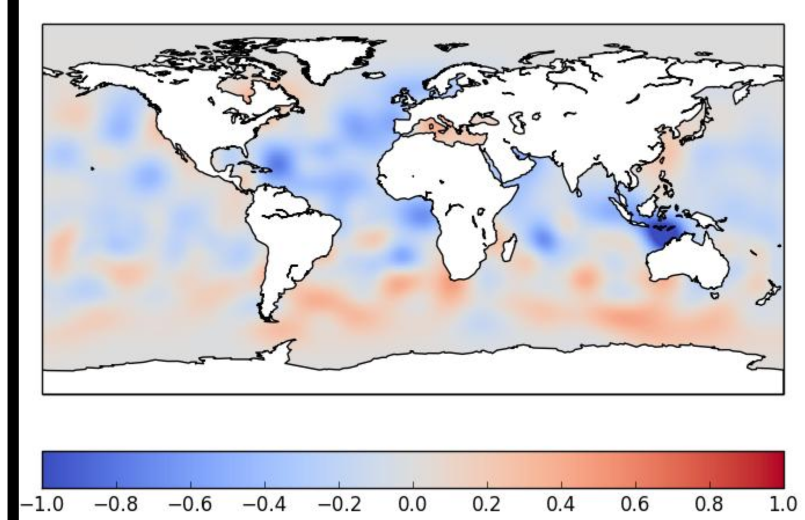
Variational Bias Correction of Sea Surface Temperature Observations in Ocean Data Assimilation

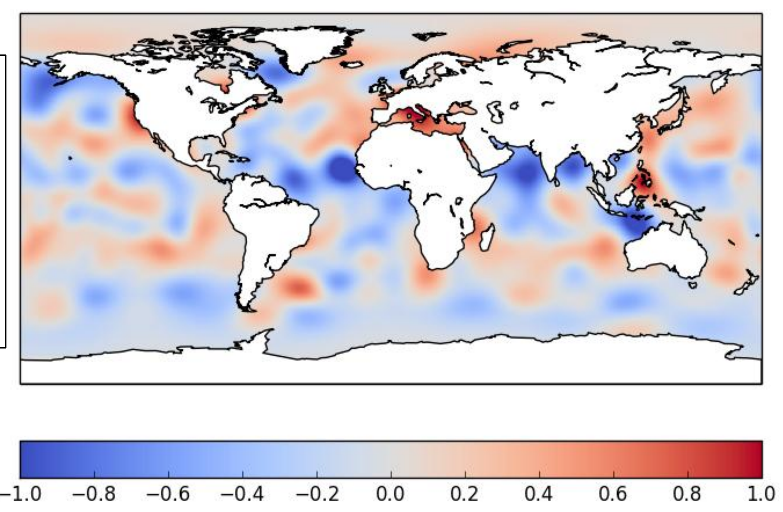
James While, Matthew Martin

1. Introduction

As generally formulated, data assimilation systems assume that observations are unbiased. In reality this is not the case. In particular, satellite observations of Sea Surface Temperature (SST) are expected to contain systematic error principally due to atmospheric contamination of the radiative signal of the sea surface. Such biases change over time and are typically of the order of 1°C.



Offline estimate of the SST bias (°C) for the NOAA-AVHRR (left) and METOP-AVHRR (right) instruments for 5th July 2016



As part of the ERA Clim2 project, the Met Office is developing a new variational technique for bias correction of SST. A key component of this new system is the use of 'observations of bias' to constrain the bias calculation.

This poster describes the theoretical aspects of our methodology and also some test results using the Lorenz 63 model. Work to implement the bias correction system operationally, using the NEMOVAR assimilation framework, is currently ongoing.

2. Variational Bias Correction Method

In our variational bias correction methodology we modify the standard cost function to become:

$$\begin{aligned} J &= 0.5(\mathbf{x}-\mathbf{x^f})^{\mathrm{T}}\hat{\mathbf{B}}^{-1}(\mathbf{x}-\mathbf{x^f}) + 0.5(\mathbf{y}-(\mathbf{H_y x}+\mathbf{H_b b}))^{\mathrm{T}}\hat{\mathbf{R}}^{-1}(\mathbf{y}-(\mathbf{H_y x}+\mathbf{H_b b})) \\ &\quad +0.5(\mathbf{b}-\mathbf{b^f})^{\mathrm{T}}\hat{\mathbf{O}}^{-1}(\mathbf{b}-\mathbf{b^f}) + 0.5(\mathbf{z}-\mathbf{H_z b})^{\mathrm{T}}\hat{\mathbf{L}}^{-1}(\mathbf{z}-\mathbf{H_z b}). \end{aligned}$$

Where:

$\mathbf{x}$:-	state vector	$\mathbf{x^f}$:-	background (forecast) vector
$\mathbf{b}$:-	state bias vector	$\mathbf{b^f}$:-	background (forecast) bias vector
$\mathbf{y}$	observation vector	$\mathbf{z}$:-	observations of bias vector
$\hat{\mathbf{B}}$:-	prescribed background error covariance	$\hat{\mathbf{R}}$:-	prescribed observation error covariance
$\hat{\mathbf{O}}$:-	prescribed bias error covariance	$\hat{\mathbf{L}}$:-	prescribed 'obs of bias' error covariance
$\mathbf{H_y}, \mathbf{H_b}, \mathbf{H_z}$:-	observation operators		

In deriving this equation we assume that model bias has already been accounted for and is negligible.

Minimising J provides both an analysis of **x** and an analysis of the bias **b.**

A key feature of the bias correction system is the presence of 'observations of bias' **z.** These help to constrain the bias analysis. In practice observations of bias will be the difference between matchups of co-located biased and (assumed) unbiased observations. We assume that most satellite data are biased, but that in-situ data and data from duel view satellite instruments are unbiased.

To prevent double counting. Observations used to calculate the matchups in **z**, will not be used as direct observations in **y**.

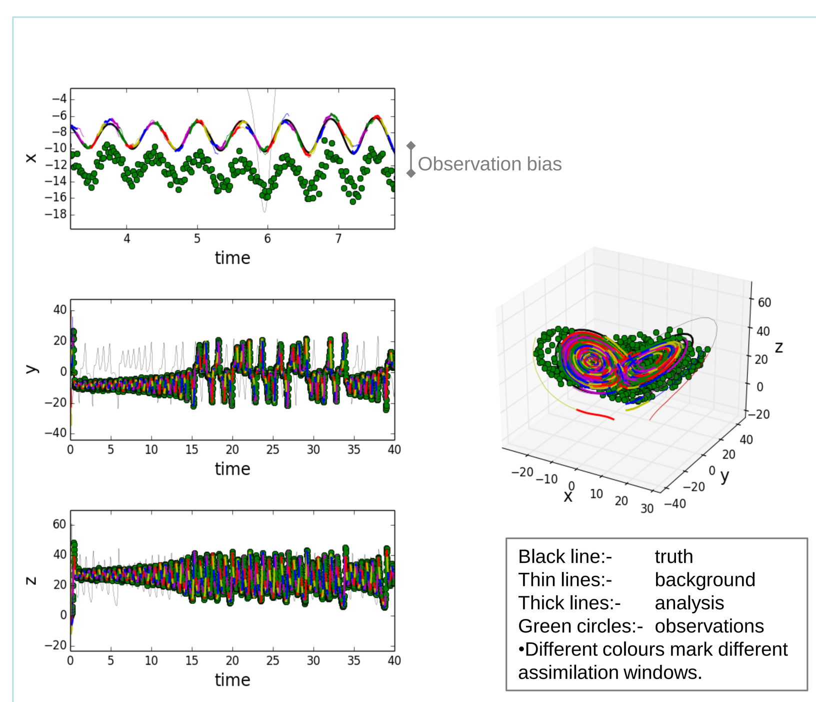
3. Theoretical errors with bias correction

For assimilation of biased observations into a linear system with no model bias it is possible to calculate theoretical expressions for the analysis error covariance (i.e. E[($\mathbf{x^a}$-$\mathbf{x^t}$) ($\mathbf{x^a}$-$\mathbf{x^t}$)$^{\mathrm{T}}$], where $\mathbf{x^a}$ is the analysis and $\mathbf{x^t}$ is the 'true' state of the system).

For a number of different bias correction methodologies the errors are as follows:

Bias correction methodology and cost function	Analysis error covariance expression
No bias correction, $J = 0.5(\mathbf{x}-\mathbf{x^f})^{\mathrm{T}}\hat{\mathbf{B}}^{-1}(\mathbf{x}-\mathbf{x^f}) + 0.5(\mathbf{y}-\mathbf{H_y x})^{\mathrm{T}}\hat{\mathbf{R}}^{-1}(\mathbf{y}-\mathbf{H_y x})$	$E[(\mathbf{x^a}-\mathbf{x^t})(\mathbf{x^a}-\mathbf{x^t})^{\mathrm{T}}] = (\mathbf{I}-\hat{\mathbf{K}}\mathbf{H_y})\mathbf{B}(\mathbf{I}-\hat{\mathbf{K}}\mathbf{H_y})^{\mathrm{T}} + \hat{\mathbf{K}}\mathbf{R}\hat{\mathbf{K}}^{\mathrm{T}} + \hat{\mathbf{K}}\mathbf{H_y b^t}(\hat{\mathbf{K}}\mathbf{H_y b^t})^{\mathrm{T}}$ If **b**$^{\mathbf{t}}$**=0** then this is the expected error of the Kalman Filter, and gives the smallest error of any of the bias correction methods Unfortunately errors are directly proportionally to **b**$^{\mathbf{t}}$ and rise rapidly as the bias increases.
Offline bias correction using observations of bias. This involves the sequential minimisation of two cost functions:- one for the bias (J_b), and then one for the state (J_x). This is what is currently done at the Met Office for SST data. $J_b = 0.5(\mathbf{b}-\mathbf{b^f})^{\mathrm{T}}\hat{\mathbf{O}}^{-1}(\mathbf{b}-\mathbf{b^f}) + 0.5(\mathbf{z}-\mathbf{H_z b})^{\mathrm{T}}\hat{\mathbf{L}}^{-1}(\mathbf{z}-\mathbf{H_z b})$ $J_x = 0.5(\mathbf{x}-\mathbf{x^f})^{\mathrm{T}}\hat{\mathbf{B}}^{-1}(\mathbf{x}-\mathbf{x^f}) + 0.5(\mathbf{y}-\mathbf{H_y}(\mathbf{x}+\mathbf{b^a}))^{\mathrm{T}}\hat{\mathbf{R}}^{-1}(\mathbf{y}-\mathbf{H_y}(\mathbf{x}+\mathbf{b^a}))$	$E[(\mathbf{x^a}-\mathbf{x^t})(\mathbf{x^a}-\mathbf{x^t})^{\mathrm{T}}] = (\mathbf{I}-\hat{\mathbf{K}}\mathbf{H_y})\mathbf{B}(\mathbf{I}-\hat{\mathbf{K}}\mathbf{H_y})^{\mathrm{T}} + \hat{\mathbf{K}}\mathbf{R}\hat{\mathbf{K}}^{\mathrm{T}} + \hat{\mathbf{K}}\mathbf{H_y}(\mathbf{I}-\hat{\boldsymbol{\Upsilon}}\mathbf{H_z})\mathbf{O}(\mathbf{I}-\hat{\boldsymbol{\Upsilon}}\mathbf{H_z})^{\mathrm{T}}\mathbf{H_y}^{\mathrm{T}}\hat{\mathbf{K}}^{\mathrm{T}} + \hat{\mathbf{K}}\mathbf{H_y}\hat{\boldsymbol{\Upsilon}}\mathbf{H_z}\mathbf{L}\mathbf{H_z}^{\mathrm{T}}\hat{\boldsymbol{\Upsilon}}^{\mathrm{T}}\mathbf{H_y}^{\mathrm{T}}\hat{\mathbf{K}}^{\mathrm{T}}$ This method uses some of the information in the observations to estimate the bias. As a result the solution is degraded when **b**$^{\mathbf{t}}$ is small The benefit is that the error is no longer proportional to **b**$^{\mathbf{t}}$. It is instead related to the error covariances **O** and **L**.
Variational bias correction. $J = 0.5(\mathbf{x}-\mathbf{x^f})^{\mathrm{T}}\hat{\mathbf{B}}^{-1}(\mathbf{x}-\mathbf{x^f}) + 0.5(\mathbf{y}-(\mathbf{H_y x}+\mathbf{H_b b}))^{\mathrm{T}}\hat{\mathbf{R}}^{-1}(\mathbf{y}-(\mathbf{H_y x}+\mathbf{H_b b})) +0.5(\mathbf{b}-\mathbf{b}^f)^{\mathrm{T}}\hat{\mathbf{O}}^{-1}(\mathbf{b}-\mathbf{b^f}) + 0.5(\mathbf{z}-\mathbf{H_z b})^{\mathrm{T}}\hat{\mathbf{L}}^{-1}(\mathbf{z}-\mathbf{H_z b}).$ **Definitions:** $\hat{\mathbf{K}} = (\hat{\mathbf{B}}^{-1}+\mathbf{H_y}^{\mathrm{T}}\hat{\mathbf{R}}^{-1}\mathbf{H_y})^{-1}\mathbf{H}^{y\mathrm{T}}\hat{\mathbf{R}}^{-1}$ $\hat{\boldsymbol{\Upsilon}} = (\hat{\mathbf{O}}^{-1}+\mathbf{H_y}^{\mathrm{T}}\hat{\mathbf{L}}^{-1}\mathbf{H_y})^{-1}\mathbf{H_y}^{\mathrm{T}}\hat{\mathbf{L}}^{-1}$ $\hat{\boldsymbol{\Lambda}} = (\hat{\mathbf{O}}^{-1}+\mathbf{H_y}^{\mathrm{T}}\hat{\mathbf{R}}^{-1}\mathbf{H_y}+\mathbf{H}_\kappa{}^{\mathrm{T}}\hat{\mathbf{L}}^{-1}\mathbf{H}_\kappa)^{-1}$ $\hat{\boldsymbol{\Psi}} = (\mathbf{B}^{-1}+\mathbf{H_y}^{\mathrm{T}}\hat{\mathbf{R}}^{-1}\mathbf{H_y}+\mathbf{H_y}^{\mathrm{T}}\hat{\mathbf{R}}^{-1}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{R}}^{-1}\mathbf{H_y})^{-1}\mathbf{H}^{y\mathrm{T}}\hat{\mathbf{R}}^{-1}$ A hat (^) symbol denotes that a matrix is a suboptimal estimate of the true matrix. The absence of the hat denotes the true matrix.	$E[(\mathbf{x^a}-\mathbf{x^t})(\mathbf{x^a}-\mathbf{x^t})^{\mathrm{T}}] = (\mathbf{I}-\hat{\boldsymbol{\Psi}}\mathbf{H_y}+\hat{\boldsymbol{\Psi}}\mathbf{H_y}\hat{\boldsymbol{\Lambda}}\mathbf{H_y}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H_y})\mathbf{B}(\mathbf{I}-\hat{\boldsymbol{\Psi}}\mathbf{H_y}+\hat{\boldsymbol{\Psi}}\mathbf{H_y}\epsilon^y\mathbf{H_y}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H_y})^{\mathrm{T}} + (\hat{\boldsymbol{\Psi}}-\hat{\boldsymbol{\Psi}}\mathbf{H_y}\hat{\boldsymbol{\Lambda}}\mathbf{H_y}^{\mathrm{T}}\mathbf{R}^{-1})\mathbf{R}(\hat{\boldsymbol{\Psi}}-\hat{\boldsymbol{\Psi}}\mathbf{H_y}\hat{\boldsymbol{\Lambda}}\mathbf{H_y}^{\mathrm{T}}\mathbf{R}^{-1})^{\mathrm{T}} + (\hat{\boldsymbol{\Psi}}\mathbf{H_y}\hat{\boldsymbol{\Lambda}}\mathbf{H_y}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H_y}+\hat{\boldsymbol{\Psi}}\mathbf{H_y}\hat{\boldsymbol{\Lambda}}\mathbf{H_z}^{\mathrm{T}}\mathbf{L}^{-1}\mathbf{H_z}-\hat{\boldsymbol{\Psi}}\mathbf{H_y})\mathbf{O}\times(\hat{\boldsymbol{\Psi}}\mathbf{H_y}\hat{\boldsymbol{\Lambda}}\mathbf{H_y}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H_y}+\hat{\boldsymbol{\Psi}}\mathbf{H_y}\hat{\boldsymbol{\Lambda}}\mathbf{H_z}^{\mathrm{T}}\mathbf{L}^{-1}\mathbf{H_z}-\hat{\boldsymbol{\Psi}}\mathbf{H_y})^{\mathrm{T}} + (\hat{\boldsymbol{\Psi}}\mathbf{H_y}\hat{\boldsymbol{\Lambda}}\mathbf{H_z}^{\mathrm{T}}\mathbf{L}^{-1})\mathbf{L}(\hat{\boldsymbol{\Psi}}\mathbf{H_y}\hat{\boldsymbol{\Lambda}}\mathbf{H_z}^{\mathrm{T}}\mathbf{L}^{-1})^{\mathrm{T}}.$ For variational bias correction without 'obs-of-bias', the terms depending on **L** should be neglected This method uses even more information in the observations to estimate the bias. Consequently it has the largest errors when **b**$^{\mathbf{t}}$ is small However, the level of error due to **b**$^{\mathbf{t}}$ and its covariances are down weighted relative to the offline bias correction method. Consequently for large biases this method is expected to produce the best results

4. Tests with Lorenz 63 system

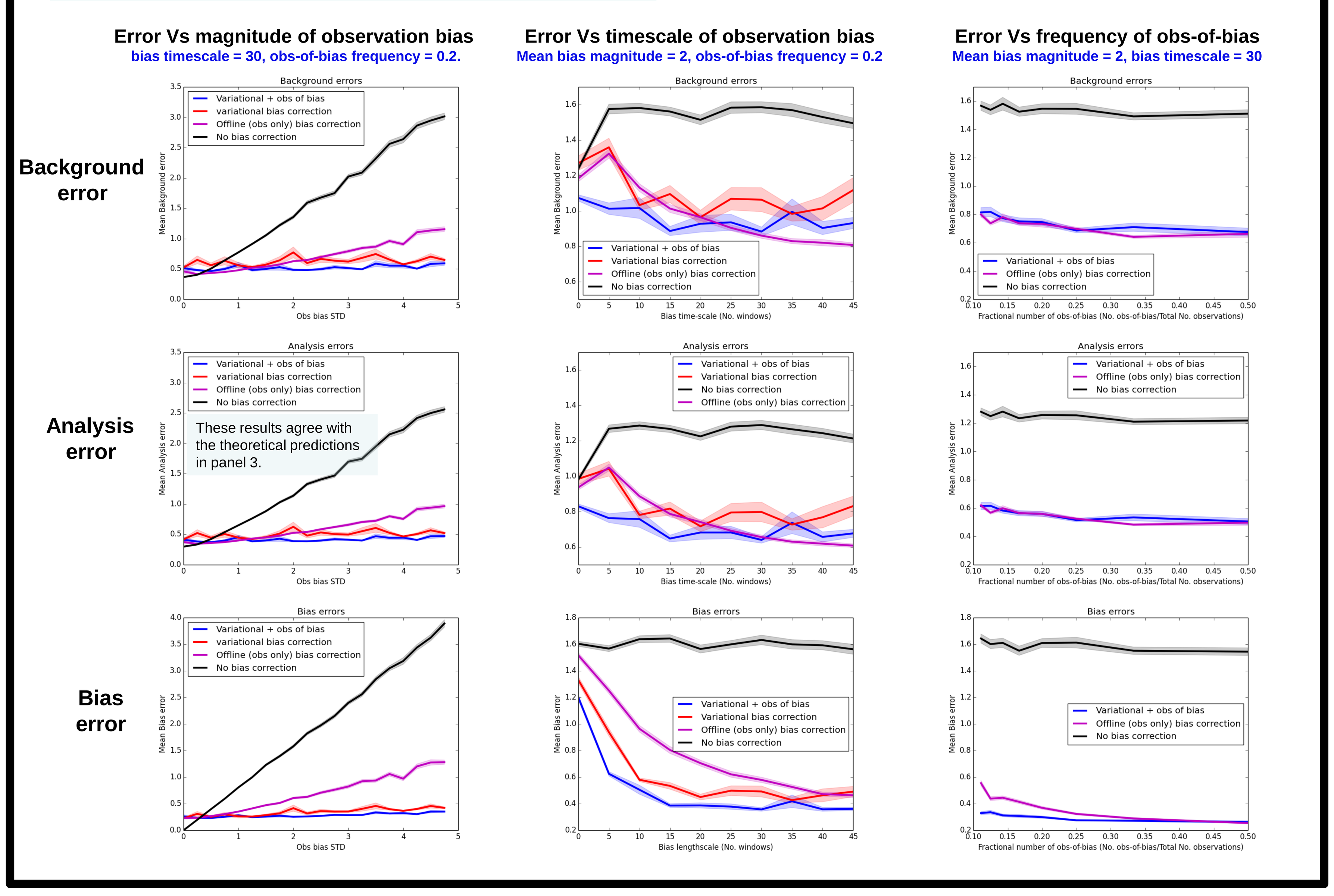


We have tested our variational bias correction system in a 3DVar context using the Lorenz 63 model.

Noisy observations are assimilated into all 3 axes, but only observations in the x-axis are biased.

The observation bias is not constant, but instead slowly varies on a user determined time-scale.

For each data point in our experiments, an ensemble of 100 model runs with perturbed parameters and initial state were used to calculate the statistics.



Met Office Fitzroy Road, Exeter, Devon, EX1 3PB United Kingdom
Tel: 01392 885680 Fax: 01392 885681
Email: james.while@metoffice.gov.uk