Maths@work in Meteorology

Jeffrey D. Kepert*

The past few decades have seen a steady and dramatic improvement in our ability to forecast the weather. The Australian Bureau of Meteorology today issues seven-day forecasts for maximum temperature that are of similar accuracy to the four-day forecasts of 20 years ago, while modern four-day temperature forecasts are of comparable accuracy to the earlier one-day forecasts. Similar improvement is found for almost all forecasts — for example, the US National Hurricane Center extended the range of its tropical cyclone track forecasts to five days in 2003, and the verification skill scores show that these are as good as the two-day forecasts of the early 1980s.

Many would be aware, at least anecdotally, of these improvements, and could name improvements in satellites and computers as having played a large role. Using the ever-growing computer power to make the best use of the data from the satellites and other sources to produce forecasts in a timely manner remains a substantial problem, and one where mathematics plays a large role. The aim of this issue of maths@work is to outline one of the many areas in which mathematics is essential to modern meteorology, namely data assimilation.

The improvement in computer guidance, or numerical weather prediction (NWP), is illustrated in Figure 1, which shows the time series of the anomaly correlation[1] scores of three-, five- and seven-day forecasts at 500 hPa (about 5.5 km altitude) from the European Centre for Medium-Range Weather Forecasting (ECMWF) for the Northern and Southern Hemispheres. A significant part of the recent forecast improvements is due to improvements in the analysis algorithms, and particularly the treatment of satellite data. The growing utility of satellite data is illustrated by the elimination of the NWP skill difference between hemispheres in recent years — the Northern Hemisphere, with larger land masses and population, relies more heavily on in situ data from ground stations, balloons and aircraft, while the Southern Hemisphere contains much larger data-void oceans, so improvements in satellite data and its analysis have had a larger impact.

The numerical weather prediction problem falls into two parts. The prediction side of the problem is essentially computational fluid dynamics. Regional-scale models forecast out to 2–3 days and usually employ a finite-difference representation. Global-scale models are needed for longer forecast horizons, and use either a

*Bureau of Meteorology Research Centre, BMRC Data Assimilation Group, GPO Box 1289, Melbourne, VIC 3001. E-mail: J.Kepert@bom.gov.au

[1]The anomaly correlation is the correlation coefficient between the forecast and verifying analysis, both with climatology subtracted off. As a rule-of-thumb, 60% marks the threshhold of useful skill.

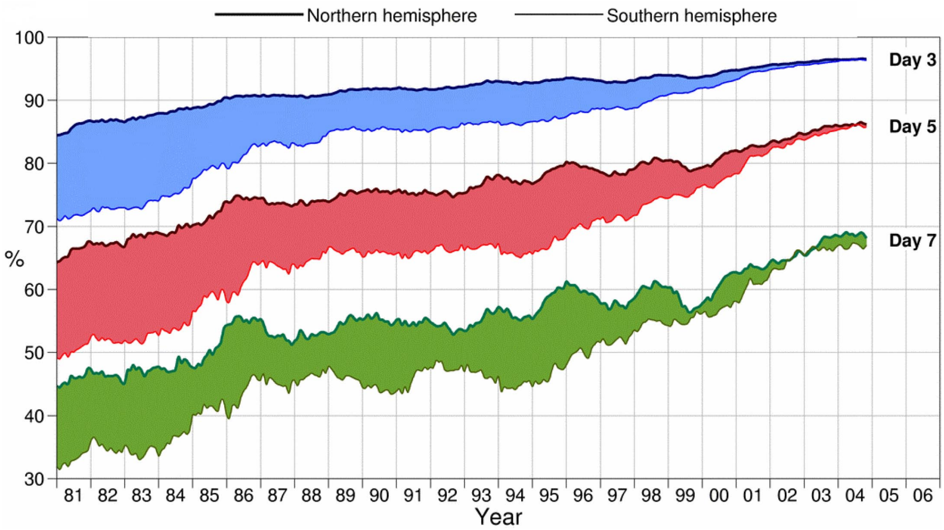

Figure 1. Anomaly correlation skill scores for the past 25 years for three-day (top), five-day (middle) and seven-day (bottom) forecasts by the ECMWF operational forecasting system, over the Northern (dark curves) and Southern (lighter curves) Hemispheres. Figure courtesy of Adrian Simmons, ECMWF.

finite-difference or a spectral (spherical harmonic) representation. As with other applications of computational fluid dynamics, it is necessary to include representations of physical processes that are not resolved by the fluid dynamics part of the model. In the atmosphere, these parametrised processes include turbulent diffusion, the fluxes of heat and water vapour to and from the underlying surface, radiation, and latent heat release due to evaporation and condensation of water. Accurately representing these processes is challenging. For instance, clouds come in many shapes and sizes, consist of some mixture of (possibly supercooled) liquid water droplets with various forms of ice, and interact strongly with radiation.

The other part of the prediction problem is finding the initial condition. Numerical weather prediction is a mixed initial value/boundary value problem — the forecast depends upon the initial atmospheric state, fixed boundary conditions such as topography, and variable ones such as sea surface temperature. Analysing the many disparate data sources to obtain the initial condition is known as data assimilation.

Data assimilation

In the Southern Hemisphere, satellite data are of paramount importance, since the large areas of ocean are nearly devoid of conventional observations. The satellite images seen on the television news are usually at 11μm wavelength, in the infrared. This is away from any absorption/emission lines in the atmospheric spectrum, and known as a window channel since the satellite sees essentially black-body radiation from the underlying surface or intervening cloud. In contrast, for frequencies within an atmospheric absorption band, upwelling radiation from the earth's surface is absorbed and re-emitted as it passes through the atmosphere. The outgoing radiation near the centre of the band where the atmosphere is opaque will be representative of the temperature in an upper layer of the atmosphere, whereas frequencies towards the edge of the band will be more sensitive to conditions lower in the atmosphere. Thus by taking many measurements at closely spaced frequencies across a peak in the atmospheric emission spectrum, information about the vertical temperature structure can be acquired.

Retrieving temperature profiles from such multiple radiance measurements is difficult. The radiative transfer equation is nonlinear, and each frequency is sensitive to a rather thick layer of the atmosphere. Mathematically, the problem is poorly conditioned and underdetermined, so a strategy of using extra information is beneficial. In fact, it is best to utilise the radiance data directly in the assimilation, rather than inverting the radiative transfer calculation to obtain temperatures and analysing those, as direct use implicitly utilises all the other data to help constrain the retrieval of temperature profiles from the radiances.

So how does data assimilation work? At each analysis time, we have two sources of information: a short-term numerical forecast from the previous analysis, and some observations. The problem of combining these disparate sources of information is approached by a least-squares minimisation,

$$J(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{x}_f)^{\mathsf{T}} \boldsymbol{B}^{-1} (\boldsymbol{x} - \boldsymbol{x}_f) + (\mathcal{H}(\boldsymbol{x}) - \boldsymbol{y})^{\mathsf{T}} \boldsymbol{R}^{-1} (\mathcal{H}(\boldsymbol{x}) - \boldsymbol{y}) \tag{1}$$

Here, the vector $\boldsymbol{x}$ is the analysis, $\boldsymbol{x}_f$ the short-term forecast, $\boldsymbol{y}$ are the observations, and the first term on the right-hand side measures the fit of the analysis to the short-term forecast, while the second measures the fit to the observations. $\mathcal{H}$ is an operator that produces the analysis estimate of the observed values. For observations of temperature, humidity or wind, $\mathcal{H}$ is just an interpolation from the model grid, but for satellite radiance measurements, $\mathcal{H}$ includes a radiative transfer calculation. The matrix $\boldsymbol{R}$ contains the variance of the random error in all the observations[2], together with the covariance of the error between all pairs of observations. Errors in observations are mostly independent, so $\boldsymbol{R}$ is nearly diagonal. Matrix $\boldsymbol{B}$ contains the error covariance of the short-term forecast, and is definitely not diagonal. Firstly, if the forecast is in error at a particular location, it is likely that similar errors apply nearby. Secondly, a forecast error in, say, pressure, is likely to be accompanied by errors in the wind, since wind and pressure are strongly coupled by atmospheric dynamics. Specifying $\boldsymbol{B}$ is a difficult but important task, since a well-formulated $\boldsymbol{B}$ allows the analysis to produce dynamically consistent results — wind observations are used not just to analyse the wind, but also to improve the depiction of the temperature and pressure fields.

Thus (1) is a very standard equation in statistics, that for finding the minimum variance estimate. With further assumptions, the analysis becomes the maximum likelihood estimate, with links to a large body of statistical theory and practice. Atmospheric data assimilation is different to other applications of these ideas in perhaps two ways: the enormous number of observations (currently millions per day), and the need to deal with highly correlated errors in the background forecast.

One approach to solving (1) is to differentiate and solve directly $\nabla J = 0$. This approach has gradually fallen into disfavour, since it involves directly inverting

[2]Random observation error includes instrument error, errors in the observation operator $\mathcal{H}$, and errors due to the instrument being affected by smaller scales of atmospheric motion than are resolved by the NWP system.

matrices whose dimension is the number of observations, and because it essentially replaces $\mathcal{H}$ by its first-order Taylor-series expansion, while the radiative transfer parts of $\mathcal{H}$ are quite nonlinear. Nowadays, we directly minimise (1) by a conjugate gradient algorithm or similar, called variational assimilation. Because we solve for the full three-dimensional structure of the atmosphere simultaneously, the specific algorithm is known as 3D-Var. Note that because observations of all types are considered simultaneously, together with the short-term forecast, a large amount of extra information is available to help constrain the poorly conditioned and underdetermined inversion of the satellite radiances.

Representing $\boldsymbol{B}$ is now important on two fronts. Apart from being necessary to produce a dynamically realistic result, $\boldsymbol{B}$ has a big influence on the conditioning of the problem, and hence the speed with which the minimisation algorithm will converge. Unfortunately, the naïve approach of calculating $\boldsymbol{B}$ in the model variables fails miserably. In this space, $\boldsymbol{B}$ is rank deficient to within numerical accuracy because it is representing highly correlated variables, and so direct minimisation of (1) will fail to converge. In addition, $\boldsymbol{B}$ contains the square of the number of model variables elements, and hence is too large to store, let alone operate on (the inverse is not required by minimisation algorithms, but it is necessary to be able to calculate the effect of multiplying a vector by $\boldsymbol{B}$).

Methods for representing $\boldsymbol{B}$ typically involve the following components:

- *Transform to less-correlated variables.* Writing the horizontal wind velocity (u, v) in terms of streamfunction ψ and velocity potential χ via

$$u = -\frac{\partial \psi}{\partial y} + \frac{\partial \chi}{\partial x} \tag{2}$$

$$v = \frac{\partial \psi}{\partial x} + \frac{\partial \chi}{\partial y} \tag{3}$$

  is helpful since forecast errors in ψ and χ are more isotropic and less cross-correlated than those in u and v. Similarly, there are quite accurate approximate balance relationships between ψ and the atmospheric pressure field, so replacing pressure by the residual unbalanced pressure eliminates the strong correlation between pressure and ψ.
- *Transform to spectral space.* Either a double Fourier representation (for a limited area model) or spherical harmonics (for a global model) are used in the horizontal. The vertical transformation may use empirical modes of some type, such as the leading eigenvectors of a covariance matrix calculated from a large sample of atmospheric columns. The variable transformations, and horizontal and vertical transformations, together make meteorologically reasonable parameterisations of $\boldsymbol{B}$ diagonal, giving the ultimate in good conditioning and computational efficiency. In fact, in variational assimilation, $\boldsymbol{B}$ is never defined in physical space, but rather in the transformed space.
- *Truncate the small scales.* The errors in the background forecast are known to be correlated over length scales of a few hundred kilometres or more. Equivalently, the error spectrum is red, with little power at small scales. Thus a lot of computer power can be saved by simply truncating $\boldsymbol{B}$ at some suitable scale, especially early in the iterative minimisation.

In practice, it is usual to replace (1) by the *incremental formulation*,

$$J(\boldsymbol{\delta x}) = \boldsymbol{\delta x}^{\mathsf{T}} \boldsymbol{B}^{-1} \boldsymbol{\delta x} + (\mathcal{H}(\boldsymbol{x}_f) + \boldsymbol{H}\boldsymbol{\delta x} - \boldsymbol{y})^{\mathsf{T}} \boldsymbol{R}^{-1} (\mathcal{H}(\boldsymbol{x}_f) + \boldsymbol{H}\boldsymbol{\delta x} - \boldsymbol{y}) \tag{4}$$

where $\boldsymbol{\delta x} = \boldsymbol{x} - \boldsymbol{x}_f$ and $\boldsymbol{H}$ is the Jacobian of $\mathcal{H}$. Here, the analysis estimate of the observation $\mathcal{H}(\boldsymbol{x})$ has been replaced by a first-order Taylor-series expansion $\mathcal{H}(\boldsymbol{x}_f) + \boldsymbol{H}\boldsymbol{\delta x}$, which facilitates the truncation of the small scales in $\boldsymbol{B}$ since $\boldsymbol{\delta x}$ is only required at the reduced resolution. Note that this formulation does not involve extra coding, since calculating ∇J from (1) already required $\boldsymbol{H}$. A further refinement is to update the linearisation once or twice during the minimisation by replacing $\mathcal{H}(\boldsymbol{x}_f)$ with the current best estimate $\mathcal{H}(\boldsymbol{x}_f + \boldsymbol{\delta x}_n)$.

A matter of time

So far we have implicitly assumed that all the data occur at the analysis time. In practice, assimilation is usually done four times a day and all data in a six-hour window is assumed to occur at the middle of that window. This introduces some errors — weather systems move and develop! These errors can be reduced by assimilating more frequently, but that has its own problems. A better way is to introduce the time dimension into the assimilation, so-called four-dimensional variational assimilation (4D-Var). Suppose we have observations at two times (Fig. 2). The black curves show the state trajectory with independent 3D-Var assimilation at each time. But suppose we could use the observation $\boldsymbol{y}_2$ at time t_2 to adjust the state at t_1 in such a way that a forecast from t_1 to t_2 is now closer to the observation? We add another term to the cost function

$$J(\boldsymbol{x}) = \cdots + (\mathcal{H}_2(\mathcal{M}(\boldsymbol{x})) - \boldsymbol{y}_2)^\mathsf{T} \boldsymbol{R}_2^{-1} (\mathcal{H}_2(\mathcal{M}(\boldsymbol{x})) - \boldsymbol{y}_2) \tag{5}$$

where $\mathcal{M}$ is the model forecast from t_1 to t_2 and the subscripts 2 refer to the time t_2. Now the gradient ∇J, needed for the minimisation, has an additional term

$$\nabla J = \cdots + 2\boldsymbol{M}^\mathsf{T} \boldsymbol{H}_2^\mathsf{T} \boldsymbol{R}_2^{-1} (\mathcal{H}_2(\mathcal{M}(\boldsymbol{x})) - \boldsymbol{y}_2) \tag{6}$$

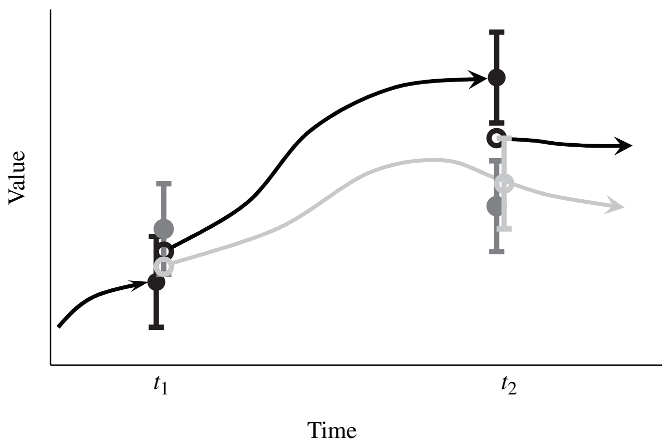

Figure 2. Schematic assimilation with two observation times. The 3D-Var assimilation is in black, with the curves representing the short-term forecast between analysis times. The analysis (open black circle) simultaneously minimises the sum of squared distances from the observation (filled dark-grey circle) and the short-term forecast (filled black circle). 4D-Var assimilation is in light grey. The analysed value at t_1 is close to the short-term forecast and observation at t_1, and initialises a forecast that is close to the observation at t_2.

that, by the chain rule of differentiation, contains the adjoint of the Jacobian of the observation operator, $\boldsymbol{H}_2^\top$, and the adjoint of the Jacobian of the model, $\boldsymbol{M}^\top$. $\boldsymbol{H}_2^\top$ takes information about the degree of misfit from radiance space back to analysis space, while $\boldsymbol{M}^\top$ propagates this misfit information backwards in time from t_2 to t_1. Minimising this J will produce an analysis at time t_1 that is close to the background and observation at that time, and that initialises a (linearised) forecast that is close to the observation at time t_2. Adding additional time levels is a straightforward extension, as is the incremental formulation.

To get 4D-Var to work on an atmospheric model in the order of 10^6 to 10^7 variables, assimilating millions of observations per day, within the limited time available under operational forecast constraints, is a major undertaking. The nonlinear atmospheric model consists of several hundred thousand lines of code, and 4D-Var requires the development of codes to represent operations by the Jacobian of the model and its adjoint. The minimisation itself benefits from research into very large optimisation problems. Even with all the computational tools in place, good results require careful attention to estimating the necessary statistics and to quality control of the observations.

Limitations of the current 4D-Var algorithm include the limited time frame over which the linearisation of the model is valid, and the failure to explicitly account for random error in the model. Accounting for the second will help with the first, and open up the possibility of performing assimilation over time windows of the order of a week long, with (hopefully) further significant improvements in forecast accuracy. The benefits include some theoretical links to other branches of mathematics: it can be shown with some reasonable assumptions that such 4D-Var schemes produce identical analyses to an extended Kalman filter (EKF). Although such a 4D-Var doesn't give the estimate of the analysis error covariance that the EKF does, our inability to even store this for the atmosphere means that it is hardly a limitation.

Maths@work

I have given a broad outline of one component of NWP that mathematics is having a big impact upon. Obviously it is not the only one — mathematics has a strong influence on the computational fluid dynamics aspects, as well as the efficient representation of radiative transfer, turbulence and clouds. A further area is in the growing field of probability forecasts, which require what is essentially Monte Carlo simulation of the atmosphere. Many questions arise here, and a great deal of work has been done on working out how best to perturb the initial conditions, represent model uncertainty, extract the probabilistic information, communicate it to users, and validate the forecasts. One successful approach has been to use the linearised and adjoint models to iteratively solve the eigen-problem of finding the most rapidly growing modes of the current atmospheric state. These unstable modes have many uses, including as a basis for Monte Carlo initial condition perturbations. A further promising field is in using an ensemble of background forecasts to produce 'errors of the day' that is, a $\boldsymbol{B}$ which reflects the varying accuracy of the background forecast in space and time, according to the difficulty of forecasting today's particular meteorological situation.

As well as having here considered only one aspect of NWP, it is important to remember that NWP is only one area in which mathematics is used in meteorology. In my career, apart from topics in data assimilation, I have worked on boundary layer flows, turbulence, tropical cyclone dynamics, forecast verification and the effects of evaporating sea-spray droplets. Mathematical tools used include the analytical and numerical solution of differential equations, Taylor and Fourier series expansion, linear algebra, multiple linear and nonlinear regression, Monte Carlo simulation, statistical significance testing, and symbolic algebra software. In my experience, many interesting problems have required several of these tools, so a breadth of knowledge has been important. I would also encourage young mathematicians that there remain many more interesting problems in meteorology upon which mathematics will have a large impact, and that the chaotic nature of the atmosphere and the ever-growing demand for meteorological services means that they should never be out of either a challenge, or a job!

Dr Jeff Kepert works in the Bureau of Meteorology Research Centre (BMRC), which undertakes research to improve scientific understanding of, and the ability to forecast, Australia's climate and weather.

Jeff originally studied at the University of Western Australia, majoring in pure mathematics and statistics. After joining the Bureau of Meteorology in 1984, he worked as a forecaster for a couple of years before returning to study for a MSc in dynamical meteorology at Monash University. This led to a couple of years working as an instructor in the Bureau's training centre, before moving to BMRC in 1992 to study the representation of clouds in numerical weather prediction systems. Since then, Jeff has worked on tropical cyclone dynamics, air–sea exchanges, turbulence, high-resolution wind prediction, and data assimilation. During this period, he completed a part-time PhD on tropical cyclone boundary layer winds at Monash University.