Chapter 4

Quantifying Study Error Correlations

Introduction

In ch-2 we explain the variational formulations of operations data adaptation algorithms, where the information provided by the observations and a first estimated model background is weighted by the inverse of their respective error covariance matrices. We discussed how the error characteristics of remotely sensed observation types are typically not perfectly express in data adaptation algorithms. So the effect on result is a negative impact on forecasts correctness and an inefficient non use of observations [14]. But quantifying observation error correlations is not a straight forward problem and they can only be estimated in a statistical sense, not observed directly. However trials have been made to quantify error correlation structures for different observation types like as atmospheric motion vectors and satellite radiances.

Two methods of investigated the correct covariance matrices using post analysis. We Described in section 2.6. The Hollingworth method is typically used when the background errors carry spatial correlations while the observation error does not.

The work we do in this chapter to answer the first thesis question posed in first chapter, what is the true structure of observation error correlations? Using infrared atmospheric sounding Interferometer (IASI) data as the observation type, we will apply the diagnostic proposed to quantify the cross channel correlations between measurements. First of all we will introduce the IASI instrument and describe the likely origin of IASI observation error correlations. We then apply the diagnostic to IASI measurements processed using the Met Office incremental 4D–Var data adaptation scheme. Technical and practical details of the process will be given, including the pre – processing of IASI observations through 1D Var retrievals. The diagnosed error for covariance will be produced for both the 1D var retrieval procedure and the 4D var adaptation. The results described have previously declare as a technical report.
4.1 Infrared Atmospheric Sounding Interferometer (ISAI) Data

IASI is the main payload device for the reason of sustaining Numerical Weather prediction. It provides information on the vertical structure of the atmospheric temperature and humidity in an unparalleled accuracy of 1 Km and a vertical resolution of 1 km, which is needed to positively improve NWP. The use of Metop data in NWP accounts for 40% of the impact of all space based observations in NWP forecasts.

Dieter Klaes, EPS Programme Scientist at EUMETSAT, explains: "When first launched on Metop-A, in 2006, IASI was a world-leading instrument. It was the first polar-orbiting interferometer providing huge amounts of operational data. The impact on NWP was huge, much superior than expected."

The IASI instrument is an infrared Fourier spectrometer which measures the infrared radiation release by the earth’s surface and atmosphere. The first IASI instrument was launched on the METOP A satellite in 2006 as part of the EUMETSAT European polar System (EPS). Its spectral interval of 645 – 2776 cm\(^{-1}\) with a spectral sampling of 0.25 cm\(^{-1}\), and divided into three bands and sampled by 8461 channels at a resolution of 0.5 cm\(^{-1}\). Band one from 645–1210 cm\(^{-1}\), is used primarily for temperature humidity and ozone sounding, band two 1210-2000 cm\(^{-1}\) for water vapor sounding and the retrieval of N\(_2\)O and CH\(_4\) (methane) column amounts, and band three 2000-2760 cm\(^{-1}\) for temperature sounding and the retrieval of N\(_2\)O and CO column amounts.

ISAI has been developed for operational meteorological sounding with a very high level of accuracy being devoted to improve medium range weather forecast. ISAI system includes the 3 instruments, data processing software integrated in the EPS ground segment.

ISAI measurement of radiances, \(r\) are expressed as black body equivalent brightness temperatures \(T\), through Planck’s functions
where \( k \) is Boltmann’s constant (define the relation between exact temperature and the kinetic energy) and \( h \) is planck’s constant, \( (h \approx 6.626176 \times 10^{-34} \text{ joule – seconds}) \), \( c \) is the speed of light and \( v \) is wave number. Planck’s function is used in the radiative transfer equations. The radiative transfer model used in the adaptation of ISAI radiances is the radiative transfer model.

ISAI data is an important element of the global observing system (GOS). The adaptation of ISAI radiances is operational at the Met Office, Meteo – France and ECMWF, and is in the testing stage at several other national weather centres. At the Met Office, forecast accuracy has improved through the adaptation of ISAI radiances from the general channel subset determine. However the adaptation of channels sensitive to water vapor has only shown a weak impact on forecast accuracy. The suspected cause of this underperformance is attributed partially to the mis-representation of the cross channel observation error correlation structure.

4.1.1 Observation Error Correlations

The IASI observation errors are treated as horizontally and vertically uncorrelated. The assumption of horizontally independent observation errors is supported by intelligent thinning of the data ensuring that no observations are adaptation at a higher density than model resolution. This is clearly a very inefficient use of the data. But it reduces the complexity of the subsequent adaptation of the radiances.

Ensuring vertically independent observation error is more difficult, because of the nature of IASI instrument; radiance measurements are sensitive to the temperature profile over several atmosphere levels. This distribution is expressed by the broad channel weighted functions of the instruments. So the errors in nearby channels (those close to each other in
wavelength) can potentially be correlated: for eg, if the sensitivity of the signal to trace gas present in several nearby channels is mis-represented. The current IASI channel selection procedure deals with this issue by avoiding the adaptation of nearby channels. However this cannot be carefully enforced because nearby channels in certain wavelength bands are needed to provide fine scale information on atmospheric profiles. For eg channels in the long wave CO$_2$ band provide information on temperature and humidity, therefore some level of error correlation structure will exist between selected channels.

![IASI weighting functions](image)

Figure 4.1: IASI weighting functions (Provided by Fiona Hitlton)

Additional correlated errors of representative are preset between channels that observe spatial scales features that the model can’t. Although the IASI observation spacing of 25 km is similar to the Met Office NWP model grid spacing of 40 km, IASI is sensitive to small variations within its 12 km field of view which the NWP model does not attempt to present. For example, the NWP model may be unable to represent.
So at the end, errors in the forward model may be correlated between channels. These include errors in the spectroscopy, a wrong criticize of the radiative transfer equation, and mis-representation of the gaseous contribution in certain channels.

### 4.1.2 Processing For Adaptation

Any preprocessing performed on the original IASI radiances prior to their adaptation is likely to generate errors. At the Met office before IASI observations are adapted directly into the NWP model, they are subject to prescreening and quality control procedures. This is performed in the observation processing system (OPS).

The observation processing system (OPS) reads the raw observations, manipulates the observations, performs quality control on the observations and reformats the observations ready for use in the numerical models.

There are two main components to the OPS (observation processing system termed) extract and process. The extract component retrieves the row data from the central observation database, identified as the MetDB, and calculates background values at the observation locations. The process factor then performs the quality control, converts the observations if necessary and then reformats them and ready for use in the operational forecast.

There are various facilities available to switch off a particular observation type in case a problem develops, such as corrupt data. Also, a record of the number of observations from each type is kept and if the expected number of observations is not received, the forecasters may be alerted as it may have implications on how the model forecast is interpreted.

After the forecast has finished the OPS is run again on the same set of observations. This time however, the analysis values are calculated at observation locations and this
information along with the results from the quality control is merged back into the MetDB so that monitoring of the observations can be performed. A schematic (diagram) of the IASI observations processing path is shown in figure 4.2. IASI has the potential to provide observations in so many channel (8461), but at present only observations from a subset of 314 are used. IASI measured brightness temperatures from this subset are fed into the OPS and processed using a code specifically written for satellite measurements. This code known as the SatRad code, implements a 1D Var adaptation on the bias-corrected brightness temperature measurements $y$, and an accurate first guess model profit from a short range forecast $x_b$. The solution is the state vector $x$ that minimizes the cost function,

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(y - h(x))^T R^{-1}(y - h(x)),$$  \hspace{1cm} (4.1)$$

where $h$ is the observation operator mapping from state space to measurement space, $B$ is the background error covariance matrix and $R$ is the observation covariance matrix. Operator $h$ is comprised of a Radiative Transfer for TIROS Ops Vertical Sound (RTTOV) radiative transfer model, it accurately predicts brightness temperatures given first-guess model fields of temperature and humidity on 43 fixed pressure levels between 0.1 and 1013 hpa, as well as surface air temperature, skin temperature and surface humidity.
The OPS has two main functions: the first is quality control on the brightness temperature measurements, and second is providing an accurate estimate of the model variables not adapted in 4D–Var. The adaptation performs a local analysis of the model state at the location of every satellite observation is appropriate for 4D-Var adaptation, if its 1D Var analysis generates a best convergence and a appropriate a posteriori cost. Each of the observation has a associated cost which is scaled to be ideally one, and if the distribution of the costs about 1 were plotted, the quality control procedure would be equivalent to eliminating those observation for which the cost lie in the tails of the distribution.

Unacceptably high costs and slow convergence are caused by inconsistencies between the background and the observation: for example, if the background assumes a clear sky but the observation is unnatural by cloud. So if we consider the previous and likelihood distribution of the background and the observations, then the 1D var adaptation finds the solution with maximum probability that satisfied both the background and observation distribution. If the distribution is high overlapping then the solution state will exist with a high probability. If the distribution has small overlapping, then the solution state is improbable, and their costs become high. So indentifying and eliminating these observations in 1D var enables a state and fast convergence in 4D-var.

4.2 Application Of The Desrozier’s Investigative

We now describe the methodology for generating the Desrozier’s diagnostic:

\[ E(d_a, d_b)^T \approx R \]  

(4.2)
where \( \Delta_b^0 = y - h(x_b) \) is the background innovation vector and \( \Delta_a^0 = y - h(x_a) \) is the analysis innovation vector. The suitability of the diagnostic (derived from 3D-Var adaptation theory) for 4D-Var adaptation is shown in Appendix B. The diagnostic is calculated for two situations: firstly using the analysis output from the OPS and secondly using the analysis output from the incremental 4D-Var adaptation. The background and analysis changes in covariance statistics are generated from the adaptation if and only if for clear sky, sea surface IASI observations. Observations will be collected from both day and night time, with the exception of daytime observations from shortwave channels which will be eliminated[21]. Using only IASI observations in the adaptation avoids the difficulties of attributing the diagnosed error correlation structure to different observation types. We will now discuss the technical and practical details of the procedure.

### 4.2.1 Technical Facts

First we calculate the diagnostic for the OPS retrievals. As previously mentioned, before satellite radiances are assimilated into 4D-Var, they are passed through the OPS for quality control. Within the OPS, a 1D-Var adaptation is performed on the equivalent brightness temperatures and a first-guess background, producing analysis retrieval. The first set of statistics will be generated using the background, \( \Delta_b^0 \), and analysis, \( \Delta_a^0 \), innovations from the 1D-Var analysis.

Calculating the diagnostic using the 4D-Var retrievals is more complicated. The initial 76 OPS run analyses those atmospheric quantities not present in the 4D-Var state vector, and passes them to 4D-Var with a quality controlled set of brightness temperatures; these are used to produce an optimal analysis increment. Along with the forecast value at the start of the time window, the increment is run through the Unified Model (UM) over a 6 hour time window to generate an analysis trajectory (is the path that a moving object follows through space as a function of time). Using the same observation set, the analysis fields can be passed back through OPS (the second OPS run), only this time as the background input. We can therefore use the background innovations generated by OPS as the \( \Delta_a^0 \) innovation statistics for the 4D-Var adaptation. This process is shown in Figure 4.3.
Figure 4.3: Met Office adaptation process: $y_0$ is the initial observation set, $\hat{y}_0$ is the quality control observation subset, $x_b$ is the background, $\hat{x}_b$ is the quality control background, $x_a$ is the analysis increment, and $x_a$ is the analysis. The yellow boxes represent adaptation steps and the pink boxes represent adaptation inputs and outputs.
Clearly we only want to generate our statistics from those observations that are deemed suitable to process in the Var system, i.e, those that pass the OPS quality control. These are easily identifiable since OPS assigns all observations a quality control flag 77 value: zero if the observation is passed to Var, one if the observation was accepted by Var but spatially thinned out, and greater than one if the observation was rejected. However, the observations passed to Var in the second OPS run will not be the same as those passed in the initial run, because the backgrounds are different. To ensure that the same observations are used to generate the $d_b^o$ innovations in the initial OPS run and the $d_a^o$ innovations in the second OPS run, we match observations using their latitude and longitude values.

### 4.2.2 Practical Calculations

The aim of this chapter is to use the background and analysis increment statistics generated from the adaptation of IASI data, to provide a consistency check on the observation error covariance’s used in the adaptation. We are interested in the correlations between channels used in (i) the 1D-Var adaptation in OPS (183 channels), (ii) the 4D-Var adaptation (139 channels). Using the diagnostic (4.2), for each channel $i$, we compute the observation error covariance with channel $j$ by averaging the product of the background and analysis innovations over the total number of observations used in the adaptation $N$,

$$ R(i,j) = \frac{1}{N} \sum_{k=1}^{N} [(d_a^o)_i (d_b^o)_i]_k - \left( \frac{1}{N} \sum_{k=1}^{N} [(d_a^o)_i]_k \right) \left( \frac{1}{N} \sum_{k=1}^{N} [(d_b^o)_i]_k \right) $$

$$ = \frac{1}{N} \sum_{k=1}^{N} (y_i^o - y_{i}^f) (y_j^o - y_{j}^f) \left( \frac{1}{N} \sum_{k=1}^{N} (y_i^o - y_{i}^f) \right) \left( \frac{1}{N} \sum_{k=1}^{N} (y_j^o - y_{j}^f) \right) $$

(4.3)
where $y_i^o$ is the brightness temperature value in channel $i$, and $y_i^a$ and $y_j^b$ are the analysis and background counterparts respectively. We subtract the mean innovation values to ensure our diagnostic is unbiased. The diagnostic (4.3) is only an approximation of the observation error covariance matrix. It is just strictly valid when $E[e_o(e_o)^\dagger] = R$ and $E[e_b(e_b)^\dagger] = B$, i.e., when the errors assumed in the adaptation are equal to those found in reality. Under such circumstances the resulting matrix would be exact and symmetric. However, we are knowingly using an incorrect specification of the observation error covariance’s, and so by construction the matrix may not be symmetric. Since an error covariance matrix is required to be symmetric positive definite, we could approximate $R$ with the symmetric component of our diagnosed matrix

$$R_{sym} = \frac{1}{2}(R + R^T)$$  \hspace{1cm} (4.4)

### 4.3 Results for Practical Calculations

We now perform a set of experiments using the techniques described in the previous section. First we consider the diagnostic (4.2) applied to the OPS analyses. A set of analyses are produced by the 1D-Var adaptation of IASI data from the 17th July 2008 at 00z, 06z, 12z and 18z, within the Observation Processing System. The total number of observations used to produce the statistics is 27,854; 9,131 of which suitable for use in the 4D-Var adaptation and 18,723 of which are thinned out. Figures 4.4 and 4.5 show the global location of all the observations used in the OPS adaptation, and the size of their background innovations for MetDB channel 1 (sensitive to stratospheric temperature) and MetDB channel 279 (sensitive to water vapour), respectively. Using the formula (4.3), we calculate the observation error covariance for this data.

Background innovations IASI channel 1
Figure 4.4: Global location and background innovation value C-B (degree kelvin) for observations in MetDB channel 1

Background innovations IASI channel 279
background innovation value C-B (degree kelvin) for observations in MetDB channel

Observation variance: operational and diagnostics
Figure 4.6 shows the operational observation error variances used in the 1D-Var adaptation (black line) and the error variances diagnosed by (4.3) (red line) for all the 183 channels used in the OPS. The error standard deviations used in the OPS (square root of the variances) are comprised of the instrument noise plus a forward model error of 0.2K. The channels numbers correspond to the index of the MetDB channel used in the OPS, i.e. MetDB channel number 1 has OPS channel index 0 (the first channel used in the OPS) and MetDB channel 280 has OPS channel index 182 (the last channel to be used in the OPS). Figure 4.7 shows a typical IASI spectrum for all 314 channels; the channels used in OPS are highlighted by the red asterisks (the full list of corresponding channel numbers and indices can be found in Appendix A). The structure of the operational and diagnostic error variances in Figure 4.6 is very similar. The diagnosed error variance is significantly lower than the current operational
IASI Spectrum for 314 chosen channels

Figure 4.7: channels used in OPS (Red asterisks) on a typical IASI spectrum (black line)

Diagnostic R matrix 1D- var
Figure 4.8: Diagnosed observation error covariance matrix for the $\text{OPS}(K^2)$

Diagnostic C matrix 1D-var

Figure 4.9: Diagnosed observation error correlation matrix for the OPS
variance for all channels. The largest difference is in the OPS indexed channels 145 – 180 which are highly sensitive to water vapour. In conclusion, the results suggest that the error variances are being overestimated especially in channels sensitive to water vapour.

The error correlation matrix can be determined easily from the error covariance matrix using the identity \( R = D^{1/2} CD^{1/2} \) from Chapter 2.5; the diagnosed error correlation matrix is shown in Figure 4.9. The correlation structure shown in Figure 4.9 is not uniformly symmetric, suggesting that the iterative procedure for updating the error variances (as proposed by Desroziers) could be beneficial.

From the results in Figure 4.9 we can conclude that when the IASI observations are analysed in the OPS, observation error cross-correlations are very small for most channels. This can be explained by recalling that only forward model and adjacent channel error correlation is expected to appear in the observation error covariance matrix. When we use the analyses from the 4D-Var adaptation, we expect the cross-channel correlations to be larger because correlated error of representative will also be contained in \( R \).

### 4.4 Summary and Conclusions

In order to form successfully observation error correlations, an true knowledge of the true correlation structure is needed. This structure of observation error correlation is varies with observation type and instrument. In this chapter we have successfully used a post-analysis diagnostic derived from variational data adaptation theory to obtain the cross-channel error correlations for IASI observations. We first introduced IASI data and commented on its current usage at the Met Office. The technical and practical details of applying the Desroziers’ diagnostic to the adaptation of IASI observations were then discussed. Background and analysis innovation statistics were acquired through the adaptation of IASI observations.
The current treatment of vertical IASI observation errors is to assume independence between channels, i.e., the observation error covariance matrix is diagonal. The new results in this chapter have challenged the validity of this assumption. The statistics from the 4D-Var adaptation showed large off-diagonal error covariances in channels highly sensitive to water vapour, and additional correlation structure in channels in the temperature sounding band. Observation error correlations were shown to be significant between neighbouring channels with similar spectral properties, leading to a block structure in the error covariance and correlation matrix. However, the statistics from the 1D-Var adaptation identified mostly uncorrelated errors between channels, with some weak correlation in those channels sensitive to water vapour. These findings advise that correlated observation errors in IASI data can largely be attributed to errors of representatives.

The application of the post-analysis diagnostic to both the 1D- and 4D-Var adaptation procedures recorded observation error variances considerably smaller than those currently being used operationally. We can attribute this over-inflation to the assumption of uncorrelated errors. In the 4D-Var adaptation, the diagnosed error covariances between certain channels are very large, and ignoring these will end to a mis-weighted representation of the observations in the analysis. Therefore inflating the variances is necessary if all observation errors are assumed independent. If we are to change this assumption, a suitable representation of the error correlation structure is needed. The diagnosed values of observation error covariances and correlations generated here provide a realistic starting point for future work on including observation error correlation structure in variational data adaptation. The block diagonal structure in the error correlation matrix highlights the potential use of Markov representations for each of the blocks, for example. Although the diagnosed matrices are not entirely symmetric, the data provides us with an approximation of the ‘true’ correlation structure, and an approximating symmetric matrix (4.4) can be generated. Against this matrix it is possible to compare analytic error correlation structures by examining features such as information content and analysis accuracy.