

# The Australian community climate and earth system simulator global and regional ensemble prediction scheme

Terence J. O’Kane<sup>1</sup>

Michael J. Naughton<sup>2</sup>

Yi Xiao<sup>3</sup>

(Received 12 August 2008; revised 3 November 2008)

## Abstract

We report on progress in the development of the Australian Community Climate and Earth Systems Simulator Global and Regional Ensemble numerical weather Prediction Scheme at the Australian Bureau of Meteorology. Based on the UK Met Office ensemble, AGREPS implements an Ensemble Transform Kalman Filter to generate independent initial perturbations as fast growing disturbances with structures and growth rates typical of the analysis errors. This method allows information about the fast growing errors to be incorporated into the initial perturbations for the forecast. An ensemble of model states is propagated, using the numerical weather prediction system and observing network at the Australian Bureau of Meteorology, from which covariances are constructed then localized and inflated to minimize the effect of small sample size.

---

<http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/1421>

gives this article, © Austral. Mathematical Soc. 2008. Published November 21, 2008. ISSN 1446-8735. (Print two pages per sheet of paper.)

# Contents

<b>1</b>	<b>Introduction</b>	<b>C386</b>
<b>2</b>	<b>Ensemble transform Kalman filter</b>	<b>C389</b>
2.1	AGREPS ensemble transform Kalman filter implementation	C391
2.2	Horizontal localization and inflation . . . . .	C391
<b>3</b>	<b>Discussion and future plans</b>	<b>C393</b>
	<b>References</b>	<b>C395</b>

## 1 Introduction

Using deterministic forecasts the early pioneers of numerical weather prediction tried to establish the theoretical limits to atmospheric predictability in terms of the divergence of pairs of initially close states [4, 11, 7]. Weather forecasting has come to be regarded as a statistical problem of predicting the probability density function of atmospheric states or, equivalently, of calculating the moments of meteorological variables. The inherently chaotic nature of the system combined with errors in the observed initial conditions lead to the failure of deterministic forecasts after a number of days (dependent on the flow dynamics).

In numerical weather prediction and data assimilation for large numerical prediction models, with millions of variables, calculating the full covariance matrices of the background errors as they evolve with time remains a very difficult problem. Ensemble methods estimate the flow dependent background error covariance matrices by construction from ensembles of short term forecasts with slightly differing initial conditions. In the ensemble Kalman filter (ENKF) approach an ensemble of model states is propagated with a fully nonlinear model allowing the error covariance matrix to be calculated with

no moment closure required. This allows the construction of the forecast error covariance at any given time by averaging over the ensemble. Unlike single realization deterministic forecasts, ensemble forecasts are able to provide both improved estimates of the mean and estimates of the forecast error variance. For some cases ensembles may also give information about higher order moments.

Typically, sampling error arises due to an insufficient number of realizations, as occurs when the sample size is less than the number of degrees of freedom, which in some cases is further exacerbated through the use of perturbed observations such as is the case for the stochastic ENKF. For global ensembles, covariance localization [3, 5], in which one assumes that background error correlations vanish for points separated by a distance greater than several thousand kilometers is a common approach to removing spurious long range correlations which arise due to small ensemble size. In ensemble data assimilation it is generally necessary to employ an empirically selected covariance inflation factor in order to ensure that the filtering solution does not diverge from the observations and to keep prior covariances small [1, 5]. For ensemble prediction, inflation is also necessary to account for an underestimation of the analysis error covariances.

For typical ensemble sizes currently available in numerical weather prediction ( $\leq 100$ ), Monte Carlo methods such as the ENKF, in which random initial perturbations are sampled isotropically, were found to grow too slowly, leading to underestimated error variances. More recent approaches endeavor to generate independent initial perturbations as fast growing disturbances with structures and growth rates typical of the analysis errors such as the ‘breeding’ method developed by Toth & Kalnay [13]. This method of bred initial forecast perturbations allows information about the fast growing errors to be incorporated into the initial perturbations for the forecast [8]. For particularly dynamic flows, such as when emergent coherent structures are developing, errors arise due to fast growing large scale instabilities. Toth & Kalnay [13] argue that the bred vectors are stochastically and nonlinearly

modified versions of the leading Lyapunov vectors (LLV's). They also note that after an initial transient period ( $\approx$  a week) initial random perturbations growing on tropospheric flows converge toward the structure of the LLVs. In 'breeding' the perturbations are periodically rescaled using a global (or regional) scaling factor so that they approximate fast growing errors within assimilation schemes. Tracton and Kalnay [14] describe the implementation of an operational ensemble prediction scheme based on the breeding of growing modes.

The Ensemble Transform Kalman Filter (ETKF) [12, 17] was developed in an effort to reduce the loss of spread associated with the rapid convergence of the initial random perturbations onto the LLV structures. The ETKF employs a transform matrix to mix perturbations from different members, which are taken to be the difference between the perturbed forecast and the ensemble mean (or in some cases the control forecast). The ETKF falls into a class of deterministic ensemble square root filters based on the use of unperturbed observations [12]. Wang and Bishop [15] compared the breeding method to the ETKF, while Wei et al. [17] implemented the ETKF in an operational global prediction system.

In this article we describe progress on the development of the ACCESS Global & Regional Ensemble Prediction Scheme (AGREPS) at the Bureau of Meteorology, Centre for Australian Weather & Climate Research. AGREPS is based on the United Kingdom Met Office ensemble (MOGREPS), employs an ensemble transform Kalman filter and critically uses global and regional observations to determine the transform matrix and inflation factors.

## 2 Ensemble transform Kalman filter

The ETKF is based on an application of the Kalman filter methodology in which  $k$ -ensemble forecast and analysis perturbations are

$$\begin{aligned}\mathbf{Z}^f &= \frac{1}{\sqrt{k-1}}[\mathbf{z}_1^f, \mathbf{z}_2^f, \dots, \mathbf{z}_k^f], \\ \mathbf{Z}^a &= \frac{1}{\sqrt{k-1}}[\mathbf{z}_1^f, \mathbf{z}_2^f, \dots, \mathbf{z}_k^a]\end{aligned}\quad (1)$$

and where the state vectors are  $\mathbf{z}_i^f = \mathbf{x}_i^f - \mathbf{x}^f$  and  $\mathbf{z}_i^a = \mathbf{x}_i^a - \mathbf{x}^a$  which are  $n$ -dimensional in model space. Typically  $\mathbf{x}^f$  is the mean of  $k$ -ensemble forecasts while  $\mathbf{x}^a$  is the analysis from the ACCESS 4D-variational operational data assimilation system. As we are principally concerned with ensemble prediction and not data assimilation the mean or first moment is unchanged between analysis and forecast. The ETKF methodology acts to choose appropriate initial forecast perturbations consistent with error covariance update equations within the vector subspace of ensemble perturbations,

$$\mathbf{P}^a = \mathbf{P}^f - \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R}^{-1}) \mathbf{H} \mathbf{P}^f \quad (2)$$

where  $\mathbf{P}^f$  and  $\mathbf{P}^a$  are  $n \times n$  forecast and analysis covariance matrices formed as  $\mathbf{P}^f = \mathbf{Z}^f \mathbf{Z}^{fT}$  and  $\mathbf{P}^a = \mathbf{Z}^a \mathbf{Z}^{aT}$  where superscript  $T$  denotes the matrix transpose.  $\mathbf{R}$  is the  $p \times p$  observational error covariance matrix for  $p$  observational values,  $\mathbf{H}$  is the linearized observational operator mapping the forecast grid point values onto the observational points. Equation (2) is the Kalman filter error covariance update equation; however, unlike the ENKF, the ETKF does not use perturbed observations. O’Kane & Frederiksen [9] and Tippett et al. [12] discussed relevant details of both the Kalman filter and ETKF equations. In order to calculate the normal ETKF transform matrix  $\mathbf{T}$ , we are required to calculate the matrix of forecast ensemble perturbations in normalized observation space

$$\mathbf{E} = \left( \mathbf{R}^{-1/2} \mathbf{H} \mathbf{Z}^f \right)^T \left( \mathbf{R}^{-1/2} \mathbf{H} \mathbf{Z}^f \right) \quad (3)$$

where  $\mathbf{Z}^f$  are the forecast perturbations.

We now find the eigenvectors  $\mathbf{C}$  and eigenvalues  $\Gamma$  of equation (3) which is equivalent to the matrix  $\mathbf{Z}^{fT}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{Z}^f$ . The transform matrix  $\mathbf{T}$  is now defined in terms of the  $k \times (k-1)$  matrix of non-zero eigenvectors  $\mathbf{C}$  and the  $(k-1) \times (k-1)$  diagonal matrix with non-zero eigenvalues such that

$$\mathbf{T} = \mathbf{C}(\Gamma + \mathbf{I})^{1/2}\mathbf{C}^T \quad (4)$$

where  $\mathbf{I}$  is the identity matrix. Equation (4) corresponds to the transform matrix in spherical simplex form as described in articles by Purser [10] and Wang et al. [16]. By centering the analysis perturbations about the analysis, the spherical simplex form allows nearly twice as many uncertain directions to be spanned as the traditional positive-negative centering; however, this approach is only appropriate where the number of uncertain directions is larger than the ensemble size.

In breeding, each analysis perturbation is produced by a straightforward rescaling of the forecast perturbation by a constant factor  $\mathbf{z}_i^a = \mathbf{z}_i^f\mathbf{c}$ . The ETKF on the other hand generates new analysis perturbations via the transformation matrix

$$\mathbf{Z}^a = \mathbf{Z}^f\mathbf{T}, \quad (5)$$

that is, the transform matrix tells how to mix perturbations from different members. Equation (3) shows that in order to generate  $\mathbf{T}$  we require a good knowledge of the observational error covariances. In that regard satellite radiance data from the Advanced TIROS Operational Vertical Sounder (ATOVS) and data from dropsonde (weather reconnaissance devices containing a GPS receiver), along with pressure, temperature, and humidity (PTH) sensors to capture atmospheric profiles and thermodynamic data, are critical to the functionality of the ETKF. In addition, observations are not used directly: instead observations are extracted, processed and quality controlled, and the model forecast of the observations used thereby removing the need for the ETKF code to know the observational operator  $\mathbf{H}$ .

## 2.1 **AGREPS ensemble transform Kalman filter implementation**

As stated, the purpose of the ETKF code is to provide initial conditions for the AGREPS forecasts. This is achieved by producing perturbation fields for wind, temperature, humidity and pressure on all model levels (currently 50 vertical levels) for each ensemble member (currently limited to 23 due to the large computational cost). The perturbation fields are then combined with the reconfigured analysis provided by 4D variational data assimilation while the control forecast uses a reconfigured analysis without perturbations and hence receives no input from the ETKF. Reconfiguration refers to changing resolution to the required domain. The ETKF generates analysis perturbations by mixing and scaling evolved perturbations valid at the new analysis time (in our case time+12 hrs) from the previous forecast cycle. Given error estimates for each assimilated observation and the background error covariance, represented by the evolved perturbations from the previous forecast, the ETKF transform matrix produces an  $n$ -dimensional representation of the analysis error covariance matrix of an optimal data assimilation system, where  $n$  is the number of perturbations. Model observations are required for each observation for each ensemble member in order to provide estimates of background uncertainty in observation space. The required pseudo-observations are calculated separately by an additional observation processing system.

## 2.2 **Horizontal localization and inflation**

The large computational costs involved in running a single operational numerical weather forecast model, let alone many, imposes a severe upper bound on the dimension of our ensemble (currently 23 plus control). Unrealistically small ensemble sizes often results in overestimated background error covariances and as a consequence overestimation of the impact of each observation. To ameliorate this sampling error we employ the horizontal localiza-

tion methodology of Houtekamer & Mitchell [6] in which 92 predetermined equally spaced localization centers are specified over the global domain. For each center, only observations falling within a 5000 km radius are used to produce a local transform matrix. The final transform matrix for each point on the model grid is obtained by interpolation between the transform matrices for the nearest localization centers. This approach effectively removes any spurious long range correlations present in the ensemble estimated background error covariance, and has also been shown to improve the relationship between ensemble spread and error as a function of latitude. Disadvantages are that the localization process can remove information due to large scale flow dependent inhomogeneities. Future plans are to investigate flow adaptive methods to reduce spurious ensemble correlations through the use of moderation functions generated by smoothed ensemble correlations raised to a higher even ordered power [2].

For the ETKF inaccuracies in estimation of the observation errors can lead directly to an over/under spread ensemble. In data assimilation, covariance inflation is commonly used to increase error variances for correlated observations. For the ETKF, the small number of perturbations compared to the nominal degrees of freedom of model space results in the analysis error covariance being greatly under estimated by the covariance of the transformed ensemble. Therefore, we choose to inflate the analysis perturbations using a variant of the method of Wang & Bishop [15]. The inflation procedure assumes that the global sum of squares of the forecast and observation differences at the same time does not depend on the initialization of the forecast and that the number, quality and location of observations are similar at all analysis times. Equation (6) states that the square of the ensemble spread depends on the difference between the mean squared innovation and the summed observation error variances:

$$\text{Tr}(\mathbf{H}\mathbf{P}^f\mathbf{H}^T) = \text{Tr}[(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^f)(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^f)^T] - \text{Tr}(\mathbf{R}). \quad (6)$$

Based on this, Wang & Bishop [15] proposed that the transform matrix

calculation be inflated using

$$\pi_n = \pi_{n-1} \sqrt{\frac{\text{Tr}[(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^f)(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^f)^T] - \text{Tr}(\mathbf{R})}{\text{Tr}(\mathbf{H}\mathbf{P}^f\mathbf{H}^T)}}, \quad (7)$$

where  $\pi_n$  and  $\pi_{n-1}$  correspond to new and old inflation factors and the terms in the  $\sqrt{\quad}$  correspond to the rms error of the ensemble mean for the selected observations (numerator) and the rms spread of the corresponding model observations (denominator). However, ensemble spread calibration requires knowledge of the true observation error variance  $\mathbf{R}$ , which is only available using dropsonde and ATOVS data. Equation (7) was found to lead to spurious oscillatory behavior not present in the more stable second order variant equation

$$\pi_n = \pi_{n-1} \sqrt{\frac{\{[\text{Tr}[(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^f)(\mathbf{y} - \mathbf{H}\bar{\mathbf{x}}^f)^T] - \text{Tr}(\mathbf{R})]_n \text{Tr}(\mathbf{H}\mathbf{P}^f\mathbf{H}^T)_{n-1}\}^{1/2}}{\text{Tr}(\mathbf{H}\mathbf{P}^f\mathbf{H}^T)_n}}. \quad (8)$$

The Kalman filter methodology implicitly assumes near linear perturbation growth which will be problematic for cases where errors induced by barotropic and baroclinic instabilities grow rapidly and then tend to saturate. By including information about flow dependent ‘errors of the day’ the ETKF approach is able to capture some information about nonlinear error growth. O’Kane & Frederiksen [9] considered cases for which fast growing disturbances occur and examined the role of higher order moments in non-Gaussian flows. Such cases are typified by the reduced predictability associated with the formation and decay of large scale coherent structures in the atmosphere evident as high-low blocking dipoles [8].

### 3 Discussion and future plans

To date our efforts have been focused toward the implementation of AGREPS on the supercomputing system at the Bureau of Meteorology. This has in-

AGREPS P robability map for 24h precip 10mm (magenta), and 10m wind speed > 15m/sec (green)  
DT: 12Z 20080110 V alid from + 24h to + 48h  
(Ensemble mean MSL overlain in contours)

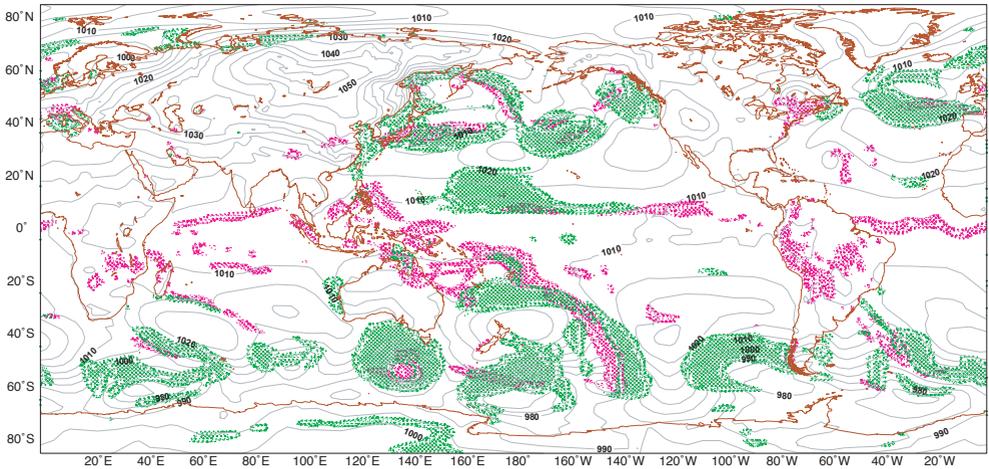


FIGURE 1: Combined high impact weather risk map for intense rainfall and high near-surface winds.

volved setting up cold start and forecast ensemble suites for both global (80 km) and a high resolution limited area Australian regional (37.5 km) domain. The regional ensemble cannot be run independently as it requires lateral boundary conditions provided by the global ensemble. Additionally AGREPS has been integrated into a 4D variational operational data assimilation scheme. Figure 1 describes an example of a forecast product that AGREPS produces, namely a combined high impact global weather risk map for precipitation exceeding 10 mm (magenta) and 10 m winds exceeding 15 m/sec (green) which in this particular case are valid on day two of the given forecast period beginning 10 January 2008. Future efforts are likely to be focused on validation and progressing the ensemble to become an official Bureau operational suite. Research efforts are also ongoing, with particular interests in methods of covariance localization, inflation, and the effect of sample size on forecast accuracy.

**Acknowledgments** The authors acknowledge the contributions of Dr Neil Bowler (UK Met Office), Dr Kamal Puri, Asri Sulaiman and Dr Ilia Bermous (CAWCR ABM) to the AGREPS project.

## References

- [1] Anderson, J. L., 2001 An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.* **129**, 2884–2903  
[doi:10.1175/1520-0493\(2001\)129;2884:AEAKFF;2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129;2884:AEAKFF;2.0.CO;2) C387
- [2] Bishop, C. H., Hodyss, D., 2007 Flow adaptive moderation of spurious ensemble correlation and its use in ensemble-based data assimilation. *Q. J. R. Meteorol. Soc.* **133**, 2029–2044 [doi:10.1002/qj.169](https://doi.org/10.1002/qj.169) C392
- [3] Buehner, M., 2005 Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP

- setting. *Q. J. R. Meteorol. Soc.* **131**, 1013–1043 doi:10.1256/qj.04.n C387
- [4] Charney, J. G., 1966 The feasibility of a global observation and analysis experiment. *Bull. Amer. Meteor. Soc.* **47**, 200–220 C386
- [5] Hamill, T. M., Whitaker, J. S., & Snyder, C., 2001 Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.* **129**, 2776–2790 doi:10.1175/1520-0493(2001)129;2776:DDFOBE;2.0.CO;2 C387
- [6] Houtekamer, P. L., & Mitchell, H. L., 1998, Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.* **126**, 796–811 doi:10.1175/1520-0493(1998)126;0796:DAUAEK;2.0.CO;2 C392
- [7] Kasahara, A., 1972 Simulation experiments for meteorological observing systems for GARP. *Bull. Amer. Meteor. Soc.* **53**, 252–264 C386
- [8] O’Kane, T. J., & Frederiksen, J. S., 2008a, A comparison of statistical dynamical and ensemble prediction during blocking. *J. Atmos. Sci.* **65**, 426–447 doi:10.1175/2007JAS2300.1 C387, C393
- [9] O’Kane, T. J., & Frederiksen, J. S., 2008b Comparison of statistical dynamical, square root and ensemble Kalman filters. *entropy* In Press C389, C393
- [10] Purser, J., 1996 Arrangement of ensemble in a simplex to produce given first and second moments, *NCEP Internal Report*. Available from the author at <mailto:Jim.Purser@noaa.gov.au> C390
- [11] Smagorinsky, J., 1969 Problems and promises of deterministic extended range forecasting. *Bull. Amer. Meteor. Soc.* **50**, 286–311 C386
- [12] Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., & Whitaker, J. S., 2003, Ensemble square root filters. *Mon. Wea. Rev.*

- 131**, 1485–1490 doi:10.1175/1520-0493(2003)131;1485:ESRF;2.0.CO;2  
C388, C389
- [13] Toth, Z., & Kalnay, E., 1997, Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.* **125**, 3297–3319  
doi:10.1175/1520-0493(1997)125;3297:EFANAT;2.0.CO;2 C387
- [14] Tracton, M. S., & Kalnay, E., 1993, Operational ensemble prediction at National Meteorological centre: Practical aspects. *Weather and Forecasting* **8**, 379–398  
doi:10.1175/1520-0434(1993)008;0379:OEPATN;2.0.CO;2 C388
- [15] Wang, X., & Bishop, C. H., 2003, A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.* **60**, 1140–1158  
doi:10.1175/1520-0469(2003)060;1140:ACOBAE;2.0.CO;2 C388, C392
- [16] Wang, X., Bishop, C. H., & Julier, S. J., 2004, Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble. *Mon. Wea. Rev.* **132**, 1590–1605  
doi:10.1175/1520-0493(2004)132;1590:WIBAE0;2.0.CO;2 C390
- [17] Wei, M., Toth, Z., Wobus, R., Zhu, Y., Bishop, C. H. and Wang, X., 2006 Ensemble Transform Kalman Filter-based ensemble perturbations in an operational global prediction system at NCEP. *Tellus*, **58A**, 2006, 28–44 doi:10.1111/j.1600-0870.2006.00159.x C388

## Author addresses

1. **Terence J. O’Kane**, Centre for Australian Weather and Climate Research, Bureau of Meteorology, Docklands, AUSTRALIA.  
mailto:t.okane@bom.gov.au

2. **Michael J. Naughton**, Centre for Australian Weather and Climate Research, Bureau of Meteorology, Docklands, AUSTRALIA.
3. **Yi Xiao**, Centre for Australian Weather and Climate Research, Bureau of Meteorology, Docklands, AUSTRALIA.