School of Mathematical
and Physical Sciences

Department of Mathematics and Statistics

Preprint MPS-2012-01

07 February 2012

Integration of a 3D Variational data assimilation scheme with a coastal area morphodynamic model of Morecambe Bay

by

G.D. Thornhill, D.C. Mason, S.L. Dance,
A.S. Lawless, N.K. Nichols and H.R. Forbes

Integration of a 3D Variational data assimilation scheme with a coastal area morphodynamic model of Morecambe Bay

G. D. Thornhill[1]*, D. C. Mason[1], S. L. Dance

1. Introduction

Understanding the processes acting in near-shore environments is increasingly important as the potential risk to coastal environments from the combination of sea-level rise and more frequent and intense storms resulting from global climate change needs to be assessed. In many coastal areas the local conditions can change rapidly, and tracking these changes is crucial for assessing flood risks, environmental impacts, safety for shipping and recreational uses of the coastal zones. In areas where the bathymetry evolves rapidly over time the collection of data from surveys and in-situ measurements with the frequency required to monitor the changes is not feasible due to cost and logistical difficulties. Newer approaches have used remote sensing to provide data over a larger area than local surveys typically provide, and allow the overall changes in local near-shore and inter-tidal bathymetry to be measured over the time-frame of the observations (e.g. Mason et al., 2010). In order to compensate for the general scarcity of data, complex computational morphodynamic models have been developed, including 3D morphodynamic models (e.g. Lesser et al 2004). However, the ability of these models to predict the state of the near shore environment can be limited (de Vriend et al 1993, Sutherland et al 2004b) due to inadequate modelling of all the hydrodynamic processes acting, and imperfect knowledge of the initial conditions. The integration of remotely-sensed data with morphodynamic models offers the possibility to make predictions which are more accurate than using the models alone, and provide more insight into the evolution of the near-shore zones than intermittent data collection alone. Previous work using data assimilation to improve bathymetric modelling includes the Beach Wizard assimilation system, which used a combination of wave roller dissipation and inter-tidal bathymetry derived from video sources and radar-derived wave celerity as data sources. These were incorporated into the model using an optimal least squares estimator to update the model (van Dongeren 2008), which resulted in improved skill for their predictions of near-shore bathymetry.

The work described here follows from the previous work of Scott and Mason (2007), where a simple coastal area morphodynamic model was coupled with an Optimal Interpolation (O.I.) data assimilation scheme. The objective of the current work is to implement a more powerful avtionar and the Isle of Man and the spits at the mouth of the bay limit wave size by restricting fetch (Mason et al 1999). The duration of the semi-diurnal flood and ebb tides are unequal, with the ebb running for about 40 min longer than the flood at Heysham (Coomber and Hansom 1994). The spring tide attains a maximum velocity of $1.5 ms^{-1}$ in the large sub-tidal channels, with currents being higher on the flood than the ebb. The sediments in the inter-tidal zone are predominantly composed of very fine and fine sand (0.06   0.2 mm), with coarser sand and fine gravel at the mouth of the bay and silts in the inner bay (SMP, 1996). The channels and sandbanks in the bay constantly migrate and in the case of the Ulverston Channel there is evidence for migration of the order of 5

3. Datasets

An extensive data set of inter-tidal bathymetry for Morecambe Bay was assembled for the study, b inter-tidal bathymetry, and do not provide observations across the whole of the inter-tidal zone. They are also irregular in terms of the temporal resolution, in some cases there are only a few hours between waterline observations, and in other cases a gap of up to 6 months may occur.

Fig. 2: Datasets: a)

laser altimetry (LiDAR) survey that covered the entire inter-tidal zone of the bay. The spatial resolution was 2m, and the error on the mean heights was found to be ±5 cm when compared with independently measured heights in urban areas around the bay. The complete data set included almost 200 million height samples. Such surveys are rare due to the difficulty of over-flying the bay at low tide, and the cost of acquiring and processing the data. These data were used solely for validation purposes in this study.

4. Data Assimilation

Data assimilation is a technique widely used in environmental science to combine information from a variety of sources of observational data with dynamic models. The purpose of this is to improve a model's predictive ability, and to permit a rigorous assessment of the relative accuracy of the model and the observations. A forecast is produced (referred to as the 'background') by running the model to the point where observations are available, then observations valid at the simulated time of the forecast are the model state and the observations and initial state:

where:

z is the model state vector (dimension n, where n is the number of model grid cells)
z_b is the background state vector (dimension n)
y is the observation vector (dimension p, with $p << n$)
H is the observation operator that maps model variables to observation space (in this case a simple linear interpolation from the model grid to the locations of the observations) (dimension p x n)
B is the covariance matrix of background errors (dimension n x n)
R is the covariance matrix of observation errors (dimension p x p)

This cost function is minimized with respect to z, and the vector z_a where the function is a minimum is referred to as the 'analysis'. The vector z_b is the background vector, representing the bathymetry predicted by the model. The vector z is usually taken to be the same as the background vector for the first iteration. In our case the observation operator **H** simply interpolates nearby model bathymetry values to the observation position. The matrix **B** is an error covariance matrix which represents the errors), and **R** is the error covariance matrix for the observation errors. Choosing appropriate values and structures for these covariance matrices is important for the correct weighting between the observations and the background. In our implementation, the minimisation is done using a conjugate gradient method performed to calculate the cost function above. The scaling factor must be calculated such that the maximum element of

$$, \qquad (5)$$

where

$$A_s = (A_{ss}+A_{sb})F \quad , \qquad (6)$$

$$A_{ss} \quad \frac{0.012 d_{50} \quad D^{0.6}}{[(s \quad 1) g d_{50}]^{1.2}} \quad , \qquad (7)$$

$$A_{sb} \quad \frac{0.005 h (d_{50}/h)^{1.2}}{[(s \quad 1) g d_{50}]^{1.2}} \quad , \qquad (8)$$

$$D \quad \frac{g(s \quad 1)}{^2} \; ^{1/3} d_{50} \quad , \qquad (9)$$

A_{ss} = suspended sediment load
A_{sb} = bed sediment load
u = depth-averaged current
u_{cr} = critical current for sediment movement
h = water depth
d_{50} = median grain diameter

5.2 Initial bathymetry

In general it may be difficult to provide a correct initial bathymetry in most operational cases and the most likely situation would be an imperfect or out-of-date bathymetry for the beginning of a model run. In this case, we wanted to provide the best estimate of the likely bathymetry at the start of our model run thus reducing the probable errors for the initial conditions and providing a more stringent test for the comparison of predictions with and without assimilation. The differences between the previously available bathymetry valid for 1997 (Mason et al 1999) and the LiDAR data for November 2005 show that substantial changes in the channel positions had occurred over this period. In order to update the 1997 data to provide a new initial state for 2004, we ran the model using the earlier data and assimilated the waterlines available for May 2003 December 2004. This was done using default parameters for the model and assimilation, and was the simplest method for updating the old bathymetry to a state which would be more likely to for the beginning of our model runs in January 2004. Fig. 3(a) and (b) show the differences between the 1997 bathymetry and our updated initial bathymetry.

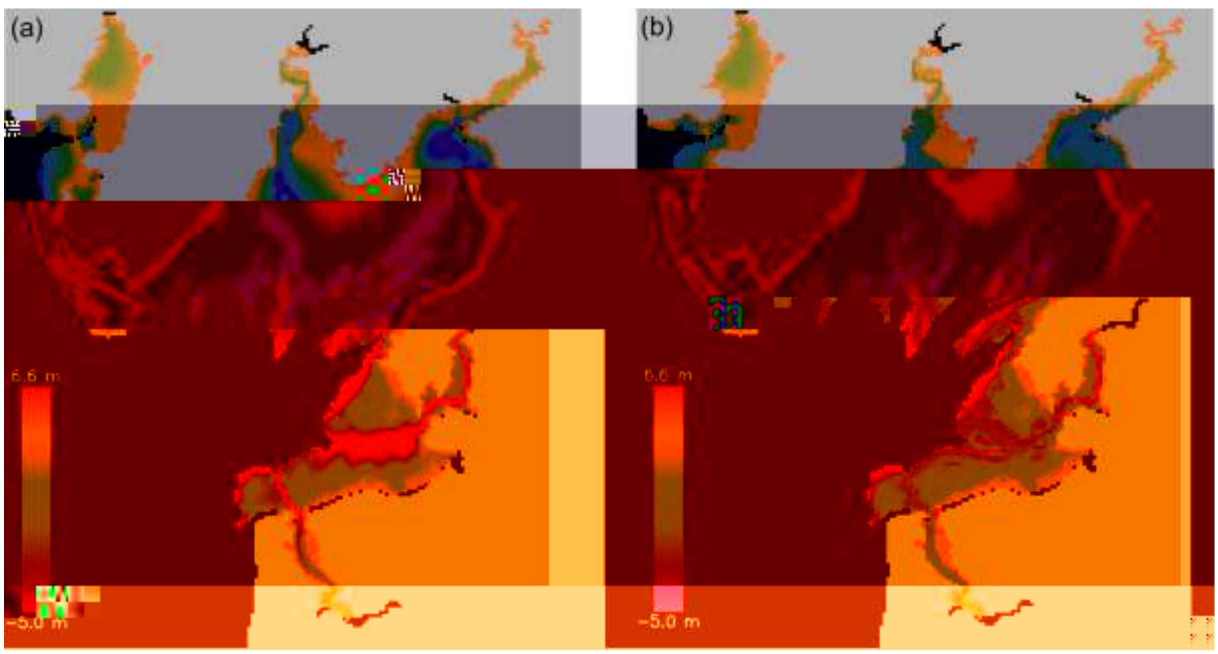

Fig. 3: (a) Bathymetry for 1997 (after Mason et al 1999) (b) Initial bathymetry created as initial bathymetry for January 2004

6. Calibration of the Assimilation System

The system needs to be calibrated for the sediment transport parameter F, and also to find the best value of the covariance length scale to produce the best results for the assimilation scheme. This was done by changing the values of the parameters over a reasonable range, and using performance measures to select the values providing the best predictions of the bathymetry.

6.1 Performance measures for the calibration

The calibration was carried out against the swath bathymetry from February 2005, as this covered a larger part of the domain than the individual waterlines. The model was run to the time of the swath dataset, and then the predicted bathymetry was compared to the swath bathymetry wherever they overlapped, using interpolation to match the model points to the data. Measures of model performance were calculated so that the impact of different values of model parameters could be assessed. The main measures calculated from the set of height differences were the mean difference, the Mean Square Error (MSE), the variance, and the Brier skill score (Sutherland et al 2004a). The Mean

Fig. 4: Background analysis showing the effect of different length scales: (a) $l = 0.5$ (b) $l = 3.0$

6.3 Calibration of sediment transport factor F

The Soulsby-van Rijn equation (5) is used to calculate the sediment transport and includes a calibration factor F. Several runs of the model were carried out using different values of F, ranging from F = 0.1 to

2.0	2.0	0.5	-0.090	2.330	2.320	0.311
2.0	2.0	1.0	-0.170	2.090	2.060	0.382
2.0	2.0	2.0	-0.360	2.430	2.300	0.281
2.0	2.0	3.0	-0.400	2.770	2.610	0.181
2.0	0.8	1.0	-0.170	2.150	2.120	0.151

*Fig. 5: Plots of differen-7(i3.48 ref*W*n0.54 5F)-2(ig)]TJETBT1 0 0 1 98.905(ts)4( )Pn0.54 5thaeWhots of271 09 Tm[(P)-2(l)13(o)-5(t*

P3904 140.88 Tm[(.)8107.54 653.26 47360re1IE717 639.

${}_b^2$ = 2.0 after 23 months. The plot shows the bathymetry for ${}_b^2$ = 2.0 subtracted from the bathymetry for ${}_b^2$ = 0.8.

6.4 8p1 0-0078 Tc[6 8oTm]-4(8)7(b)BTF7 9.96 Tf2078(8)7(b)BThe b[.).96 TfkBTgrou v512.ariBT7(nc(th)-427 716 (8p1

(waterlines and swath data)				
With assimilation (excluding swath data)	0.480	2.300	2.080	0.262

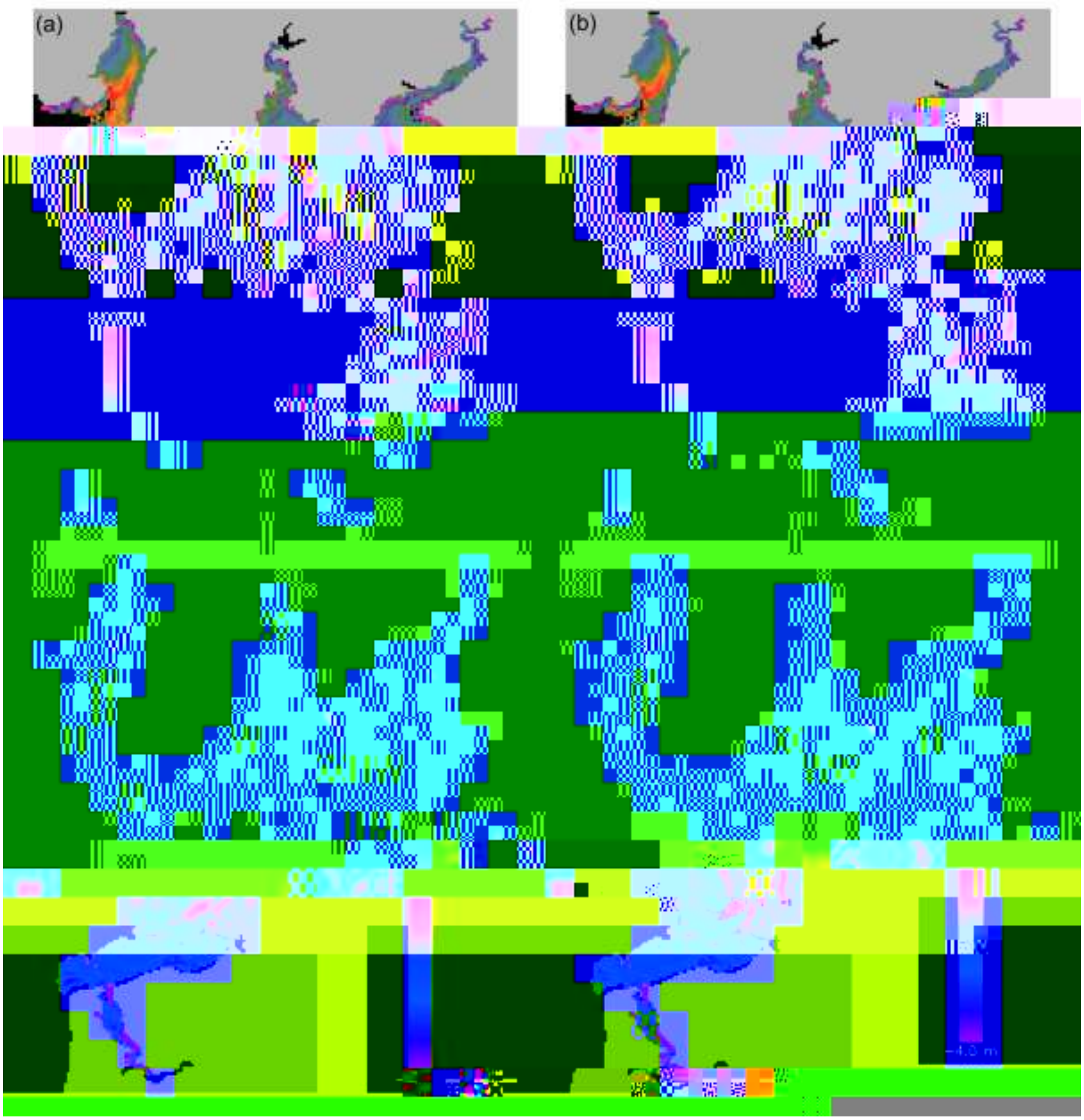

Fig. 6: Plots of differences between predicted bathymetry and bathymetry measured by LiDAR (prediction data). The letters in each plot are referred to in the text.
(a) without assimilation (b) with assimilation. c) Differences between model output bathymetry and LiDAR data for model run excluding assimilation of swath data. d) Differences between weighted ensemble mean bathymetry and LiDAR data.

The advantage of the reduced chi-squared compared to the normal chi-squared test is that it normalizes for the number of degrees of freedom.

The variance (^{2} cell is taken as the weighted standard deviation of the model runs squared, added to the variance for the sub-sampled LiDAR observations (0.05 m^2). This gives us the combined variance of the difference between the observation and the combined model runs at each grid cell. On the assumption that the null hypothesis is true, the values of the predicted differences $f(x_i)$ are zero. In principle, a reduced chi-squared value of 1 indicates that the extent of the match between observations and predictions is in provide an improvement over the best individual run, and reduce the differences between the validation data and the predicted bathymetry. This may be due to combining the parameter values so that there is effectively a variable value of the parameters from the different individual runs providing a better overall match. A plot showing the differences between the LiDAR data and the weighted mean of the runs (Fig. 6(d)) shows where the ensemble prediction has improved over the nominal run shown in Fig. 6(b). The ensemble still does not correctly predict the deposition at the northern edge of the Ulverston Cha(ers)4(ed)-BT1 0 0

enTBT1 0 0 1 538.92 1 0 0 597[(s)-5(ho)4(w)pa(n)-9( )-10(f)hion trijnals-57()-3(u)91.944 654.7 e(j)] TJf-58(c)-53 R fohtoT

The changes in the bathymetry over the course of the 23 month run shown in Fig. 9(a) are based on comparing the initial bathymetry from January 2004 used for the model run, and the bathymetry as measured by the LiDAR data in November 2005. The LiDAR data are averaged to produce a single point for each model grid cell for this comparison and this is subtracted from the initial bathymetry. This is not a comparison of actual observed changes, as the initial bathymetry is a modelled bathymetry as described in Section 5.2. The comparison shows where the assimilation system should produce changes to the initial bathymetry in order to match the final LiDAR data. The LiDAR data is subtracted from the initial model estimated bathymetry, so areas appearing positive (red and orange) are where erosion has occurred, and the initial estimate is higher than the LiDAR data; conversely, the blue and purple areas show where there has been deposition over the model run period. The main changes are in the migration of the Ulverston Channel, which shows as the red and orange area between the Leven and Kent estuaries (A), and the purple areas to the south of this (B). Along the Kent estuary there has been considerable deposition, which extends across to Wharton Sands (C). The Lune Estuary (D) is also higher as measured by the LiDAR data, while the channels leading up to Wharton Sands are deeper (E). The differences between the initial bathymetry and the final model bathymetry produced by the assimilation scheme are shown in Fig. 9(b) for comparison, using the values of $F = 2.0$, $= 1.0$. If the assimilation system was perfect, the two figures would be identical. Although the two figures show changes occurring in similar areas, the results from the assimilation scheme do not show the larger areas of deposition along the Cartmel Wharf (A), and the removal of sediment along the Ulverston Channel has been over-estimated. The changes in the Lune Estuary are also not apparent in the assimilation system. This may be in part because the model itself is simplistic and is not reflecting fully the sediment transport processes active in the bay, particularly secondary currents which are responsible for meander migration. This means that changes in the channel positions are not well represented by the model alone. The use of the waterline assimilation should correct for this in part, but can only do so if the changes are adequately captured by the waterline data. Also the temporal spacing of the data assimilated can be large in some cases, and is certainly not regular. In some cases waterlines are separated only by a matter of hours, in others several months elapse between observations. Where the model runs without assimilation for a period of months, it may drift away from the correct bathymetry. If the next assimilation is a waterline that does not cover the areas where the change has been greatest, it may not be able to correct the model sufficiently over the required areas, leaving discrepancies between the true state of the bathymetry and the prediction from the assimilation scheme.

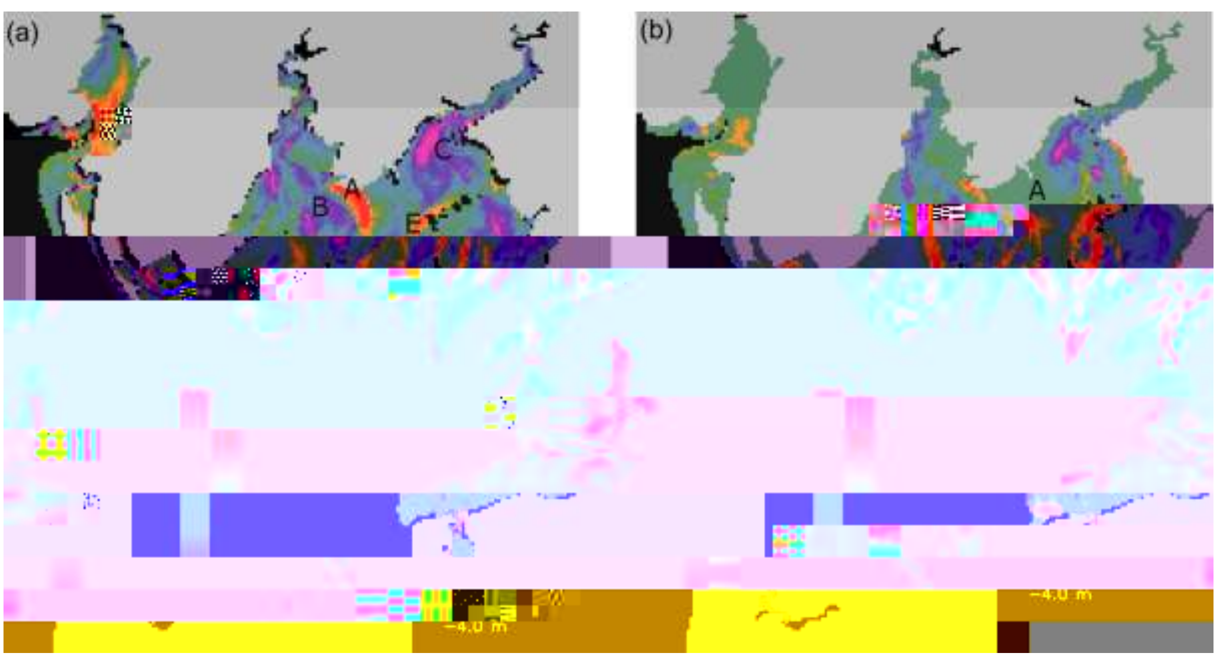

Fig. 9: Change in bathymetry over 23 months from initial model bathymetry: (a) compared to LiDAR data

(b) compared to final model bathymetry prediction. The letters in each plot are referred to in the text.

7.6 The impact of storms

Storms in Morecambe Bay may be expected to affect the bathymetry due to storm surges and increased wave action. This can result in more rapid changes in the bathymetry during a storm, and would result in the model being less effective at predicting these changes, as it includes neither storm surges nor wave action. If this assumption is correct, the assimilation system would be expected to provide a larger correction to the predicted bathymetry after a storm than after a relatively calm period. As the observations th39 T-311(A)4(2)-1 d6tiont Tc[(c)-5(-34(t)-10T(bo(the)-10T4s)-17(y)3prges)-4(m[(ti)-3(v)6(e)-95(on)A)4

9.

Conservation, Morecambe Bay. Coastal Research Group, University of Glasgow, 79pp.

de Vriend, H.J., 1987. 2DH mathematical modelling of morphological evolutions in shallow water. Coastal Engineering 11, 1-27.

de Vriend, H.J., Zyserman, J., Nicholson, J., Roelvink, J. A., Pechon, P., Southgate, H. N. 1993. Medium-term 2DH coastal area modelling. Coastal Engineering 21, 193-224.

Järvinen, H. 1998. Observations and diagnostic tools for data assimilation, http://www.ecmwf.int/newsevents/training/rcourse_notes/DATA_ASSIMILATION/OBS_AND_DIAG_TOOLS/Obs_and_diag_tools.html

Johnson, C., 2003. Informational content of observations in variational data assimilation. PhD Thesis, Department of Mathematics, University of Reading.

Johnson, C., Hoskins, B.J., Nichols, N. K., 2005. A singular vector perspective of 4D-Var: Filtering and interpolation. Quarterly Journal of the Royal meteorological Society, 131 part A, 1, 19.

Laub, C., Kuhl, T. L., 2005. How Bad is Good? A Critical Look at Iterative Fitting of Reflectivity Models using the Reduced Chi-Square Statis-10(os)-3(k)-194ta6Inftlterin