# An Improved Method for Generating Ensembles of Tropical Cyclone Tracks

Jeffrey D. Kepert *
*Centre for Australian Weather and Climate Research*

Extended abstract, AMS 31st Conference on Hurricanes and Tropical Meteorology

April 24, 2014

## 1. Introduction

The Australian Bureau of Meteorology uses a Monte Carlo ensemble of TC forecasts (track, intensity, structure) to predict wind exceedance probabilities, based on the method used in the USA and described by DeMaria et al. (2009). The Bureau of Meteorology is also seeking to upgrade its tropical cyclone storm surge prediction system. Storm surge is very sensitive to cyclone track and other details of the atmospheric forcing, so surge forecasts should ideally be probabilistic. Thus an ensemble of atmospheric forcings is required. Forcing a storm surge model with the DeMaria track ensemble would be an elegant solution, since it would mean that the storm surge system would inherit the uncertainty information from another part of the forecast system and thereby benefit from the development and verification effort already made, as well as ensuring consistency between the probabilistic wind and surge forecasts. However, we shall see that these tracks are not smooth. While this irregularity is acceptable for the calculation of wind probabilities, it is likely a problem for storm surge prediction, since the ocean has a "memory" through its inertia.

This report outlines preliminary results from a revised track generation method, which does not alter those properties of the ensemble that are presently used operationally, but does improve the smoothness of individual tracks and hence their utility for forcing a storm surge ensemble prediction system. At present, the new method has only been developed for the track component of the system, but we expect that the same methods will also prove useful for the intensity and structure components. The approach is to modify the serial correlation in the DeMaria track ensemble to make individual tracks smooth, without substantially altering the probability distributions at each time. The new method also has some further advantages:

- The new method is continuous (rather than discrete) in time, so arbitrary time-interpolation is not needed.
- The new method has fewer parameters that must be fit from data, so a smaller training dataset is needed and the parameters can be upgraded more frequently, or different sets can be used for different geographical regions or times of the year.
- The parameters to the new method have somewhat clearer physical meanings than the old, so it would be more straightforward to estimate them from other sources. For example, they could be estimated directly from an NWP ensemble, or from such an ensemble and then blended with the climatological values.

* *Corresponding author address:* Centre for Australian Weather and Climate Research, Bureau of Meteorology, 700 Collins St, Docklands Vic 3008, Australia.
E-mail: J.Kepert@bom.gov.au

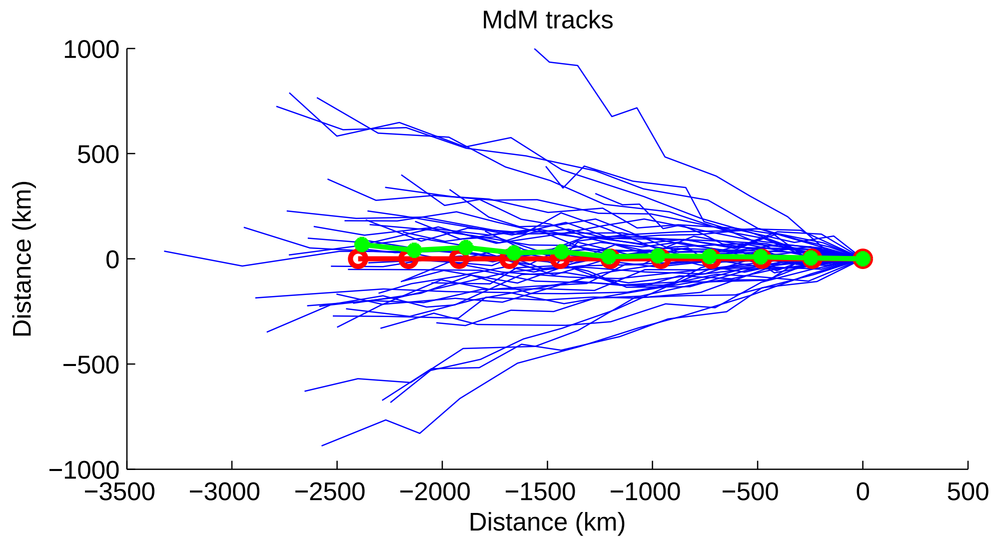


FIG. 1. An ensemble of 50 perturbed tropical cyclone tracks calculated by the method of DeMaria et al. (2009) (blue), together with the input track (red) and the ensemble mean calculated from a much larger ensemble (green).

## 2. Analysis of the DeMaria et al. (2009) method

The DeMaria et al. (2009) method models the track errors by a first-order autoregressive process. The along-track perturbation at time $t$ is given by

$$AT_t = a_t AT_{t-12} + b_t + \epsilon_t \quad (1)$$

where $a_t$, $b_t$ are adjustable parameters derived from an analysis of historical errors, $AT_{t-12}$ is the track perturbation at the preceding time 12 hours before, and $\epsilon_t$ is a random number. Other perturbations (cross-track, intensity, structure) are modelled similarly, albeit with their own set of adjustable parameters and some additional refinements to account for landfall, etc.

Figure 1 displays an ensemble of 50 such tracks (blue curves)[1]. For convenience, they are here drawn as perturbations about a straight track to the west with a uniform motion of 20 km hr$^{-1}$ (thick red curve). The ensemble mean (calculated from a 10000-member ensemble to minimise sampling error) is shown in green, and it can be seen that there is a small systematic perturbation from the deterministic forecast track (red). Note that the ensemble tracks contain unrealistic kinks; these are not present in the input track and are undesirable for storm surge forecasting since the ocean circulation will require time to adjust to the new storm motion after each such kink.

Figure 2 shows the covariance matrices for the along- and cross-track perturbations, and the corresponding standard deviations and correlation matrices, again calculated from the 10000-member ensemble. Note that the standard deviation does not increase smoothly with time, and that the correlation matrix has cusps at zero-lag (that is, along the diagonal of the matrix)[2].

## 3. The new covariance model

The covariance matrices in Fig. 2 characterize the distribution of the track ensemble; sampling

[1]This figure uses the model parameters listed in Table 1 of DeMaria et al. (2009), but approximates the random draws from the distribution shown in their Fig. 1 by random draws from a Gaussian with mean 0 km and standard deviation 50 km.

[2]We can show empirically that the cusps – that is, points where the first derivative of the correlation is discontinuous – are associated with the kinks in the track, but have not yet completed a full mathematical analysis.

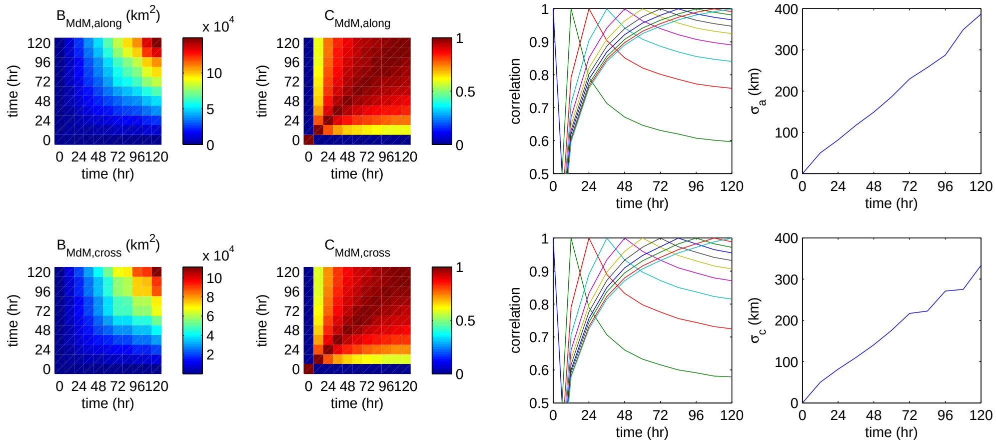


FIG. 2. Some statistical properties of the DeMaria et al. (2009) ensemble, all calculated from a 10000-member ensemble. First column: Covariance matrices for the along- (top) and cross-track (bottom) perturbations. Second column: The corresponding correlation matrices. Third column: Line plots of the rows of the correlation matrices. Note the cusps at zero lag (i.e. where the correlation equals 1). Fourth column: The standard deviations (i.e. the square roots of the diagonals of the covariance matrices).

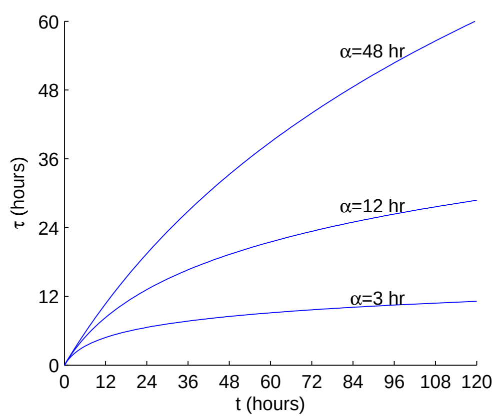


FIG. 3. The time transform (2) plotted for various $\alpha$.

from a distribution with these covariances would be another way of generating a track ensemble. The new method makes modest changes to the covariance matrices, so as to remove the kink-causing features, while not significantly changing those properties of the ensemble that are presently used operationally for wind probability prediction. In particular, this means that the variance at each time must remain almost unaltered, but that the serial correlation structure can be changed.

In making these modifications, it is important that the new matrices be covariance matrices; that is, that they represent the time-lag covariances from a statistical process. This is important for two reasons: firstly to ensure that the new method has a suitably rigorous statistical underpinning, and secondly to enable the calculation of random elements from the distribution so defined. There is an extensive body of theory on the calculation of such matrices, used for example in meteorological data assimilation. We follow that theory and define the covariance matrices in terms of a suitable analytic correlation function, except written in terms of time differences rather than distances. Gaspari and Cohn (1999) contains much of the relevant theory.

It is apparent from Fig. 2 that the perturbations are more strongly correlated later in time, than initially. We model this effect by using a transformed time variable $\tau$,

$$\tau = \alpha \log(t/\alpha + 1) \qquad (2)$$

This function is plotted for various $\alpha$ in Fig. 3. where $t$ is time and $\alpha > 0$ determines the strength of the transformation. In the limit $\alpha \to \infty$ the transformation reduces to $\tau = t$.

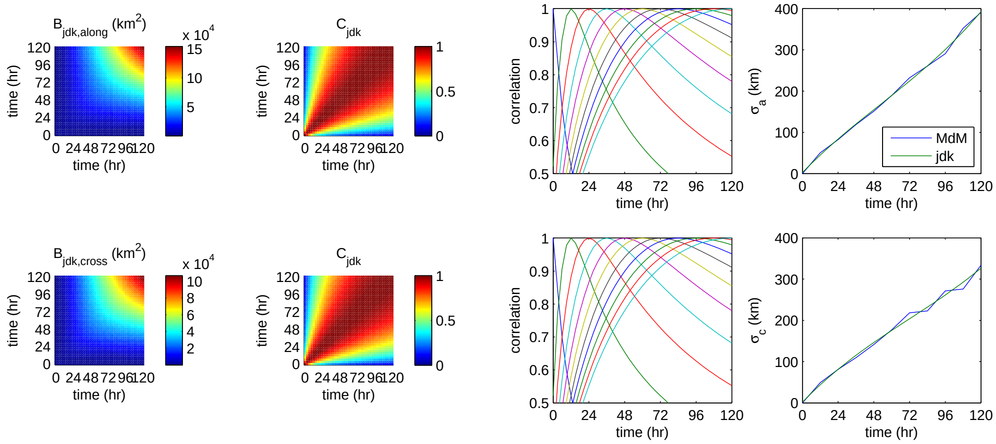


FIG. 4. As for Fig. 2, except calculated using the new method described in the text. These calculations were done with a time interval of 3 rather than 12 hr, and only every fourth line of the correlation matrices is plotted in the third column for clarity. Note that the correlation matrices for the along- and cross-track perturbations are identical. The fourth row plots standard deviations from both the original DeMaria et al. (2009) ensemble (blue) and the polynomial fit used in the new method (green).

It is apparent also from Fig. 2 that the correlation function has reasonably fat tails. The widely-used Gaussian correlation function is therefore unsuitable, and after some experimentation we chose a second-order autoregressive (SOAR) function,

$$C_{ij} = (1 + |d_{ij}|) \exp(-|d_{ij}|) \tag{3}$$

where

$$d_{ij} = (\tau_i - \tau_j)/L \tag{4}$$

is the difference between transformed times $\tau_i$ and $\tau_j$ normalized by a time scale $L$. The second derivative of (3) is defined at all $d_{ij}$ (including $d_{ij} = 0$), so the correlation function lacks cusps. The correlations in Fig. 2 are quite similar for the along- and cross-track perturbations, so we use the same matrix for each in the new model.

The time-stretching function (2) is monotonic, and the SOAR function is a correlation function, so by the results in Gaspari and Cohn (1999) the matrix with elements $C_{ij}$ is a correlation matrix.

We smooth the standard deviations and interpolate in time by fitting a cubic polynomial, as shown in Fig 4. The use of a cubic is probably excessive and linear growth may well be an adequate approximation.

The resulting matrices with $\alpha = 3\,\mathrm{hr}$, $L = 3\,\mathrm{hr}$ are plotted in Fig. 4, and it is apparent that they are a reasonable approximation to those shown in Fig. 2. The parameter values for $\alpha$ and $L$ used here are rough estimates, and a more objective method should be used operationally. Note that the diagonals of the covariance matrices are changed from those in Fig. 2 only by the slight smoothing of the standard deviations, so the marginal distribution of the pdf at any time will be almost unchanged. Thus, the use of this model to estimate wind probabilities will yield answers nearly indistinguishable from those obtained using DeMaria et al. (2009).

We generate random draws from the pdf with covariance matrix $\mathbf{B}$ as follows. $\mathbf{B}$ is a covariance matrix, so has a real symmetric square root $\mathbf{B}^{1/2}$ satisfying $\mathbf{B}^{1/2}\mathbf{B}^{1/2} = \mathbf{B}$ and $\mathbf{B}^{1/2\,T} = \mathbf{B}^{1/2}$.[3] Generate a column vector $\mathbf{x}$ of indepen-

[3] Any real matrix $\mathbf{A}$ may be written in terms of its singular value decomposition $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal and $\mathbf{\Sigma}$ is diagonal; the diagonal elements of $\mathbf{\Sigma}$ are called the singular values of $\mathbf{A}$. If $\mathbf{A}$ is square, then so are $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{\Sigma}$; if it is symmetric, we may choose $\mathbf{U} = \mathbf{V}$; and if it is positive semidefinite, the singular values are all nonnegative. A covariance matrix $\mathbf{B}$ possesses each of these properties, and so $\mathbf{U}\mathbf{\Sigma}^{1/2}\mathbf{U}^T$ is a real symmetric square root; here $\mathbf{\Sigma}^{1/2}$ denotes (in con-

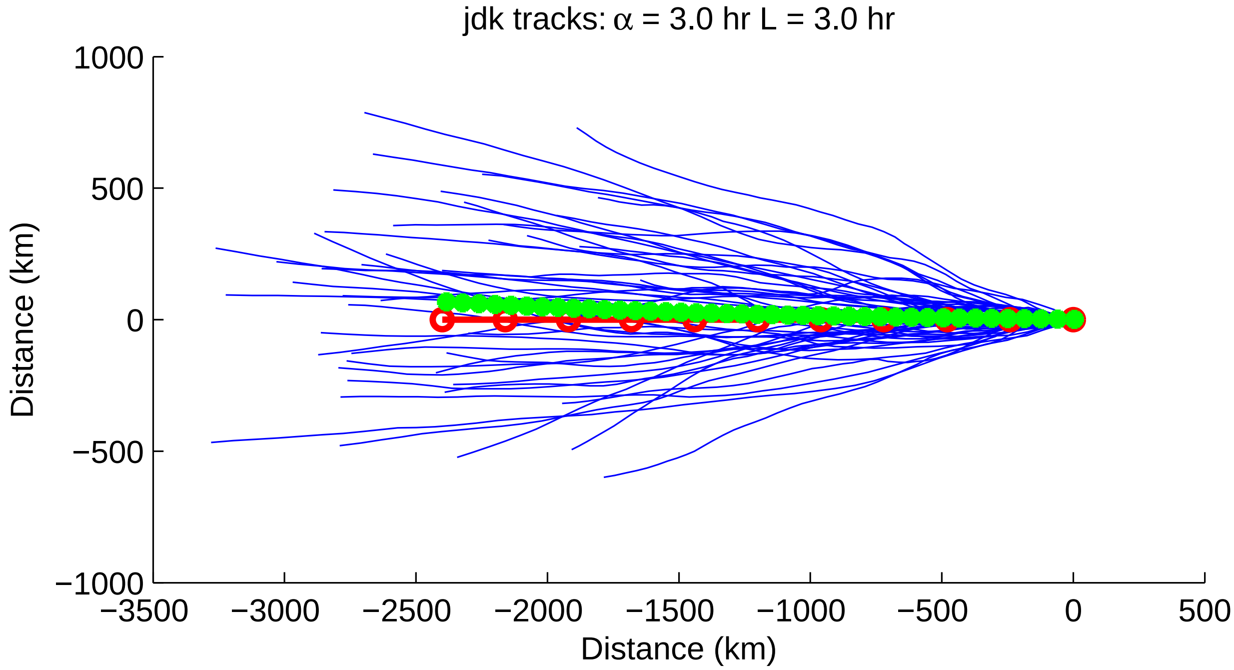


FIG. 5. As for Fig. 1, except calculated using the new method with $\alpha = 3\,\mathrm{hr}$, $L = 3\,\mathrm{hr}$.

dent Gaussian random numbers of unit variance, and let

$$\mathbf{y} = \mathbf{B}^{1/2}\mathbf{x} + \mathbf{c} \tag{5}$$

where $\mathbf{c}$ is the mean deviation of the track ensemble from the input track in DeMaria et al.'s (2009) method (i.e. the difference between the red and green curves in Fig. 1). The expected value of $\mathbf{y}$ is

$$\langle \mathbf{y} \rangle = \mathbf{B}^{1/2}\langle \mathbf{x} \rangle + \mathbf{c} = \mathbf{B}^{1/2}\mathbf{0} + \mathbf{c} = \mathbf{c} \tag{6}$$

where the angle braces $\langle \cdot \rangle$ denote the expected value and $\mathbf{0}$ is a vector of zeros. The covariance matrix of $\mathbf{y}$ is

$$\begin{aligned} \langle (\mathbf{y}-\mathbf{c})(\mathbf{y}-\mathbf{c})^T \rangle &= \langle \mathbf{B}^{1/2}\mathbf{x}\mathbf{x}^T\mathbf{B}^{1/2} \rangle \\ &= \mathbf{B}^{1/2}\langle \mathbf{x}\mathbf{x}^T \rangle \mathbf{B}^{1/2} \\ &= \mathbf{B}^{1/2}\mathbf{I}\mathbf{B}^{1/2} \\ &= \mathbf{B}. \end{aligned} \tag{7}$$

so $\mathbf{y}$ has the desired properties.

Figure 5 shows a track ensemble generated using the new method. Clearly, the tracks are much smoother than in Fig. 1, but the ensemble mean (green curve) is similar – in fact, it is identical apart from sampling error. Figure 6 shows the marginal pdf of along- and cross-track position every 12 hours for this method and DeMaria et al.'s (2009), and it is clear that the distributions are virtually unchanged.

## 4. Discussion

A new method for calculating a synthetic tropical track ensemble has been described. By construction, it yields results which have very similar marginal pdfs at each time to the method described by DeMaria et al. (2009), and hence will not interfere with the current application, namely of predicting wind probabilities. However, it also yields smoother tracks, without the kinks that are apparent in ensembles generated using the DeMaria et al. (2009) method. We expect that these will be more suitable for other applications, specifically for providing the atmospheric forcing to a storm surge model.

The new method may have some further advantages. The statistical model is a relatively

trast to the above useage) an element-wise square root of the diagonal matrix $\mathbf{\Sigma}$.

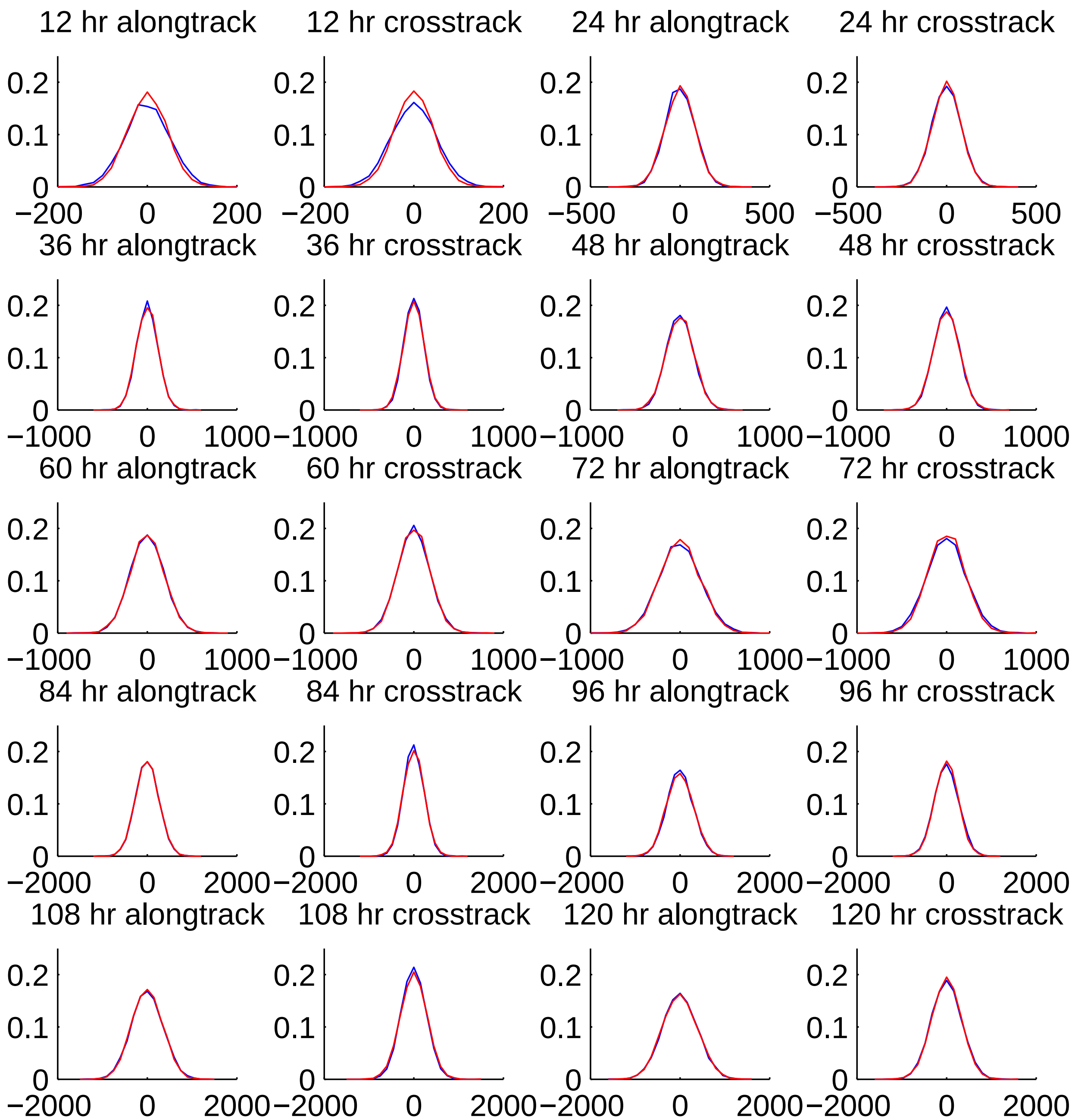


FIG. 6. The marginal pdfs of the along-track and cross-track position perturbations, calculated every 12 hours for 10000-member ensembles according the the method of DeMaria et al. (2009) (blue) and as described in the text (red).

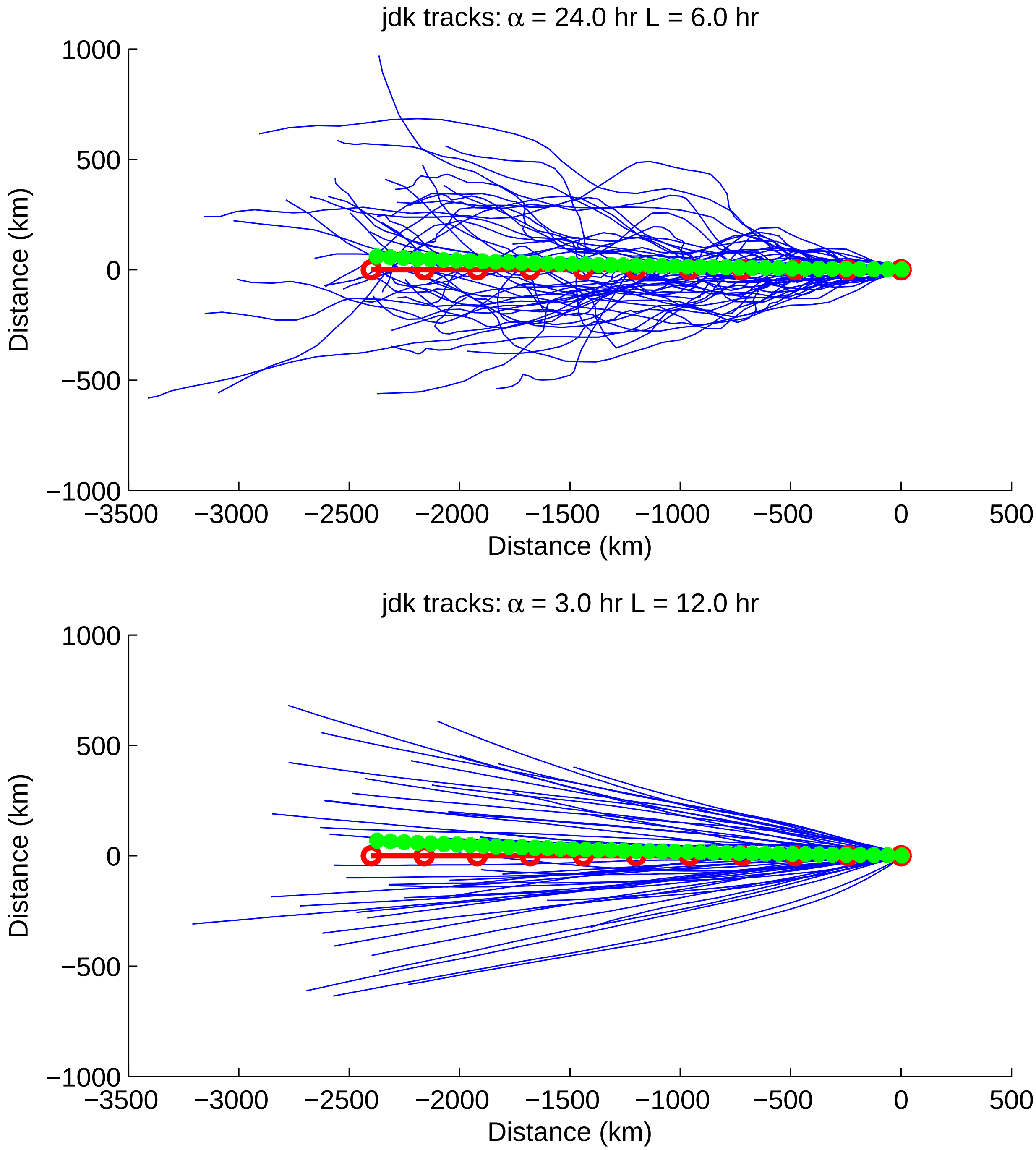


FIG. 7. Two somewhat unrealistic examples illustrating the flexibility of the new model. Top: weak temporal correlation leading to quite "wriggly" tracks ($\alpha = 24\,\text{hr}$, $L = 6\,\text{hr}$). Bottom: strong temporal correlation ($\alpha = 3\,\text{hr}$, $L = 12\,\text{hr}$). Both were calculated at 3 hr, rather than 12 hr, intervals.

parsimonious model, with many fewer parameters than the original. It should thus be easier to estimate these parameters, and less training data will be needed. In addition, the parameters have a clear physical meaning. We expect that they could be estimated from a numerical weather prediction ensemble prediction system, opening the possibility of using "errors of the day" in generating the ensemble. For this application, they could also be blended with climatology before use, perhaps obtaining the advantages of both. The requirement for less training data also opens up the possibility of more frequent updates to the parameters. It may even prove advantageous to use different parameter sets for different geographical regions, cyclone intensities (many forecasters regard intense storms as generally easier to forecast than weak ones), times of the year, or even pre- and post-recurvature.

Another advantage is that the new model is continuous in time, rather than the fixed 12-hour intervals in DeMaria et al. (2009). Generating more frequent data will increase the size of the matrices, but hourly intervals out to 120 hours would still only require operations on matrices which are several hundred elements square, easily manageable on modern desktop computers. There is thus no need for an arbitrary time interpolation, and more frequent data may allow for more accurate modelling in situations of rapid change such as at landfall, a possibility we aim to explore further.

The covariance model allows for a large degree of flexibility. For example, the two ensembles in Fig. 7 have the same spread at each time, but different correlation structures. Reducing the time-correlation leads to "wigglier" tracks, while increasing it results in quite linear perturbation growth. While neither of these examples is particularly realistic, they do illustrate the range of possibilities.

In summary, the new method preserves the good features of the method described by DeMaria et al. (2009) and used operationally in the USA and Australia, but produces smoother tracks which are likely more suitable for forcing a storm surge ensemble. The next step in this research is to expand the analysis described here to include the other components of the DeMaria et al. (2009) method, with the aim that this should eventually become part of the Australian Bureau of Meteorology's planned new storm surge prediction system.

## REFERENCES

DeMaria, M., J. A. Knaff, R. Knabb, C. Lauer, C. R. Sampson, and R. T. DeMaria, 2009: A New Method for Estimating Tropical Cyclone Wind Speed Probabilities. *Weather and Forecasting*, **24**, 1573 1591.

Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.