gamma < \varepsilon$, so the updates are smaller than and in the opposite direction to those to the height field h^b.

Denoting the difference between the model and the true fields by superscript 'e', and subtracting (2) - (4) from (5) - (7) we obtain

$$\frac{\partial u^e}{\partial t} - fv^e = -g\frac{\partial}{\partial x}(h^e + \eta^e) + \tau^e_x, \tag{9}$$

$$\frac{\partial v^e}{\partial t} + fu^e = -g\frac{\partial}{\partial y}(h^e + \eta^e) + \tau^e_y, \tag{10}$$

$$\frac{\partial h^e}{\partial t} + \varepsilon h^e + H_e\left(\frac{\partial u^e}{\partial x} + \frac{\partial v^e}{\partial y}\right) = 0, \tag{11}$$

$$\frac{\partial \eta^e}{\partial t} - \gamma h^e = 0. \tag{12}$$

As $\gamma \neq 0$, (12) implies that any stationary or time-mean solution has $h^e = 0$, which means that the model surface height is the same as the true surface height and the time-mean data assimilation updates will be zero. Using $h^e = 0$ in (11) we see also that the horizontal divergence $\partial u^e/\partial x + \partial v^e/\partial y$ is zero for these solutions. For the 3D problem this implies that the vertical velocities will also be zero. It is also possible to show that there are no spurious sources of energy in the augmented equation set (5) - (8) and that they do not support spurious wave amplification or propagation.

Initial integrations showed that the scheme reduced the time-mean temperature increments and improved the equatorial undercurrent. The method was implemented in the Met Office ocean assimilation system and shown to have a significant impact on the forecast [2]. In Figure 1a we see the substantial overturning deep ocean vertical velocities induced by the assimilation and in Figure 2a we see the mean temperature updates from the assimilation that generate this flow. In Figures 1b and 2b,

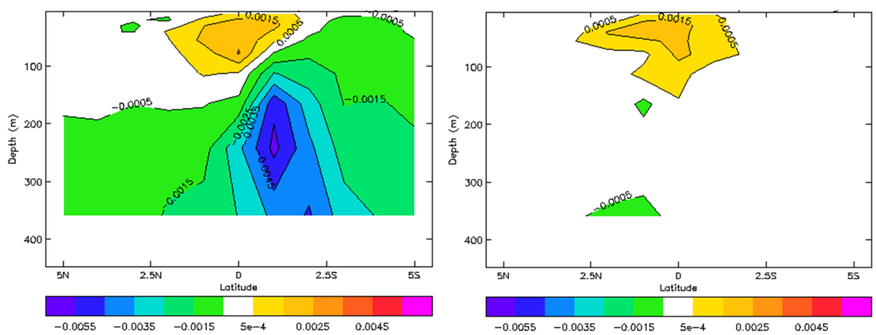


Fig. 1 Annual mean vertical velocities (cm s^{-1}) cross-section across the equator at 110°W: (a) with standard data assimilation and (b) with the pressure correction scheme.

we see that the pressure correction technique provides a major improvement in the assimilation, eliminating the spurious circulation and reducing the temperature updates.

5 Impacts

Ocean data assimilation systems are used to initialise predictions of the ocean and climate from days to decades ahead, as well as providing information about the past through reanalyses. These systems use a vast quantity of input data collected from satellites, surface drifting buoys, sub-surface profiling floats from the Argo array, moored buoys, ships and various other observing platforms. Improvements to data

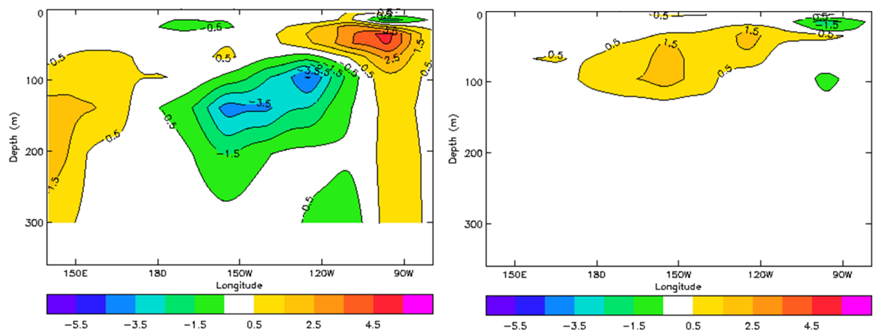


Fig. 2 Annual mean potential temperature increments (°C month^{-1}) cross-section along equator between 140°E-90°W: (a) with standard data assimilation and (b) with the pressure correction scheme.

assimilation techniques, such as those described here, enable better use of these expensively acquired data to give more accurate predictions and reanalyses.

Accurate seasonal and decadal forecasts are particularly important for understanding the effects of climate change and in developing strategies for living with changes in our environment as well as for mitigating hazardous conditions that may arise, such as flooding, drought, intense rainfall, heavy snow and ice, or excessive temperatures. Increased accuracy of ocean, weather and climate forecasting has impacts on economic, commercial and organisational elements of society as well as on our understanding of the environment.

The development of the pressure correction method by the University of Reading and the Met Office, and the improvement in accuracy it has allowed, has made an important contribution to ocean data assimilation. The technique forms an integral part of the Met Office ocean data assimilation system. It has also been extended by ECMWF and incorporated into their ocean-atmosphere assimilation system for seasonal forecasting and plays an important role in their system for re-analysis of ocean heat uptake [1], a major factor affecting climate change.

Both the Met Office and ECMWF have now transitioned to a new core ocean model, NEMO (Nucleus for European Modelling of the Oceans - http://www.nemo-ocean.eu/), which is a community ocean modelling system for oceanographic research, operational oceanography, seasonal forecasting and climate studies. The assimilation systems based on this model are used for producing short-range forecasts of the ocean and sea-ice state (out to 7 days), and are also used to initialise the ocean component of seasonal forecasts. New coupled ocean-atmosphere assimilation systems are also under development for initialising future coupled weather and climate forecasting systems. The pressure correction scheme remains an important component of these new systems and our research on data assimilation continues to bring significant benefits to the community.

References

1. Balmaseda, M.A., Trenberth, K.E. and Kallen, E.:Distinctive climate signals in reanalyses of global ocean heat content. Geophys. Res. Lett., **40**, 1754–1759 (2013).
2. Bell, M.J., Martin, M.J., Nichols, N.K.: Assimilation of data into an ocean model with systematic errors near the equator. Quart. J. Roy. Meteorol. Soc., **130**, 873–894 (2004)
3. Gill, A.E.: Atmosphere-Ocean Dynamics, Academic Press, San Diego, London (1982)
4. Nichols, N.K.: Data Assimilation: Aims and Basic Concepts. In: Data Assimilation for the Earth System, (eds R. Swinbank, V. Shutyaev and W.A. Lahoz) Kluwer-Academic Publishers, Dordrecht (2003).