Improving the information content of IASI assimilation for numerical weather prediction

Submitted for the degree of Doctor of Philosophy
at the University of Leicester

by Fiona Isobel Smith

Department of Physics and Astronomy

July 2014

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

The suboptimal analysis framework used in chapter 10 of this thesis, and the method of projecting the analysis onto eigenvectors of the background error covariance matrix, used in chapter 5 and chapter 10 have been published in the following journal article:

> Eyre, J. R. and F. I. Hilton (2013). Sensitivity of analysis error covariance to the mis-specification of background error covariance. *Q. J. R. Meteorol. Soc.*, Vol. 139, pp. 524–533. DOI 10.1002/qj.1979

Improving the information content of IASI assimilation for numerical weather prediction

Fiona Isobel Smith

Abstract

The Infrared Atmosperhic Sounding Interferometer (IASI) provides significant impact to numerical weather prediction systems despite current assimilation schemes using less than 2% of the channels. The current system does not achieve the information content predicted by earlier theoretical studies and results presented here show that the information content could be doubled if the full spectrum were exploited. There is potential to improve the vertical resolution of the humidity analysis and the stratospheric temperature in particular.

This thesis explores principal component (PC) compression and radiance reconstruction to compress the spectrum by over 90% whilst retaining almost the full information content. Theoretical calculations are shown that indicate PC scores and reconstructed radiances achieve close to the maximum information content, making them promising approaches for better exploitation of IASI. However, care must be taken because neglected error terms and matrix conditioning are problematic due to the way the information in the compressed observations is coupled in the vertical. New methods for choosing reconstructed radiance channels for assimilation are developed and tested, generating channel selections suitable for implementation in the Met Office operational system.

The final section is concerned with the interaction between the observation information and the background error covariance matrix. This matrix can only ever be estimated, which causes the analysis to be suboptimal. If the differences between true and assumed errors are large enough, the analysis may be degraded relative to the background. Guarding against exaggeration of background errors is therefore important, and for water vapour in particular, spurious vertical structures in the stratosphere must be avoided. Increasing the spectral coverage increases the information content and reduces exposure to analysis degradation. This result is encouraging because it means that there is no greater risk to the analysis if more spectral information is provided, paving the way for assimilation of reconstructed radiances.

Acknowledgements

This PhD has been funded by the Met Office, to whom I am grateful for the opportunity to conduct this research. I would particularly like to thank Steve English and John Eyre who supervised the work and somehow managed to find me the time to undertake it. They also provided much encouragement and useful discussion.

I would also like to thank John Remedios for his supervision and hugely helpful suggestions that have greatly improved this thesis. Although I did not spend much time in Leicester, I found EOS a very enthusiastic place, and an enjoyable department to be a part of. A big thank you to all the friends I have made there for making me so welcome. My thanks also to Mike Goad and Jane Hull.

I owe a lot to Sean Healy, Andrew Collard and Tim Hultberg, all of whom I have worked closely with to try to understand the complications and implications of reconstructed radiance error covariances.

My Met Office colleagues have been very understanding of my non-availability and I am grateful to them all for being such good people to work with. In particular, I am grateful for technical assistance and useful feedback provided by Nigel Atkinson and Stephan Havemann. My thanks also to Andrew Lorenc and Marek Wlasak for their thorough explanations of the ins and outs of 4D-Var and covariance modelling.

Finally, I would also like to thank Kerry Day and Katie Lean for their proofreading assistance.

Contents

List of Figures

List of Tables

Glossary of new terms introduced in this thesis

Chapter 6

Raw radiance
A fully calibrated "level 1c" radiance spectrum. This is the spectrum as measured by the instrument, but subject to ground-processing that renders it usable for assimilation or retrieval. Note that this is not a new term but it is not widely used other than by those who work on PC-compressed radiance spectra. It is used to distinguish the original observation from the reconstructed radiance spectrum.

Reconstructed radiance
A radiance spectrum that has been reconstituted from a truncated set of PC scores, such that it retains most of the signal and with most of the noise discarded. Note that this is not a new term, but it is not widely used other than by those who work on PC-compressed radiance spectra.

Suboptimal-in-H
This refers to a calculation using the linear analysis equations, or a variational analysis such as 1D- or 4D-Var that uses a forward model Jacobian matrix, $\mathbf{H}$, that is known to be incorrect. In this thesis, this scenario specifically refers to the use of a raw radiance Jacobian for the assimilation of reconstructed radiances.

Suboptimal-in-H-and-R
This refers to a calculation using the linear analysis equations, or a variational analysis such as 1D- or 4D-Var that uses a forward model Jacobian matrix, $\mathbf{H}$, that is known to be incorrect, and where the observation error covariance matrix, $\mathbf{R}$, is not adjusted to take account of this extra forward model error. In this thesis, this scenario specifically refers to the assimilation of reconstructed radiances using a raw radiance Jacobian coupled with an observation error covariance matrix for reconstructed radiances that does not take reconstruction error into account.

Chapter 7

Double difference covariance matrix
Alternative formulation of the observational method of Hollingsworth and Lönnberg (1986) as described in Cameron and Eyre (2008), used to estimate an observation error covariance matrix from observation–background statistics.

Chapter 10

Suboptimal-in-B

This refers to a calculation using the linear analysis equations, or a variational analysis such as 1D- or 4D-Var that uses a background error covariance matrix, $\mathbf{B}_A$, that is known to be incorrectly formulated relative to the true covariance matrix, $\mathbf{B}_T$.

Danger Zone

This refers to the region of background – analysis error space for the suboptimal-in-**B** scenario where the analysis error is larger than the background error. This generally occurs where the assumed background error variance is 2–4 times larger than the true variance.

List of symbols

General mathematical terms

$^{1/2}$	Superscript representing matrix square root.
$^{-1}$	Superscript representing matrix inverse.
$^{-I}$	Generalised inverse of a non-square matrix.
$^{\mathrm{T}}$	Superscript representing matrix transpose.
Tr	Trace of a matrix.
$\lvert\mathbf{X}\rvert$	The vertical lines denote the determinant of a matrix.
$\mathbf{I}$	The identity matrix.

Chapter 3

$\mathbf{A}$	Analysis error covariance matrix.
$\mathbf{B}$	Background error covariance matrix.
$\underline{\mathbf{B}}$	4-dimensional background error covariance matrix (the underscore denotes the time dimension).
$\underline{\mathbf{d}}^o$	Departure of observations from the best guess atmospheric state, the underline representing the distribution of the vector in the time dimension.
$H(\mathbf{x})$	Forward model converting the atmospheric state, $\mathbf{x}$, to a model equivalent observation.
$\mathbf{H}$	Forward model Jacobian matrix.
$\mathbf{K}$	Kalman gain matrix; the weights given to the observations in the optimal analysis.
$\underline{M}$	NWP model including trajectory in time.
$\mathbf{M}$	Jacobian of full NWP model.
$\underline{\mathbf{M}}$	4D Jacobian of full NWP model including trajectory in time.
$\underline{\tilde{\mathbf{M}}}$	Perturbation forecast linear model.
MoistCov	Background error covariance matrix calculated from an ensemble of forecast states and using a new moisture control variable designed to eliminate increments in the stratosphere.
NMC	Background error covariance matrix calculated with the NMC method (Parrish and Derber, 1992).
q_t	Total water, i.e. including liquid, vapour and ice phases.
q	Water vapour mass mixing ratio.
$\mathbf{R}$	Observation error covariance matrix.
$\mathbf{S}$	Simplification operator mapping from full NWP model to perturbation forecast model state.
$\mathbf{T}$	Variable transform mapping from perturbation forecast model variables to 4D-Var control vector variables that are uncorrelated.

$\mathbf{T}_h$	The horizontal transform mapping from model grid points to spectral functions.
$\mathbf{T}_v$	The vertical transform that maps from model levels to uncorrelated vertical modes.
$\mathbf{T}_p$	The parameter transform that maps from forecast variables such as potential temperature and wind to 4D-Var uncorrelated variables such as stream function and ageostrophic pressure.
$\mathbf{U}$	Variable transform mapping from 4D-Var control vector variables that are uncorrelated to perturbation forecast model variables.
$\mathbf{U}_h$	The horizontal transform mapping from spectral functions to model grid points.
$\mathbf{U}_v$	The vertical transform that maps from uncorrelated vertical modes to model levels.
$\mathbf{U}_p$	The parameter transform that maps from 4D-Var uncorrelated variables such as stream function and ageostrophic pressure to forecast variables such as potential temperature and wind.
$\mathbf{v}$	4D-Var control vector, equivalent to $\delta\mathbf{w}$ after the $\mathbf{T}$ transform has been applied.
$\mathbf{v}^b$	The subscript b indicates the background (short-range forecast a priori) $\mathbf{v}$ vector.
$\mathbf{w}$	Simplified model state vector on a given iteration, equivalent to $\mathbf{x}$ after the application of the simplification operator, $\mathbf{S}$.
$\mathbf{w}^b$	Background (a priori) simplified model state vector.
$\mathbf{w}^g$	Best guess simplified model state vector, after an iteration of the 4D-Var outer loop. Equal to $\mathbf{w}^b$ before the outer loop is run.
$\delta\mathbf{w}$	Perturbation forecast model state vector, equal to the perturbation of $\mathbf{w}$ from the best guess vector, $\mathbf{w}^g$.
$\delta\mathbf{w}^b$	Perturbation forecast background (a priori) model state vector, equal to the perturbation of $\mathbf{w}^b$ from the best guess vector, $\mathbf{w}^g$.
$\mathbf{x}^t$	True atmospheric state.
$\mathbf{x}$	3D NWP model atmospheric state at a single point in time.
$\underline{\mathbf{x}}$	4D NWP model atmospheric state, with the underline character incorporating a time dimension to the vector.
$\mathbf{x}^b$	Background state: typically a 6-hour forecast for an NWP model used as prior information in a variational scheme.
$\mathbf{x}^a$	Analysis state: the optimal combination of the background state with the observations sought by a variational scheme.
$\mathbf{x}^g$	Best guess atmospheric state. The atmospheric state at the start of the assimilation window after a run of the outer loop in 4D-Var. Equal to $\mathbf{x}^b$ at the start of minimisation.
$\underline{\mathbf{x}}^g$	As $\mathbf{x}^g$, but with the underline character indicating that the atmospheric state has been propagated in time through the assimilation window.
$\delta\mathbf{x}$	Atmospheric perturbation: the difference of the atmospheric state on a given iteration of variational analysis from the best guess state. Used in incremental formulations of 3D- and 4D-Var.

$\delta\underline{\mathbf{x}}$	As $\delta\mathbf{x}$, but with the underline character incorporating a time dimension to the vector.
$\delta\mathbf{x}^b$	Background perturbation: the difference of the best guess state on a given iteration of variational analysis from the background state. Used in incremental formulations of 3D- and 4D-Var.
$\mathbf{y}$	Forward modelled top of atmosphere radiance generated from the NWP model state.
$\mathbf{y}^o$	Observed top of atmosphere radiance as measured by IASI.
$\underline{\mathbf{y}}$	As $\mathbf{y}$, but the underline character emphasises the distribution of observations in time by incorporating a time dimension.

Chapter 6

$\mathbf{C}$	Covariance matrix representing variability of an observed radiance climatology (departures from a mean radiance).
$H_{obs}(\mathbf{x})$	Forward model for raw radiances.
$\tilde{H}(\mathbf{x})$	Forward model for reconstructed radiances.
$\tilde{\mathbf{H}}$	Jacobian of forward model for reconstructed radiances.
$H_{pc}(\mathbf{x})$	Forward model for principal component scores.
$\mathbf{L}$	Matrix of eigenvectors of $\mathbf{C}$, the radiance climatology.
$\mathbf{L}_p$	Truncated matrix of eigenvectors, retaining only the leading vectors.
$\mathbf{L}_s$	Truncated matrix of eigenvectors, retaining only the leading vectors for selected channels.
n	Number of channels in the raw radiance spectrum.
n_{ass}	Number of reconstructed radiance channels assimilated.
$\mathbf{N}$	Assumed instrument noise of observed radiances.
$\tilde{\mathbf{N}}$	Assumed instrument noise of reconstructed radiances.
p	Number of retained principal component scores.
$\mathbf{p}$	Vector of truncated principal component scores.
$\mathbf{P}$	Principal component score selection matrix that discards lower order PCs.
$\mathbf{\Phi}$	Extra forward model error (or reconstruction error) resulting from the use of a raw radiance forward model with reconstructed radiances as a consequence of discarding true, forward modelable atmospheric signal in the truncated PC set.
$\mathbf{R}_{pc}$	Observation error covariance matrix for PC scores.
$\tilde{\mathbf{R}}$	Observation error for reconstructed radiances.
$\tilde{\mathbf{R}}'$	In the Suboptimal-in-$\mathbf{H}$ system, the observation error for reconstructed radiances taking into account additional error from the use of a raw radiance forward model.
$\mathbf{S}$	Channel selection matrix that defines which channels are reconstructed from principal component scores.
$\mathbf{y}_m$	Climatological mean observed radiance.
$\mathbf{y}_{obs}$	Observed top of atmosphere radiance.

$\tilde{\mathbf{y}}$ Reconstructed radiance.

$\mathbf{y}_t$ Top of atmosphere radiance that is generated by the true atmospheric state.

$\tilde{\mathbf{y}}_t$ Top of atmosphere radiance that is generated by the true atmospheric state filtered to reconstructed radiance space.

$\delta\tilde{\mathbf{y}}'$ Observation increment in the Suboptimal-in-**H** assimilation where a reconstructed radiance is assimilated with a raw radiance forward model.

$\delta\tilde{\mathbf{y}}'_{fm}$ In the suboptimal-in-**H** assimilation where a reconstructed radiance is assimilated with a raw radiance forward model, the departure of the forward modelled observation from the true reconstructed radiance.

$\mathbf{Y}$ Covariance matrix representing climatological variability of the true radiance spectrum.

Chapter 7

$\mathbf{R}_{EC}$ Observation error covariance matrix derived using the double difference method, equivalent to the observational method of Hollingsworth and Lönnberg (1986), from IASI observations and ECMWF analyses interpolated to observation time.

$\mathbf{R}_{HL}$ Observation error covariance matrix derived using the double difference method, equivalent to the observational method of Hollingsworth and Lönnberg (1986), from IASI observations and Met Office forecasts interpolated to observation time.

Chapter 8

$\mathbf{H}_{pc}$ Jacobian of principal component forward model.

Chapter 10

$\mathbf{A}_{opt}(\mathbf{B}_A)$ In the suboptimal-in-**B** analysis, the optimal linear analysis covariance that would have resulted had $\mathbf{B}_A$ been the true background covariance.

$\mathbf{A}_{opt}(\mathbf{B}_T)$ In the suboptimal-in-**B** analysis, the optimal linear analysis covariance that would have resulted had $\mathbf{B}_T$ been accurately specified.

$\mathbf{A}(\mathbf{B})$ In the suboptimal-in-**B** analysis, the suboptimal linear analysis covariance that results when $\mathbf{B}_A$ is used in the analysis but is not equal to the true background covariance, $\mathbf{B}_T$.

$\mathbf{B}_T$ In the suboptimal-in-**B** analysis, the true background error covariance matrix.

$\mathbf{B}_A$ In the suboptimal-in-**B** analysis, the assumed background error covariance matrix.

$\mathbf{B}'$ A hypothetical background error covariance matrix with the same eigenvectors as $\mathbf{B}_A$, but with eigenvalues set to those of the diagonal of

the projection of $\mathbf{B}_T$ onto the eigenvectors of $\mathbf{B}_A$.

$\mathbf{B}_d$ In the scalar suboptimal-in-$\mathbf{B}$ analysis for a given $\mathbf{B}_A$, the value of $\mathbf{B}_T$ below which the true analysis error will be higher than the true background error.

$\mathbf{K}_{\mathbf{B}_A}$ In the suboptimal-in-$\mathbf{B}$ analysis, the Kalman gain matrix calculated with the assumed background error, $\mathbf{B}_A$.

$\mathbf{\Lambda}$ Matrix of eigenvalues of the assumed background error covariance matrix

$\mathbf{V}$ Matrix of eigenvectors of the assumed background error covariance matrix

Appendix D

$\delta\tilde{\mathbf{y}}_{obs}$ In the suboptimal-in-$\mathbf{H}$ assimilation where a reconstructed radiance is assimilated with a raw radiance forward model, the departure of the reconstructed radiance observation from the true raw radiance.

Chapter 1

Introduction

1.1 Observations in numerical weather prediction

Numerical weather prediction (NWP) is often described as an initial value problem: the forecast is produced by propagating forwards the estimated state of the atmosphere using dynamical and physical equations. The accuracy of the forecast is therefore dependent on the accuracy of the initial atmospheric state. NWP centres use observational data to construct the initial state via data assimilation (DA). DA is the science of combining observations with prior information to produce a statistically optimal estimate of the true state.

The observations assimilated into NWP models are usually a combination of: radiosonde ascents, measuring profiles of temperature and relative humidity; aircraft, ship, land surface and buoy observations of parameters such as wind, temperature, humidity and surface pressure; and satellite observations of temperature, humidity, wind and in some cases precipitation. The World Meteorological Organisation (WMO) aims to coordinate the development and provision of this ground and space-based collection of meteorological observations via its WMO Integrated Global Observing System (WIGOS) programme. Satellite observations are a critical part of the WIGOS, providing global coverage several times a day of key atmospheric and surface variables. The so-called conventional data from surface, ship and aircraft observations are usually limited to populous areas and routes, plus a few limited remote sites.

1.2 Satellite observations in NWP

An overview of the main types of satellite observations used in NWP is given by Collard et al. (2011). Satellite sounding data have been assimilated in operational NWP systems since the 1980s (e.g. Gadd et al., 1995). Over time, satellite data have come to form the majority of the data assimilated, with data volumes increasing rapidly year-on-year, but they also provide the largest proportion of the short-range forecast impact (Cardinali, 2009). The increase in forecast skill in the

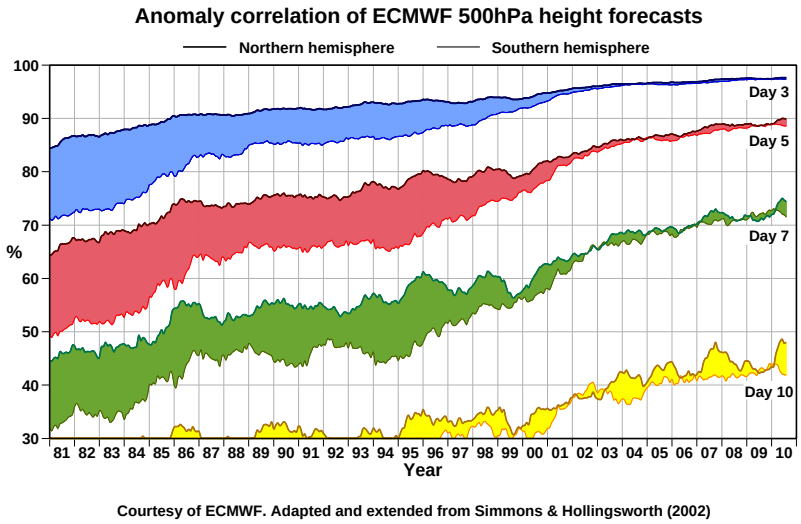


Figure 1.1: The evolution of forecast skill in time for ECMWF forecasts, as represented by 500 hPa Anomaly Correlation scores. A score of 100 % is ideal. Shown in blue, red, green and yellow are forecast lead times of 3, 5, 7 and 10 days respectively. The thick lines represent the northern hemisphere (NH) score, and the thin lines the southern hemisphere (SH). Thus it can be seen that in the 1980s the SH forecasts were considerably worse than the NH, but for the 3 day forecasts this gap had closed to almost nothing by 2001. The effects of data assimilation are reduced with increasing forecast lead time. This figure was previously published in Collard et al. (2011).

southern hemisphere – for which there are fewer conventional data – relative to the northern hemisphere (shown for ECMWF[1] in figure 1.1), has been attributed by authors such as Simmons and Hollingsworth (2002) and Rabier (2005) mostly to the impact of satellite observations, from results such as those shown in figure 1.2. With their global coverage and information on the vertical structure of temperature and humidity, satellite sounders are generally considered to give the most forecast impact of any observation type (Joo et al., 2013).

Unlike radiosondes, satellite sounding instruments do not provide direct measurements of atmospheric components, but rather they are sensitive to radiation emitted and reflected by components of the earth's atmosphere and surface (scattering processes are not generally important at infrared wavelengths). A radiative transfer forward and inverse model is required to relate the atmospheric state to the measurements and vice versa.

Much of the signal in the infrared (IR) part of the electromagnetic spectrum

[1] the European Centre for Medium-Range Weather Forecasts

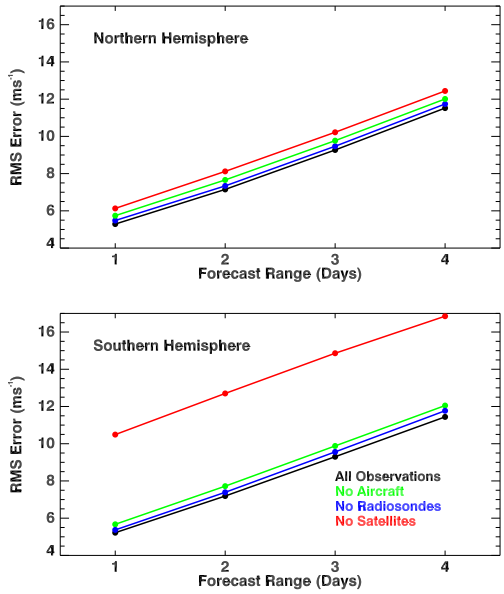


Figure 1.2: The impact of removing respectively aircraft, radiosonde and satellite observations on ECMWF 200 mbar vector wind forecast error from the full data assimilation. This figure is based on results from Kelly and Thépaut (2007) and was previously published in Collard et al. (2011).

comes from emission by carbon dioxide and water vapour. Carbon dioxide is well-mixed in the atmosphere and so the strength of emission by CO_2 molecules can be used to infer information about the atmospheric temperature. Thus channels that measure close to the spectral frequencies of CO_2 absorption lines are generally referred to as temperature sounding channels. Emission by H_2O molecules in the water vapour band means that close to H_2O absorption lines the measurements are sensitive to the atmospheric water vapour burden. In the window region, where there is little absorption and emission from atmospheric components, the radiance measurement is sensitive to the surface emissivity and temperature, and weakly sensitive to water vapour via continuum contributions. Figure 1.3 illustrates how, away from the window region, a sounding observation is the sum of emission terms from each layer of the atmosphere. For cloudy fields of view, much of the spectrum becomes sensitive to the emission by the surface of the cloud. Correct treatment of cloudy fields of view is an area of active research, as mentioned in section 2.3.

1.2.1 Hyperspectral sounders

Each channel of an IR sounder has a response to the atmospheric radiation similar to that shown in figure 1.3. The level that contributes most to the observed signal

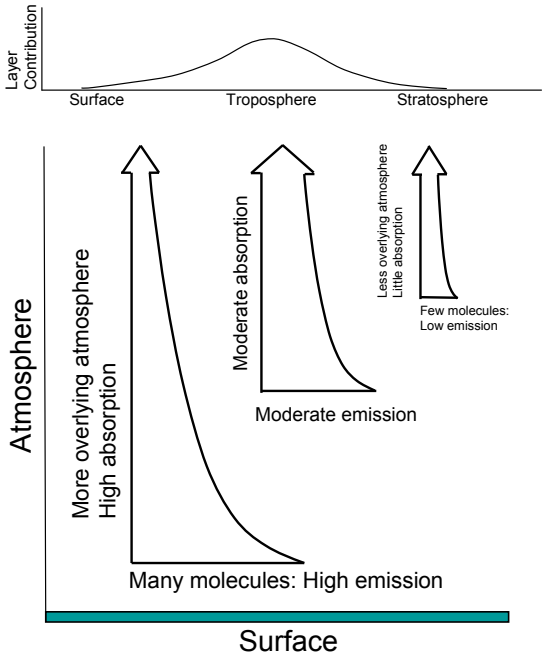


Figure 1.3: An illustration of how an observation made by a sounding instrument is composed of emission terms from each layer of the atmosphere. The greatest part of the signal comes from the region where the emission from the larger number of molecules lower in the atmosphere is balanced by the lower absorption by the atmosphere above the emitting layer. The level that contributes most to the observed signal depends on the absorption of the atmosphere at the wavelength considered. This figure was previously published in Collard et al. (2011).

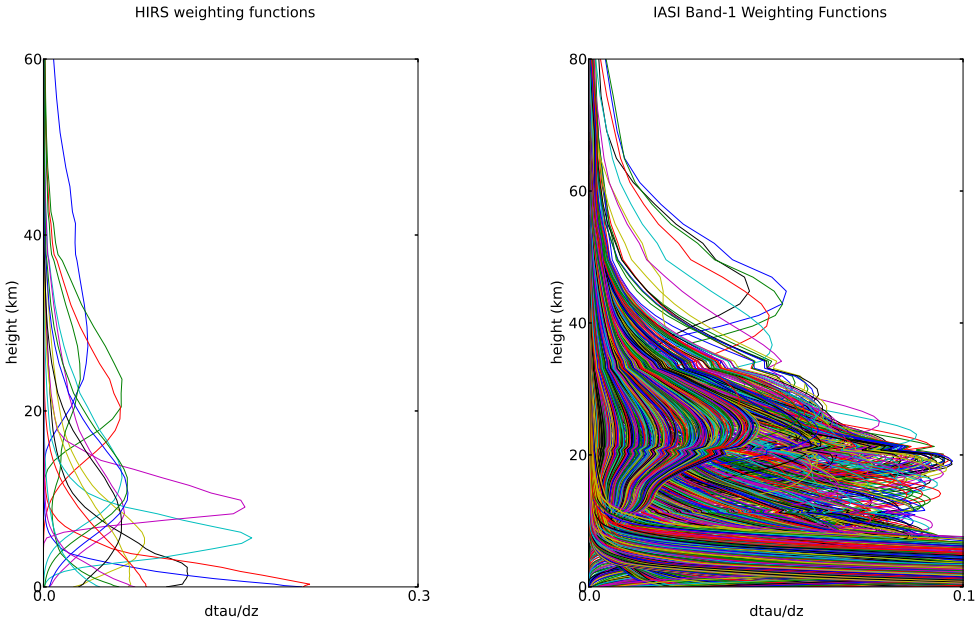
– the weighting function peak – varies depending on the frequency at which the channel measures. Close to a strong absorption line, the few molecules that exist in the upper atmosphere will be enough to saturate the measurement with their emission. Further from an absorption line the greater mass lower in the atmosphere will dominate the emission (even though the signal will be attenuated by absorption in the upper atmosphere). A family of measurements at different frequencies will thus provide information from different parts of the atmospheric column. The more measurements at different frequencies, the greater the potential for extracting fine vertical scale information on the atmsopheric column.

In the last ten years hyperspectral or advanced IR sounders have been launched, such as the Infrared Atmospheric Sounding Interferometer (IASI) and the Atmospheric InfraRed Sounder (AIRS), that have thousands of high-specral-resolution channels. The previous generation of IR sounders, such as the High-resolution Infrared Radiation Sounder (HIRS) had only 20 broad channels spanning up to $30\,\mathrm{cm}^{-1}$. Figure 1.4 compares the weighting functions of the HIRS temperature sounding channels with those of the IASI longwave CO_2 band. The increased spectral resolution of IASI provides greater vertical resolution because there are many measurements with overlapping weighting functions that span the atmospheric column.

Radiance measurements are usually expressed as top of atmosphere brightness temperatures for assimilation into NWP models. Figure 1.5 shows a typical brightness temperature spectrum measured by IASI. The range from 645 to $770\,\mathrm{cm}^{-1}$ is often referred to as the longwave CO_2 band, and the region between approximately 1200 and $2000\,\mathrm{cm}^{-1}$ as the water vapour band. The window region, where there are few strong CO_2 or H_2O absorption lines, is between 800 and $1100\,\mathrm{cm}^{-1}$. The region beyond $2200\,\mathrm{cm}^{-1}$ is often referred to as the shortwave infrared in the NWP community.

Joo et al. (2013) showed that in the Met Office global assimilation, IASI provides half of the impact of the whole Metop-A satellite, which in turn provides 25% of the total impact of all observations, more than any other platform or conventional observation type. Although it is already one of the most important observations within the assimilation system, IASI might still be under-used because only a few hundred isolated channels are currently assimilated. This thesis aims to improve access within the Met Office assimilation scheme to the information contained in each IASI observation, with particular emphasis on vertical structure information. The thesis focusses on clear sky assimilation.

The equations of the data assimilation schemes used operationally at the Met



(a) HIRS weighting functions **(b)** IASI weighting functions

Figure 1.4: Comparison of HIRS and IASI band 1 weighting functions, with colours being used simply to differentiate between channels. The weighting function shows the contribution from each layer of the atmosphere in the vertical to the overal measurement. It is defined as the derivative of the atmospheric transmittance (tau) with respect to height (z)

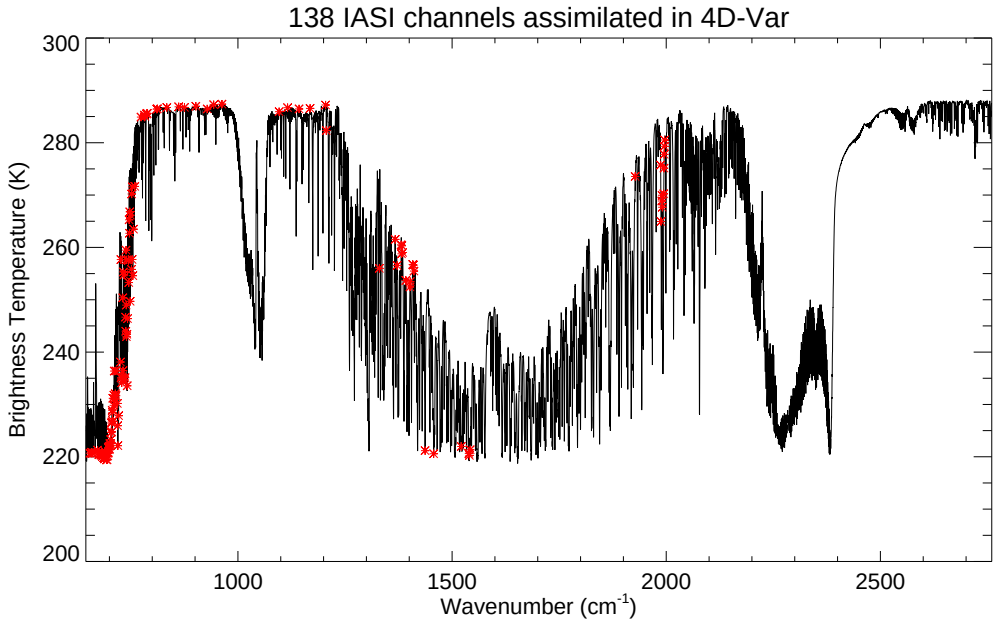


Figure 1.5: A typical IASI brightness temperature spectrum. Marked with red stars are the operationally assimilated channels from the *VAR* selection (see appendix E and section 4.3 for information on this channel selection).

Office and in this thesis are described in chapter 3. DA schemes are usually based on optimal estimation theory (e.g. Rodgers, 2000). The analysis generated will only be optimal if all assumptions made in the underlying theory hold true. Fundamental to an optimal analysis is that the background and observation error covariance statistics are well-characterised. Until the 1990s, it was common to assilimate column retrievals of satellite-derived temperature and humidity, but this method relies upon accurate knowledge of the error characteristics of these retrievals. In some cases retrievals may be incestuous if the same weather forecasting model is used to provide prior information to the retrieval and to the assimilation system, or may have inconsistencies because of differing assumptions, for example in surface temperature. In recent times, almost all NWP centres assimilate the radiances directly, using the radiative transfer model as part of the data assimilation code. Since the operational use of advanced data assimilation techniques such as 4-Dimensional Variational Assimilation (4D-Var), satellite radiance data from sounding instruments have become an increasingly important part of the observing system. The error characteristics of the brightness temperature data still need to be known, but they are generally less complex and there is minimal danger of correlation between observation and background errors.

Of equal importance in establishing the correct weighting of observations and background to produce an optimal analysis is the background error covariance matrix. The form of this is generally estimated using a procedure such as the so-called National Meteorological Center (NMC) method (Parrish and Derber, 1992) or via the use of an ensemble forecasting system (e.g. Fisher, 2003). These methods are usually used to produce a statistical estimate of the error in the background state; because the true atmospheric state is not known, background errors can only ever be approximated and may be significantly different from the true errors.

1.3 The purpose and structure of this thesis

Hyperspectral sounders provide huge volumes of data, with a large amount of redundancy between channels. It is not possible for computational reasons to assimilate all the channels, neither is it desirable because of the inefficiency of extracting a few tens of pieces of NWP information from several thousand channels. The purpose of this thesis is to establish the information content of IASI within the current assimilation system with particular reference to vertical structure, and then investigate how the information could be extracted from the observations more efficiently, so

that the maximum information can be extracted with the minimum computational effort. The focus is on clear sky situations.

One way this could be done is via principal component (PC) compression, where the observations are projected onto orthogonal vectors, and sorted by eigenvalue. The components with the lowest eigenvalues generally show no coherent spatial pattern and contain most of the random measurement error. These can be removed leaving the atmospheric signal represented in only a few hundred scores. The use of PC scores requires new methods to deal with cloudy atmospheres. Current assimilation schemes for hyperspectral sounders use channels above a diagnosed cloud top, for which the effects of the cloud below can be sufficiently well modelled by a simple grey cloud assumption. This technique does not work well with PC scores, as each component is sensitive to most of the atmospheric column, whereas the original channels have much more localised weighting functions.

An alternative technique is to reconstruct the noise-reduced spectrum from the PC scores: theory shows that it should be possible to find a set of these reconstructed radiances that contain the same information content as the set of PC scores from which they were calculated, and in the same number of channels as PC scores. The specific aim of this thesis is to investigate the potential of reconstructed radiances in a clear sky assimilation setting, and to find a set of reconstructed radiance channels with the maximum possible information content to be used in an operational setting.

A further aim is to investigate the potential effects of an incorrectly specified background error covariance structure. One possible outcome of this suboptimal system is that the analysis may have larger errors than the true background error. The final part of this thesis investigates whether increasing the weight given to the observations by presenting them in a more information-dense format increases the likelihood of degrading the analysis, or on the contrary provides some protection.

Chapter 2 presents the IASI instrument, its use in NWP, and main areas of current research that will enable better use of the instrument. Chapter 3 presents the equations of data assimilation, and discusses the main features of the implementation of 4D-Var at the Met Office, including the background error term. The optimal 1D linear analysis and experimental 1D-Var framework used in the thesis are then described.

Chapter 4 assesses the information content of the current assimilation system, using an adjoint-based technique to establish the pattern of information delivered by the observations. Information content measures are then used to determine how much extra information it may be possible to extract from the observations over the

current channel selection. Chapter 5 then predicts the impact of the assimilation of the current IASI channel selection using optimal analysis equations, and compares the prediction with the performance in a 1D-Var simulation experiment.

The next four chapters discuss principal component compression and radiance reconstruction. Chapter 6 introduces the theory and discusses the characteristics of the reconstructed radiance observation error covariance matrix. The extra errors introduced by assimilating reconstructed radiances with a raw radiance forward model are presented. Chapter 7 describes a method for selecting channels appropriate for use with reconstructed radiances, and shows the estimation of the observation error covariance matrix, which is a necessary precursor, using the observational method of Hollingsworth and Lönnberg (1986). Chapter 8 then compares the information content of several different channel selections produced using this method and the mathematical stability of the resultant observation error covariances. Chapter 9 tests the use of reconstructed radiances and principal component scores in a 1D-Var simulation framework.

Chapter 10 demonstrates the effect of misspecified background error on the analysis, and examines whether the use of more spectral information is likely to make the analysis more or less robust to this misspecification. Finally, chapter 11 presents the conclusions of this thesis and examines ways in which the work could be extended.

Chapter 2

The Infrared Atmospheric Sounding Interferometer

IASI is the first operational IR interferometer on a weather satellite, providing information on atmospheric temperature, water vapour content, ozone and trace gas concentrations. It is mounted on EUMETSAT's[1] polar orbiting satellites, the first of which (Metop-A) was launched in October 2006 and the second (Metop-B) in September 2012. The final IASI instrument is planned to be launched on Metop-C in 2018. The observations from IASI are used operationally at the Met Office to improve the knowledge of temperature and water vapour. Section 2.1 describes the Metop satellite, and section 2.2 describes the IASI instrument and its observations in more detail. Section 2.3 describes the current status of the science of IASI assimilation at the Met Office and other operational NWP centres, and section 2.4 highlights some important future instruments that will potentially benefit from the work carried out in this thesis.

2.1 The Metop satellites

The Metop satellites form Europe's contribution to the Initial Joint Polar System (IJPS) – a collaboration between EUMETSAT and the US National Oceanic and Atmospheric Administration (NOAA). The IJPS provides operational global coverage of critical meteorological observations via low-earth orbiting satellites in two sun-synchronous orbital planes. Metop-B provides coverage in the morning orbit, with a local equator crossing time of 09:30 for the descending node. The afternoon orbit is served by NOAA satellites, with the current operational satellite being the Suomi-NPP satellite. Coverage is extended by providing data from older satellites that are still functioning, such as Metop-A and NOAA-19.

IASI is one of eight meteorological instruments on board Metop; the other instruments are shown in table 2.1. Of these, the microwave and infrared sounding radiometers (AMSU, MHS and HIRS), the ASCAT scatterometer, and radio oc-

[1] the European Organisation for the Exploitation of Meteorological Satellites

	Instrument Name	Instrument Type	Main Purpose
IASI	Infrared Atmospheric Sounding Interferometer	IR interferometer	Atmospheric temperature and humidity profiles; trace gases
AMSU	Advanced Microwave Sounding Unit	Microwave radiometer	Atmospheric temperature
MHS	Microwave Humidity Sounder	Microwave radiometer	Atmospheric humidity
HIRS	High-resolution InfraRed Sounder	IR radiometer	Upper tropospheric and stratospheric temperature; tropospheric humidity
GRAS	Global Navigation Satellite System (GNSS) Receiver for Atmospheric Sounding	Radio Occultation receiver	Upper-atmosphere temperature and humidity
ASCAT	Advanced SCATterometer	Pulsed radar	Ocean surface windspeed and direction
GOME-2	Global Ozone Monitoring Experiment-2	Nadir Ultraviolet/ Visible Spectrometer	Ozone
AVHRR	Advanced Very High Resolution Radiometer	IR/Visible imager	Sea surface temperature; Cloud and surface properties

Table 2.1: The meteorological instruments on the Metop satellite platforms. More details of the instruments can be found in Klaes et al. (2007).

cultation (GRAS) observations are assimilated operationally into the Met Office analysis. The visible and infrared imagery provided by the AVHRR instrument is used by forecasters, but also forms an important part of the ground-segment processing of IASI data as described in section 2.2.

2.2 The IASI Instrument

The primary purpose of IASI is to provide high-resolution atmospheric sounding information. Mission accuracy requirements were stated as 1 K in the free troposphere for temperature and 10 % for humidity, for a vertical resolution of 1 km (Chalon et al., 2001). The instrument capabilities allowed pre-launch specification of trace gas total column products for CO_2, O_2, N_2O, CO and CH_4.

The IASI instrument (Siméoni et al., 1997) was developed by the Centre National d'Études Spatiales (CNES) in cooperation with EUMETSAT, and was built

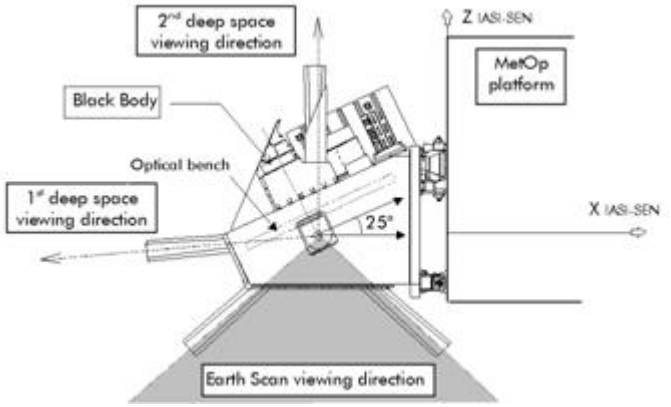


Figure 2.1: Diagram of the IASI instrument. This figure was previously published in Hilton et al. (2012a).

by ThalesAleniaSpace. It is a nadir-viewing cross-track scanning Michelson interferometer which covers the range 645–2760 cm^{-1} (3.8–15.5 µm) with 8461 spectral samples (usually referred to as channels) at 0.25 cm^{-1} intervals. Figure 2.1 is a diagram of the IASI instrument. It is the first operational instrument of its kind.

The instrument's scan geometry relative to the path of the satellite is shown in figure 2.2. IASI has a total of 30 ground fields of regard (FOR) per scan. One FOR measures a 25 km by 25 km area containing a 2×2 array of footprints (often referred to as pixels), each of which is circular with a 12 km diameter at nadir. The spectrum is measured in three wavelength bands each of which has a separate detector, with approximate coverage as follows: Band 1 – longwave CO_2 and window regions; Band 2 – water vapour band; Band 3 – shortwave.

The raw measurements obtained by the instrument are interferograms which are processed to radiometrically calibrated spectra on board the satellite using two calibration views, which can be identified on figure 2.1 as the first Cold Space (CS1) view and the internal Black Body (BB) view. The observations are further processed by the EUMETSAT ground segment, to deliver 8461 apodised radiances (known as the Level 1c or L1c product) to the end user. The apodisation function renders a Gaussian instrument spectral response function (SRF) of half-width 0.5 cm^{-1}. The advantage of the apodisation function is to remove long-range lobes from the sinc function of the inverted interferogram, improving signal to noise ratio and simplifying radiative transfer, but at the expense of widening the main spectral response of each measurement. This introduces significant correlation to adjacent channels because the SRF half-width is twice that of the spectral sampling interval. Most NWP centres try to avoid these correlations by picking sparse channel selections

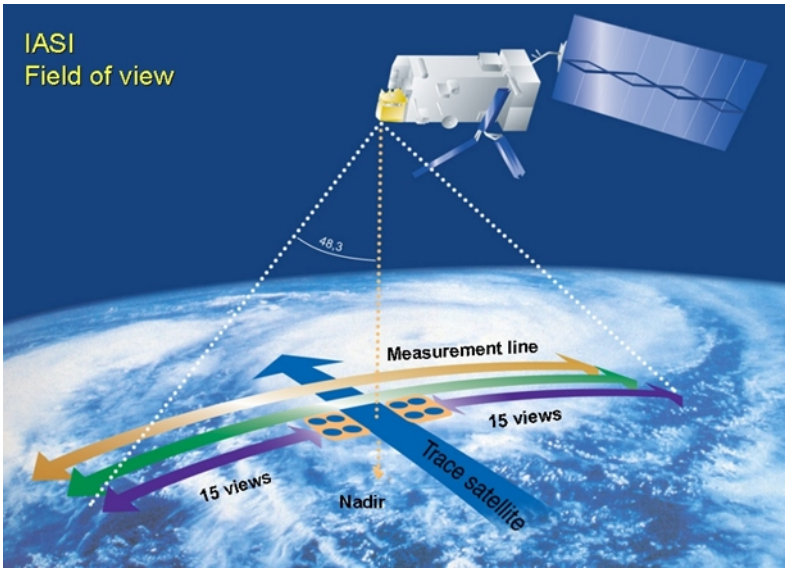


Figure 2.2: Illustration of the scan geometry of the IASI instrument, showing the four footprints within each of 15 fields of regard either side of the path of the satellite. This figure was previously published in Hilton et al. (2012a).

that avoid adjacent channels where possible.

The radiometric noise performance, in terms of noise-equivalent brightness temperature at 280 K is close to 0.16 K for most of the tropospheric temperature sounding channels of the longwave CO_2 band and 0.1–0.12 K for much of the water vapour band, which equates to 0.16 K and 0.2–0.4 K respectively at standard scene temperature. As described in section 1.1, the high spectral resolution provides channels with highly overlapping response to changes in the atmospheric profile, and the large number of these overlapping channels results in higher resolution profile information than is possible with instruments with fewer channels.

The IASI interferometer has an associated imager, the Integrated Imaging Subsystem (IIS), which is used to co-register the interferometer footprints with Metop's Advanced Very High Resolution Radiometer (AVHRR) imager. Provided with the IASI radiances is information derived from the AVHRR observations within each IASI pixel, which helps to characterize the heterogeneity of the scene, including the cloud fraction. This information can be used to help with data selection (thinning), and is used at the Met Office to select one out of four pixels corresponding to the most homogeneous scene for storage and onward processing.

The IASI instrument characteristics, performance and major scientific achievements delivered with its data are reviewed in Hilton et al. (2012a). In brief, the stability, good radiometric calibration and characterisation of IASI have resulted in an instrument ideally suited to NWP, where reliable data is of great importance.

The instrument has proved to have greater capabilities than expected in the field of atmospheric chemistry, and its data will form a valuable part of the climate record. The current status of the use of IASI for NWP is described in the next section.

2.3 The use of IASI in NWP

IASI radiances, generally expressed as brightness temperatures, are assimilated operationally at the majority of worldwide NWP forecast centres with advanced data assimilation schemes such as 4D-Var. IASI retrievals are used by several nowcasting (very short-range forecasting) applications. The large number of channels, high spectral resolution and low radiometric noise of IASI mean that much more information about the atmospheric structure is available than from previous operational sounders (Prunet et al., 1998; Collard, 1998). With improvements in the vertical resolution of forecasting models, the higher resolution information from IASI spectra is becoming increasingly important.

2.3.1 The status of operational assimilation schemes

Tests of the performance of IASI within forecasting systems have shown positive impact at least as great as for any previously introduced new instrument (Hilton et al., 2009a; Collard and McNally, 2009; Hilton et al., 2009c). Strong impacts have been found on geopotential height and mean sea-level pressure, among other variables, out to 5-6 days forecast lead time. The impact scores in global models have been particularly impressive because IASI is assimilated into a system which at the time was already very well observed by sounding data, with 10 microwave and 5 additional infrared sounders. IASI observations are increasingly used in limited area models, and have shown positive impact, including on forecasts of precipitation, in convective-scale models with 1.5–2.5 km horizontal resolution (Guidard et al., 2010).

Current operational assimilation schemes for IASI are conservative. They use up to only 200 channels, carefully chosen from the spectrum to maximize the information content of the observations, using methods such as those of Collard (2007). Operational timing constraints mean that assimilating many more channels is too computationally expensive. Chapter 4 discusses the information content of such a channel selection versus the full spectrum. Observations are also typically thinned out to minimise correlations of their errors in the horizontal domain. At the Met Office, one observation in four (the most homogeneous pixel from each field of regard) is processed via a 1D-Var scheme for quality control and retrieval of quantities

required in radiative transfer but not available in 4D-Var. After the 1D-Var step, further thinning takes place to reduce the observation density to 1 observation per 80 km in the extra-tropics and 1 observation per 154 km in the tropics.

Most centres assimilate channels where the effect of cloud in the field of view can be simulated using a simple cloud model. A common technique is to assimilate channels with weighting functions that peak above a diagnosed cloud top, and at the Met Office, channels with less than 10 % of their contribution from below the cloud top are used (Pavelin et al., 2008). Research is ongoing into more advanced use of cloud affected radiances that would allow IASI to influence analyses of cloud ice water and liquid water; early results show promise (Martinet et al., 2013).

Observations are typically assimilated over the sea surface by all centres, but treatement over sea ice and land surfaces vary. The Met Office scheme uses data over land, with spectral surface emissivity components and skin temperature retrieved during a 1D-Var pre-processing step (Pavelin and Candy, 2013). This is only done at night time because land surface temperatures are too much in error during the day to allow reasonable forward modelling of observations. Observations are not assimilated over sea ice. At ECMWF and Météo-France, observations are assimilated over sea ice but not over land.

Water vapour

The assimilation of most sources of satellite sounding data usually has little effect on humidity profiles in the boundary layer or on fine vertical structures. There are numerous factors which make assimilation of humidity information difficult, such as: ambiguity arising from water vapor channel sensitity to both temperature and humidity; non-linearity of the radiance response to the amount of water vapor present; and high spatial variability of the water vapor concentration. These issues make it a challenge for linear assimilation algorithms to handle the observations correctly if the model state is too far from the true atmospheric state.

At present, only a few (<40) IASI channels sensitive to water vapour molecules can generally be assimilated in NWP systems before the fit of the model to other observations and the numerical stability of the assimilation algorithm suffers (resulting in non-convergence or giving physically unrealistic values). To enable their assimilation, humidity sounding channels are usually assumed to have much larger errors than can be attributed to the instrument noise, and this restricts their impact on the humidity profile. The initial assimilation experiments for IASI in 2007 showed that the 31 water vapour channels provided only a 17 % increase in impact

versus observations when added to the 107 temperature sounding channels. However, recent results at the Met Office with the Cross-track Infrared Sounder (CrIS) on Suomi-NPP, an instrument producing similar data to IASI, have been more positive. Adding 44 water vapour channels to a baseline of 85 temperature sounding channels trebled the impact of CrIS when verified against conventional observations (Smith et al., 2015) (although it should be noted that the overall impact from IASI was much greater than that seen in the CrIS experiments). It is likely that incorrect observation error covariance matrices play an important part in the problems with assimilation of water vapour assimilation. This is discussed further in 2.3.2.

Ozone

Many NWP models now contain limited modelling of ozone for air quality forecasting, improved upper atmosphere winds through correlations with atmospheric dynamics, and improved radiative heating rates. Ozone is also important to the assimilation system, as many infrared channels that could be used for temperature sounding are weakly affected by the gas. Han and McNally (2010) report on the assimilation of IASI channels at the frequencies of strong ozone absorption lines. IASI observations, which have the advantage that they can be used to estimate ozone quantities at night, were found to have as strong an impact on the system as dedicated visible and ultra-violet ozone instruments, and the assimilation of the IASI ozone channels greatly improved the fit to stratospheric temperature sounding channels. Ozone is not an active variable in the Met Office 4D-Var at the present time.

2.3.2 Observation errors

One of the major conclusions of the 2009 ECMWF/EUMETSAT NWP-SAF "Workshop on the Assimilation of IASI in NWP" was that observation errors for hyperspectral sounders are not well characterised, and that it was not clear that they can be estimated with sufficient accuracy to ensure an optimal analysis (Garand et al., 2009). However, Bormann et al. (2010) and Stewart (2010) began to address ways of estimating the form of the observation error covariance structure for satellite radiances and its influence on the optimality of the NWP analysis. In 2013, during the course of this thesis, the Met Office implemented a diagnosed error covariance matrix for IASI assimilation Weston et al. (2014). The new error covariance matrix gave more weight to the observations because the diagonal no longer required

inflation to compensate for missing off-diagonal elements, yielding positive forecast benefit. This implementation paves the way for the assimilation of reconstructed radiance observations, which have a highly non-diagonal instrument noise covariance structure.

2.3.3 Current areas of research for IASI assimilation in NWP

There are four main areas of research being pursued by NWP centres at the present time relating to the assimilation of IASI observations:

- Extracting more information from the spectrum
- Improving assimilation in cloudy areas
- Increasing data coverage over non-sea surfaces
- Using the data in high-resolution limited area models

Extracting more information from the spectrum

This thesis addresses the first of these areas of research: how can we use more information from the IASI spectrum, improving vertical resolution, without significantly increasing computational costs. Although the IASI spectrum contains far fewer than 8461 pieces of information because of strong correlations in the spectral response of many channels, by using less than 2 % of the spectrum, as in current channel selections, the information content is reduced (see chapter 4). (It should be noted that even if there were 8461 pieces of information, unless the NWP profile consisted of 8461 uncorrelated variables and each channel was entirely independent, we could not hope to extract all of the information from the observation anyway.)

Principal component analysis (PCA) has received a great deal of attention as a means for compression of hyperspectral data and in NWP-related retrieval algorithms (e.g. Aires et al., 2002; Antonelli et al., 2004; Liu et al., 2007; Masiello et al., 2012). The theory is explained fully in chapter 6, but in summary PCA exploits the redundancy of information in a data set by performing a transformation to a new, uncorrelated set of variables (the PCs) where the first PC represents the spectral feature with the most variance, the second PC represents the feature with the second most variance and so on. Retaining just the first few hundred components preserves the majority of the spectral information, and the remaining components, consisting mostly of random instrument noise, may be discarded without significant loss for many applications. There are difficulties in the practical use of PCs (Collard et al., 2010), but their limited use has been demonstrated in an operational context

at ECMWF (Matricardi and McNally, 2014). This technique is explored briefly in chapter 9. An alternative is to reconstruct radiances from the PC scores, theoretically retaining the information content of the PCs themselves, but in a format for which quality control is easier to approach. This approach is explored in chapters 6 to 9.

Critical to maximising the information content will be to handle the error correlations between the channels correctly and thus the assimilation of compressed radiances relies heavily on the recent advances in this area both at the Met Office (Weston et al., 2014) and at ECMWF (Bormann et al., 2010) using observation-forecast and observation-analysis differences to derive diagnostic error covariance matrices.

Improving assimilation in cloudy areas

Cloudy data assimilation is currently only reliable in cases where a single layer grey cloud is a good approximation. The ability to analyse complex multi-layer cloud situations correctly is an active area of research. Work is also ongoing to develop more advanced schemes to make better use of cloud-affected data (e.g. Stengel et al., 2010; Guidard et al., 2010), but significant progress in this area is reliant upon developments in data assimilation schemes, radiative transfer models and NWP models themselves. Recent work by Martinet et al. (2013) has shown an encouraging start to the use of IASI observations to provide information to the assimilation system on hydrometeor profiles (in this case cloud liquid water and cloud ice) by adding in channels additional to the basic channel selection that help to differentiate these quantities. Ensemble data assimilation techniques may provide new opportunities to develop cloudy assimilation schemes.

Increasing data coverage over non-sea surfaces and using the data in high-resolution limited area models

Current data usage in the polar regions and large parts of the northern hemisphere is still rather low, owing to lack of knowledge of the surface properties over land and sea ice; progress in this area is being aided by the development of surface emissivity atlases.

The majority of NWP centres run limited area models (LAMs) covering their own country in greater detail than is possible in the global model. These LAMs usually contain a significant proportion of land points, so improving knowledge of non-sea surface properties is important to be able to use the observations effect-

ively. However, there are other significant challenges in using satellite data in these models, for example: the tops of the models are often much lower than is necessary to accurately forward model the observations; the model grid size is considerably smaller than the observation footprint and the data assimilation systems may have fundamental differences that filter out information on scales to which the instrument is sensitive.

2.4 IASI-NG, MTG-IRS and other future instruments

IASI was the first operational hyperspectral infrared sounder. The newer CrIS instrument is technologically similar, although it has broader spectral response functions and lacks full spectral coverage, so the data volumes are somewhat smaller than for IASI. The implementation of CrIS in NWP leans heavily on experience with IASI, which in turn followed experience with the research instrument AIRS (a grating spectrometer, but providing similar data to IASI and CrIS). It is important that we learn to maximise our use of these data over the next few years, because future instruments benefitting from technological advances will provide opportunities for new science, but present even greater challenges to NWP systems in terms of data volumes.

The EPS satellite programme currently has two Metop satellites in space, and the third and final satellite, Metop-C, will be launched in 2018. EUMETSAT are in the process of commissioning the EPS-Second Generation (EPS-SG) satellites, which will follow on from 2021. One of the crucial instruments for the EPS-SG satellite series will be a follow-on instrument to IASI, known as IASI-Next Generation (IASI-NG). The instrument specifications for IASI-NG require a factor of 2 increase in spectral resolution, and a factor of 2–4 increase in radiometric resolution. The main driving application for the increase in spectral resolution is atmospheric chemistry, as the new instrument will allow better definition of absorption lines, and thus better detection of weakly absorbing trace gas species.

For NWP, the major advancement will be in the improvement in radiometric noise, giving greater weight to the observations in the analysis if they are used correctly. There may be some further improvements from the increased spectral resolution, but since the CO_2 and H_2O lines are already well-resolved by IASI, this is likely to be a second-order improvement. On the other hand, the halving of the spectral resolution will result in a doubling of the number of channels. If

we were to continue with the way in which we use the IASI spectrum, this would mean assimilating around 1 % of the channels, and a significant loss of spectral information. It is therefore vital that we use IASI to learn how to condense the spectral information into a form which can be handled computationally by NWP systems.

EUMETSAT also have a geostationary satellite programme, and the development of the next generation of geostationary satellites, Meteosat Third Generation (MTG) is also under way. The satellites will have on board an infrared interferometer, the Infrared Sounder (IRS). Whilst lower in spectral resolution than IASI and without full spectral coverage, the IRS will generate vast data volumes (2500 spectra per second compared with 15 for IASI). It is already planned to use PCs to compress the spectra in order to reduce data volumes for dissemination, so understanding the implications of this process through experience with IASI will help to plan for the use of IRS data from 2020 onwards.

The US Satellite programme has undergone significant changes over the last ten years, and reorganisations and changes in funding allocation have resulted in significant delays to the Joint Polar Satellite Systems (JPSS) satellite program. The Suomi-NPP satellite has now become the first operational satellite of the JPSS series, and the future JPSS satellites also carry CrIS. It is anticipated that post-JPSS satellites will also carry hyperspectral IR sounders of similar specification to IASI-NG.

The Chinese Meteorological Agency's satellite programme is advancing at a great rate, and the FY-3D polar orbiting satellite, to be launched in 2015 will carry an IR interferometer. In addition the FY-4 satellite series, of which the first is due to be launched in 2016 will likely carry the first operational IR interferometer in geostationary orbit. The Russian meteorological satellite agency, RosHydroMet also plan to put interferometers into polar and geostationary orbit within the next five years. The World Meteorological Organisation's "Vision for the Global Observing System in 2025" calls for a hyperspectral IR sounder on each of a ring of six geostationary satellites, and on polar orbiting satellites in three sun-synchronous orbits.

2.5 Summary

IASI and other hyperspectral sounders have been shown to provide significant impact to NWP. The current assimilation schemes are conservative and by improving access to data over land and sea ice and in cloudy conditions, we may begin to exploit

the observations more fully. There are further gains to be made by using more of the observed spectrum: current assimilation schemes use less than 2% of the channels. The work presented in this thesis addresses the exploitation of the spectral information from IASI.

It is clear that hyperspectral infrared sounders will form an important part of the observing system for NWP in the coming decades. It is therefore important that we begin to make significant progress in exploiting these observations to extract the maximum benefit from their assimilation in the most efficient manner possible.

Chapter 3

Data Assimilation

As described in chapter 1, observations play a critical role in the production of NWP forecasts. The role of data assimilation is to find the best possible representation of the atmosphere from which to start the forecast runs, combining observations with prior information (the background). This statistically optimal estimate is referred to as the analysis, and in global DA is usually calculated four times a day with six-hour batches of observations. The NWP model is run forward from the analysis to produce forecast products and also the background state for the next DA cycle.

The background is a short-range forecast of the three-dimensional atmospheric state represented on some kind of grid; for global NWP models, a 6 hour forecast is usual. The background is usually already a very good estimate of the atmospheric state and contains a weighted average of information from observations gathered over approximately the previous three to ten days (Fisher and Auvinen, 2011; Fisher et al., 2006). Data assimilation is more than an exercise in interpolation, because the model state is already fairly accurate and the observations have errors associated with them. Large data gaps may exist (for example over sparsely populated land masses and in polar regions) and it is necessary for the analysis to be a smooth field without any gaps or abrupt changes in properties. Furthermore, the variables that make up the atmospheric state are not independent from each other, but instead are related by the equations of atmospheric physics, and these relationships must hold in the analysis.

One of the most common DA techniques used in operational meteorology is variational analysis (Var), which is the current formulation of the Met Office data assimilation system and will be discussed in the next section. An alternative family of methods based on approximating the Kalman filter (Kalman, 1960) with ensemble-based forecasting systems (the Ensemble Kalman Filter, or EnKF, described by Evensen, 2007) have become more prevalent in recent times. Most NWP centres, including the Met Office, are now moving towards combining EnKF based systems with Var: a review of these methods will be given in section 3.7.

3.1 Variational assimilation theory

A good background to variational assimilation is given in Daley (1991) and Kalnay (2003), in which all of the equations presented in this section are fully derived starting from Bayesian statistics. Much of data assimilation theory is the same as retrieval theory and the equations of DA are related to those presented in Rodgers (2000), but the notation of Ide et al. (1997) is generally used for DA.

The analysis state sought by the DA algorithm is an optimally weighted average of the background and observations. The basis for both Var and the Kalman filter approach is Bayesian statistics and conditional probability. The Var scheme finds the maximum a posteriori (MAP) solution, in other words the most likely representation of the state of the atmosphere given the observation and the prior information (background). If $\mathbf{x}$ is the state of the atmosphere, and superscripts b and a define the background and analysis respectively, the probability density function (pdf) of the analysis (referred to as the posterior pdf, P^a), given the observations is:

$$P^a(\mathbf{x}|\mathbf{y}^o) \propto P(\mathbf{y}^o|\mathbf{x})P^b(\mathbf{x}) \tag{3.1}$$

where $\mathbf{y}^o$ are the observations and P^b is the pdf of the prior state, i.e. the background. Full derivations are given in the standard texts, but assuming that the pdfs can be modelled by a mean and covariance (it is usual to assume a Gaussian distribution), the probability can be evaluated as follows:

$$P^a(\mathbf{x}|\mathbf{y}^o) \propto e^{-\frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b)-\frac{1}{2}(\mathbf{y}-\mathbf{y}^o)^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-\mathbf{y}^o)} \tag{3.2}$$

where $\mathbf{y}$ is the model prediction of the observation. The weights given to the information contained in the prior state (the background) and in the observations are inversely proportional to the error covariance in those quantities, respectively $\mathbf{B}$ and $\mathbf{R}$, much as in a simple least squares fit. The maximum likelihood solution is equivalent to the minimum variance solution under the assumption of Gaussian errors.

Var finds the most likely state of the atmosphere to fit these conditions by minimising a cost function, J, that is the logarithm of equation 3.2:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) + \frac{1}{2}(\mathbf{y}-\mathbf{y}^o)^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-\mathbf{y}^o) \tag{3.3}$$

3.2 3D-Var

3D-Var finds the global atmospheric state that best fits the full set of observations simultaneously. Although still in use for operational NWP, usually for cost reasons and especially for convective scale model configurations, the main disadvantage of 3D-Var is that although observations are spread out in time throughout the assimilation window, the background errors, $\mathbf{B}$, are fixed at analysis time, usually the mid-point of the assimilation window. Also, the model equivalent observations, $\mathbf{y}$, are calculated from a model state at the analysis time (unless the incremental formulation described in section 3.4 is used). 4D-Var (section 3.3) is preferred because it extends the equations to take account of the model trajectory over the assimilation time window and evolves $\mathbf{B}$ along this trajectory. The equations of 3D-Var provide the starting point, and are particularly relevant to this thesis because they also describe the single column equivalent, or 1D-Var, which is used in this work (section 3.8).

In order to calculate the observation departures, $\mathbf{y} - \mathbf{y}^o$, the model state $\mathbf{x}$ needs to be interpolated to the observation location. For many observation types the observation is not directly sensitive to model variables and a forward model, $H(\mathbf{x})$, is needed to carry out these operations:

$$\mathbf{y} = H(\mathbf{x}) \tag{3.4}$$

For satellite radiances this incorporates a radiative transfer model such as RTTOV (Saunders et al., 2010). The cost function, equation 3.3, is now written:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(H(\mathbf{x}) - \mathbf{y}^o)^{\mathrm{T}}\mathbf{R}^{-1}(H(\mathbf{x}) - \mathbf{y}^o) \tag{3.5}$$

The left-hand term of equation 3.5 is known as the background penalty, J^b, and the right hand term is the observation penalty, J_o. The solution is found by iterating numerically towards the minimum of the cost function using the gradient, which is zero at the minimum:

$$\nabla_{\mathbf{x}} J(\mathbf{x}) = \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(H(\mathbf{x}) - \mathbf{y}^o) \tag{3.6}$$

where $\mathbf{H}$ is the gradient of the observation operator, $H(\mathbf{x})$.

A numerical iterative method is used to approach the minimum of the cost function. A conjugate gradient method (Fisher, 1997) is currently used in global

4D-Var at the Met Office, and the quasi-newtonian L-BFGS[1] method is used in the UK variable resolution (UKV) configuration 3D-Var: these are described in Payne (2007). This enables the assimilation problem to cope with a degree of non-linearity, as $\mathbf{H}$ can be recalculated during the descent. If the forward model is linear such that

$$H(\mathbf{x}) = H(\mathbf{x}^b) + \mathbf{H}(\mathbf{x} - \mathbf{x}^b) \tag{3.7}$$

and assuming the variational algorithm iterates to the optimal (minimum variance) solution, an expression for the analysis, $\mathbf{x}^a$, can be obtained by setting equation 3.6 equal to 0:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - H(\mathbf{x}^b)) \tag{3.8}$$

where:

$$\mathbf{K} = (\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1} \tag{3.9}$$

$\mathbf{K}$ is the optimum weight matrix, also known as the Kalman gain, and can be equivalently formulated:

$$(\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{T})^{-1} \tag{3.10}$$

The analysis error of the linear minimum variance formulation is given by:

$$\begin{aligned} \mathbf{A} &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^{\mathrm{T}} + \mathbf{K}\mathbf{R}\mathbf{K}^{\mathrm{T}} \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} \end{aligned} \tag{3.11}$$

$\mathbf{A}$ can also be derived using the following formula, which demonstrates clearly that $\mathbf{A}$ is equivalent to the result of a least-squares fit between two parameters:

$$\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H} \tag{3.12}$$

The equations The 3.8 to 3.12 are of little practical use for a full NWP model with a billion degrees of freedom because the matrix manipulation is impossible for a system of that size (see section 3.5.2), but are often used in analysis of 1D-Var problems. Section 3.8 describes how they are used in chapters 5, 8 and 10 to examine the impact of IASI observations in various 1D-Var configurations, as a predictor for

[1] limited memory Broyden-Fletcher-Goldfarb-Shanno

behaviour in 4D-Var.

3.3 4D-Var

The 3D-Var cost function, equation 3.5, and its gradient, equation 3.6, can be extended to take account of the evolution of the modelled atmospheric state throughout the assimilation window. The background and analysis are typically defined at the start of the assimilation window, time $t0$, but the observation departures are calculated at the time of the observation by propagating the latest guess, $\mathbf{x}_{t0}$, with the NWP model, $\underline{M}$:

$$\underline{\mathbf{x}} = \underline{M}(\mathbf{x}_{t0}) \tag{3.13}$$

The underline symbol is used to incorporate a time dimension into the vector to denote a 4D field. Extending $H(\mathbf{x})$ slightly to include a time interpolation because the model state is usually sampled at intervals throughout the time window, the cost function now becomes:

$$J(\mathbf{x}_{t0}) = \frac{1}{2}(\mathbf{x}_{t0} - \mathbf{x}^b_{t0})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_{t0} - \mathbf{x}^b_{t0}) + \frac{1}{2}(H(\underline{\mathbf{x}}) - \underline{\mathbf{y}}^o)^{\mathrm{T}}\mathbf{R}^{-1}(H(\underline{\mathbf{x}}) - \underline{\mathbf{y}}^o) \tag{3.14}$$

Note that the underline symbol is also added to the observation vectors, to make it clear the the observations are spread through the assimilation window. The cost function gradient now becomes:

$$\nabla_{\mathbf{x}_{t0}} J(\mathbf{x}_{t0}) = \mathbf{B}^{-1}(\mathbf{x}_{t0} - \mathbf{x}^b_{t0}) + \underline{\mathbf{M}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(H(\underline{\mathbf{x}}) - \underline{\mathbf{y}}^o) \tag{3.15}$$

where $\underline{\mathbf{M}}$ is the gradient of the model, $\underline{M}$, and $\underline{\mathbf{M}}^{\mathrm{T}}$ represents its adjoint. The adjoint is used to evolve the weights given to the observations backwards in time to generate increments at the start of the assimilation window and thus calculate the analysis at $t0$.

Although the subscript $t0$ is used in equations 3.14 and 3.15 to indicate the start of the assimilation window, from this point on the subscript will mostly be omitted for brevity.

3.4 Incremental 4D-Var

In practice, NWP centres adopt a formulation known as incremental 4D-Var. Rather than solving for the full atmospheric state, the solution is found for the departure of the analysis state from the background, in other words, for a model perturbation. The behaviour of this perturbation is assumed to be linear for small departures, and its evolution in time is represented by a linear model. A full derivation of this approach, also known as statistical 4D-Var, is given in Lorenc (2003). The main features of the approach are summarised below.

For deterministic 4D-Var, time evolution is accounted for by the NWP model and, in theory, this should be used with its full adjoint to evolve the pdf of the background state in time through the assimilation window. This is not at all practical, and so in general a linear model is used in 4D-Var to evolve the perturbations. This is usually a linearised approximation of the full NWP model known as the tangent linear model, and is generally run at a lower resolution than the full NWP model. The model is linearised about a series of short range forecasts from the full NWP model, known as linearisation states, distributed throughout the assimilation window, which act as coefficients for the tangent linear model.

The equations of incremental variational analysis are equivalent to those presented in section 3.3, but are now expressed as a function of model perturbations away from $\mathbf{x}^g$, where the superscript g indicates that this state is our current best guess of the atmospheric state. The perturbation is written $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^g$ (defined at the start of the assimilation window), along with the background increment, $\delta\mathbf{x}^b = \mathbf{x}^b - \mathbf{x}^g$, and equation 3.5 is rewritten:

$$J(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x}-\delta\mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\delta\mathbf{x}-\delta\mathbf{x}^b)+\frac{1}{2}(H(\underline{\mathbf{x}}^g)+\mathbf{H}\delta\underline{\mathbf{x}}-\underline{\mathbf{y}}^o)^{\mathrm{T}}\mathbf{R}^{-1}(H(\underline{\mathbf{x}}^g)+\mathbf{H}\delta\underline{\mathbf{x}}-\underline{\mathbf{y}}^o) \tag{3.16}$$

and

$$\nabla_{\delta\mathbf{x}}J(\delta\mathbf{x}) = \mathbf{B}^{-1}(\delta\mathbf{x} - \delta\mathbf{x}^b) + \underline{\mathbf{M}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(H(\underline{\mathbf{x}}^g) + \mathbf{H}\delta\underline{\mathbf{x}} - \underline{\mathbf{y}}^o) \tag{3.17}$$

Note that the gradient or Jacobian of the forward model, $\mathbf{H}$, is used to propagate the increment here, along with adjoint of $\underline{\mathbf{M}}$, the tangent linear model. At the start of the minimisation, $\mathbf{x}^g$ is normally equal to $\mathbf{x}^b$. It is updated following an outer loop cycle, as explained in the next section.

3.4.1 Inner and outer loops

Incremental 4D-Var only finds a solution within the bounds of the linear model used to propagate the increments. It is possible to allow for some non-linearity in the progress towards the analysis by re-linearising the tangent linear model every few iterations. The minimisation of the cost function is performed via the linear inner loop but an outer loop can be run, where $\underline{\mathbf{x}}^g$ is recalculated by running the full NWP model from the updated model state calculated at the end of the inner loop, i.e.

$$\underline{\mathbf{x}}^g = \underline{M}(\mathbf{x}^g_{t0}) \tag{3.18}$$

Rather than recalculating the observation departure from the current state for every iteration of the inner loop, the observation departure from the guess, $\underline{\mathbf{d}}^o$, is usually calculated before the inner loop:

$$\underline{\mathbf{d}}^o = \underline{\mathbf{y}}^o - H(\underline{\mathbf{x}}^g) \tag{3.19}$$

During the inner loop, the tangent linear model is then used to propagate only the increment forward through the assimilation window:

$$\delta\underline{\mathbf{x}} = \underline{\mathbf{M}}(\delta\mathbf{x}_{t0}) \tag{3.20}$$

and this is acted upon by the forward model Jacobian to give the observation penalty, J_o, and its gradient:

$$J^o(\delta\mathbf{x}) = \frac{1}{2}(\mathbf{H}\delta\underline{\mathbf{x}} - \underline{\mathbf{d}}^o)^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\underline{\mathbf{x}} - \underline{\mathbf{d}}^o) \tag{3.21}$$

and

$$\nabla_{\delta\mathbf{x}} J^o(\delta\mathbf{x}) = \underline{\mathbf{M}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\underline{\mathbf{x}} - \underline{\mathbf{d}}^o) \tag{3.22}$$

The incremental formulation can also be applied to 3D-Var. This enables better use of the observations, as equation 3.19 can be used to provide the first guess at the appropriate time (this system is thus known as 3D-Var FGAT).

3.4.2 Theoretical justification for incremental 4D-Var

Lorenc and Payne (2007) provide a theoretical justification for incremental 4D-

Var in its own right, rather than being simply an approximation to deterministic 4D-Var. The deterministic equations presented in section 3.3 find the most likely solution to a linear approximation to the full problem, because the model error evolution and observation operator evolution are assumed to be linear. In reality, even though the starting assumption is that the background errors are Gaussian, the model is non-linear, so the final pdf is non-Gaussian and the most likely solution will not be the best solution by any mathematical definition. In the incremental implementation, the mean of the background error pdf is evolved using the full non-linear model, and the linear model provides the best estimate of the variation about this mean. The incremental 4D-Var analysis thus finds the most likely solution (i.e. the mode) of a simplified pdf, but with the correct mean. Because the final pdf remains Gaussian, the mode of the analysis error pdf is also the mean. Furthermore, the linear approximation can be designed to filter the analysis to include only scales that behave close to linearly. The Met Office implementation of 4D-Var does this using a linear model of reduced complexity, as described in section 3.5.

3.18 to 3.21 show that in incremental 4D-Var the linear model is only used for calculating the analysis corrections to the current guess, but not in the evaluation of the full guess trajectory. The linear model can be seen as a way of propagating the background pdf into four dimensions. In other words, the time evolution of the atmospheric state is calculated using the full model, whilst only the evolution of the error covariances uses the linear model.

Propagating the error covariances using a model means that the analysis is constrained by the relationships that are contained within the model: this is known as using the model as a strong constraint on the solution. If model error is incorporated into the equations, the solution is no longer required to fit the model exactly, and the formulation is known as weak constraint. However, model error is very poorly understood, and in practice grows slowly enough over the six hour assimilation window that it is usually neglected.

3.5 The Met Office 4D-Var

This section describes the way in which incremental 4D-Var is implemented at the Met Office. There are a few key differences relative to other centres, which are described in sections 3.5.1 and 3.5.2. Section 3.5.3 then describes the control variable transform that in turn explains how the background errors are modelled in the operational system.

3.5.1 Inner and outer loops, and the observation operator

Most NWP centres formulate the inner loop as described in 3.19 to 3.22. At the Met Office, the inner loop incorporates a non-linear observation operator, to give the following J^o and gradient terms:

$$J^o(\delta\mathbf{x}) = \frac{1}{2}(H(\delta\underline{\mathbf{x}} + \underline{\mathbf{x}}^g) - \underline{\mathbf{y}}^o)^{\mathrm{T}}\mathbf{R}^{-1}(H(\delta\underline{\mathbf{x}} + \underline{\mathbf{x}}^g) - \underline{\mathbf{y}}^o) \tag{3.23}$$

and

$$\nabla_{\delta\mathbf{x}} J^o(\delta\mathbf{x}) = \underline{\mathbf{M}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(H(\delta\underline{\mathbf{x}} + \underline{\mathbf{x}}^g) - \underline{\mathbf{y}}^o) \tag{3.24}$$

Although this increases the cost of the inner loop slightly because the non-linear observation operator must be used, an inner inner loop (or quadratic loop) provides cost savings by retaining a linear observation operator for ten iterations before re-calculating $\mathbf{H}$ and the observation departures.

The Met Office 4D-Var has no outer loop, so $\mathbf{x}^g$ and $\underline{\mathbf{x}}^g$ are not updated following the inner loop. Thus $\mathbf{x}^g = \mathbf{x}^b$ throughout the minimisation.

3.5.2 The perturbation forecast model

Instead of a tangent linear approximation to the full model, a new simplified linear model with lower resolution and reduced complexity is used, known as the perturbation forecast (PF) model. Incremental 4D-Var already has cost advantages because the perturbuation can be calculated at a lower resolution than is required for the full model state (Courtier et al., 1994; Lorenc et al., 2000). The less complex PF model provides further cost savings: minimisation problems may be reduced because, as described in section 3.4.2, the PF model is designed to filter the analysis increments to only those modes and scales that behave close to linearly. Furthermore, as will be described in section 3.6.1, the background errors are only sampled and approximated, and so there is little justification in using a full high resolution $\underline{M}$ to propagate this approximate $\mathbf{B}$. The linear PF model is denoted $\underline{\tilde{\mathbf{M}}}$ to show that it is not the Jacobian of $\underline{M}$

The reduced complexity of the PF model mean that the analysis increments are not calculated for the same atmospheric variables that are used in the full NWP model. A simplification operator, S, is defined, which interpolates to the lower resolution and maps full model variables into the PF model state. An example of this mapping is in the construction of the total humidity variable, q_t, from the full model's

specific humidity, cloud liquid water and cloud ice water variables. This approach avoids having to model the cross-correlation terms between the different humidity components in the $\mathbf{B}$ matrix.

The vector $\underline{\mathbf{y}}$ is calculated from the model state at the start of the assimilation window, $\mathbf{x}_{t0}$, by the following series of operations:

$$\underline{\mathbf{y}} = H(L(\underline{M}(\mathbf{x}_{t0})) + \mathbf{S}^{-I}(\tilde{\mathbf{L}}\underline{\tilde{\mathbf{M}}}\delta\mathbf{w})) \tag{3.25}$$

In this equation, L is a horizontal and time interpolation operator from the full model to the observation positions, and $\tilde{\mathbf{L}}$ is the equivalent PF operator for the model perturbations. $\mathbf{S}^{-I}$, known as the incrementing operator, is the generalised inverse of $\mathbf{S}$. Using the simplification operator S, equation 3.16 is rewritten in terms of the simplified state vector $\mathbf{w}$:

$$J(\delta\mathbf{w}) = \frac{1}{2}(\delta\mathbf{w} - \delta\mathbf{w}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\delta\mathbf{w} - \delta\mathbf{w}^b) + \frac{1}{2}(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o)^{\mathrm{T}}\mathbf{R}^{-1}(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o) \tag{3.26}$$

where

$$\delta\mathbf{w}^b = \mathbf{w}_b - \mathbf{w}_g = S(\mathbf{x}^b) - S(\mathbf{x}^g) \simeq \mathbf{S}\delta\mathbf{x}^b \tag{3.27}$$

and $\mathbf{S}$ is the tangent linear of the operator S evaluated at $\mathbf{x}^g$ and $\mathbf{B}$ is defined for the PF model background state. $\delta\mathbf{w}$ is equivalently formulated for the current iteration using $\mathbf{x}$ in place of $\mathbf{x}^b$.

In practice, it is not possible to manipulate, or even hold in memory, the full 4D error covariance, $\mathbf{B}$, in equation 3.26. Furthermore, it would not be possible to calculate, as the model would need to be run more than 10^8 times to generate sufficient statistics to represent the full matrix. If it were possible to have a full 4D $\mathbf{B}$, its structure would have to reflect the coupling of fields at different time steps that occurs through the model equations as states which do not fit these coupling relationships are very unlikely. Given that these relationships cannot be represented statistically, the practical approach taken is to constrain the solution using the model itself: the 4D state used in equation 3.25 above, $\underline{\tilde{\mathbf{M}}}\delta\mathbf{w}_0$, is constrained by the form of $\underline{\tilde{\mathbf{M}}}$, and the background error covariance is represented as $\underline{\tilde{\mathbf{M}}}\mathbf{B}\underline{\tilde{\mathbf{M}}}^{\mathrm{T}}$, using the PF model as a strong constraint on the solution.

3.5.3 The 4D-Var control vector

The control vector is the vector of quantities for which the Var system is finding a solution. In 1D-Var retrieval applications, the control vector is usually the state vector $\mathbf{x}$, composed of atmospheric variables. In the case of retrievals from satellite data, the variables in the state vector are often the same as, or similar to, those required for input to the radiative transfer equation. For example, one may retrieve a profile of temperatures, and humidity mass mixing ratios, or in the case of trace gas constituent analysis, the temperature and humidity may be assumed to be already known, and the state vector may consist of a profile or total column amount of the relevant gas. This approach is reasonable since, for single column retrievals, the complete state vector is rarely more than a couple of hundred elements long.

In 3D- or 4D-Var, where a solution is sought for the whole globe, the state vector can easily be 10^8 elements long, and this number is increasing all the time with improved resolution models. As mentioned in the previous section, it is not possible to specify a full error covariance matrix for a state vector this large. The solution is to reduce the complexity of the analysis problem. The simplification operator, S, goes some way to achieving this, but the PF model variables are still highly correlated and the dimensionality is still too large. To reduce the analysis problem to a manageable size, linear variable transforms are used to construct a control vector whose elements are uncorrelated, vastly reducing the computation required to solve the variational analysis equations.

Rather than minimise the cost function of $\delta\mathbf{w}$, a control vector $\mathbf{v}$ is used, whose elements are uncorrelated. The control vector is mapped into PF model perturbation space by the use of the control variable transform matrix, $\mathbf{U}$, where $\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{B}$:

$$\delta\mathbf{w} = \mathbf{U}\mathbf{v} \tag{3.28}$$

In this usage, $\mathbf{B}$ is not really a matrix, but is essentially a set of statistics that define relationships between observed short-range forecast errors, which are used as coefficients for the $\mathbf{U}$ transform. The transformed 4D-var cost function (equation 3.26) can now be rewritten:

$$\begin{aligned} J(\mathbf{v}) &= \frac{1}{2}(\mathbf{U}\mathbf{v} - \mathbf{U}\mathbf{v}^b)^{\mathrm{T}}(\mathbf{U}\mathbf{U}^{\mathrm{T}})^{-1}(\mathbf{U}\mathbf{v} - \mathbf{U}\mathbf{v}^b) + \frac{1}{2}(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o)^{\mathrm{T}}\mathbf{R}^{-1}(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o) \\ &= \frac{1}{2}(\mathbf{v} - \mathbf{v}^b)^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}(\mathbf{U}\mathbf{U}^{\mathrm{T}})^{-1}\mathbf{U}(\mathbf{v} - \mathbf{v}_b) + \frac{1}{2}(\mathbf{y} - \mathbf{y}^o)^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{y}^o) \end{aligned} \tag{3.29}$$

Since $\mathbf{U}^{\mathrm{T}}(\mathbf{U}\mathbf{U}^{\mathrm{T}})^{-1}\mathbf{U} = \mathbf{I}$, each element of $\mathbf{v}$ is both uncorrelated and has unit

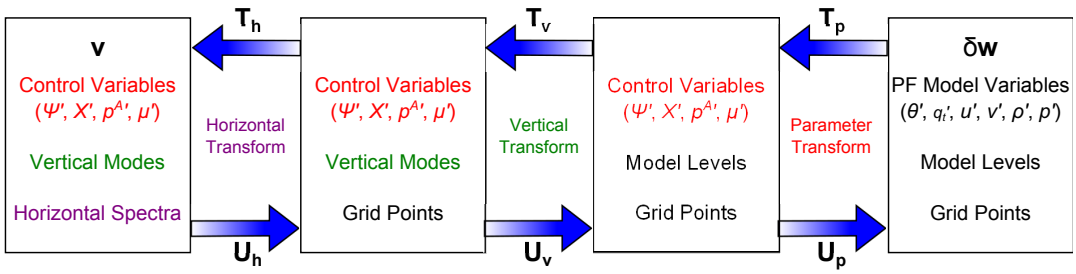


Figure 3.1: Diagramatic representation of the control variable transforms used in the Met Office 4D-Var

variance (although it is worth noting that this transform is a model, and the behaviour of $\delta\mathbf{w}$ is not perfectly represented). This further simplifies equation 3.29 to give

$$J(\mathbf{v}) = \frac{1}{2}(\mathbf{v}-\mathbf{v}^b)^{\mathrm{T}}(\mathbf{v}-\mathbf{v}^b) + \frac{1}{2}(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o)^{\mathrm{T}}\mathbf{R}^{-1}(\underline{\mathbf{y}}-\underline{\mathbf{y}}^o) \tag{3.30}$$

and the J^b term now needs no matrix multiplication steps.

The $\mathbf{U}$ transform, and its adjoint, $\mathbf{U}^{\mathrm{T}}$ are used in the minimisation, but in order to calculate $\mathbf{v}$ and $\mathbf{v}^b$, the reverse transform, mapping PF model perturbations into control vector space, is needed. The reverse transform is represented by the matrix $\mathbf{T}$, which is the left inverse of $\mathbf{U}$, i.e. $\mathbf{TU} = \mathbf{I}$:

$$\mathbf{v} = \mathbf{T}\delta\mathbf{w} \tag{3.31}$$

The control variable transform is currently under review, but the present operational scheme is split into three steps. In order of operation on $\mathbf{U}$ these are: the horizontal transform $\mathbf{U}_h$, which maps from model grid points onto horizontal spectra; the vertical transform $\mathbf{U}_v$, which maps from model levels to empirical vertical modes; and $\mathbf{U}_p$, the parameter transform which maps from control variables to PF model variables:

$$\delta\mathbf{w} = \mathbf{U}\mathbf{v} = \mathbf{U}_p\mathbf{U}_v\mathbf{U}_h\mathbf{v} \tag{3.32}$$

$$\mathbf{v} = \mathbf{T}\delta\mathbf{w} = \mathbf{T}_h\mathbf{T}_v\mathbf{T}_p\delta\mathbf{w} \tag{3.33}$$

This is summarised graphically in figure 3.1, and the three steps of the control variable transforms are explained in more detail below. The transforms are most easily described in terms of their $\mathbf{T}$ forms.

The parameter transform, $\mathbf{T}_p$

$\mathbf{T}_p$ takes correlated PF model variable increments, $\delta\mathbf{w}$, that we wish to analyse in 4D-Var, and transforms them to the non-correlated control vector space, $\mathbf{v}$. The control vector is defined through physical understanding of atmospheric processes to yield a set of variables which are independent. The transform is also designed so that the analysis finds the best solution for the balanced portion of the atmosphere; unbalanced features that do not obey the dynamical relations specified in the PF model, such as increases in cloud amount coinciding with a reduction in humidity will be both poorly modelled and unobserved. Unconstrained increments such as this would lead to an ill-conditioned solution, and it is best to omit these from the analysis.

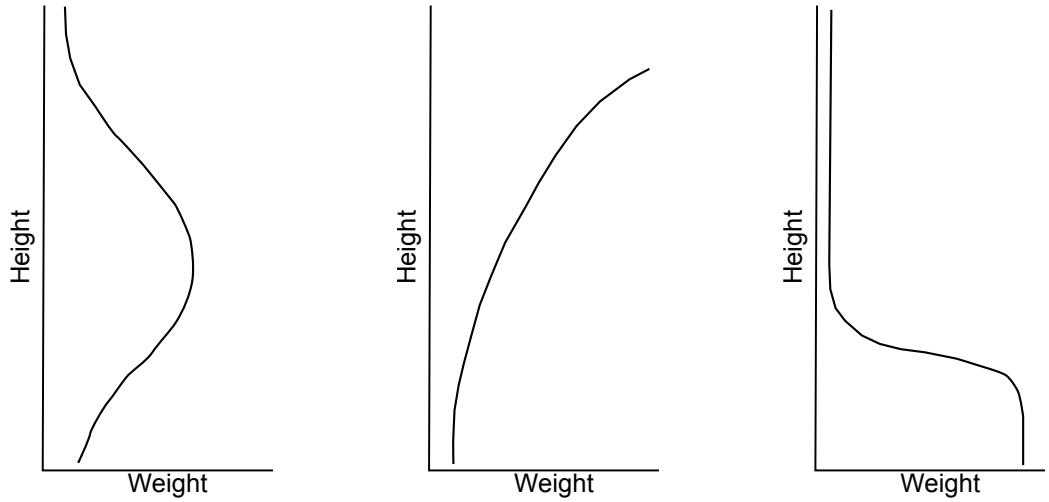
The PF model variables are the three components of the wind field u', v' and w', potential temperature θ', density ρ', pressure p', and total humidity q_t' (combining water vapour, cloud liquid water and cloud ice), where the primes indicate that the quantities are perturbations rather than full model fields. The uncorrelated control vector variables are streamfunction and velocity potential, ψ' and χ' respectively, which are a function of u' and v'; unbalanced pressure, $p^{A\prime}$, which depends upon u', v', θ', q_t', ρ' and p'; and a humidity analysis variable, μ', which is a function of p', θ' and q_t'. This transform can be summarised

$$(u', v', \theta', \rho', p', q_t') \Rightarrow (\psi', \chi', p^{A\prime}, \mu')$$

Once the parameter transform has taken place, the background error covariance matrix, $\mathbf{B}_v$, is block diagonal. The control vector is now uncorrelated between variables, but each variable has vertical and horizontal correlations across the PF model grid.

The vertical transform, $\mathbf{T}_v$

For each state vector variable, the vertical transform is estimated using statistical techniques from zonally and seasonally averaged model states, compiled either from 30h-6h forecast differences (the NMC method; Parrish and Derber, 1992) or using a collection of model states from an ensemble prediction system (Piccolo, 2011). The vertical transform projects from model levels onto vertical modes for each control variable. This is similar to an eigenvector analysis, except that the modes are defined such that atmospheric layers are given weights by a diagonal inner product matrix, $\mathbf{P}$. This is partly because the model levels are not evenly distributed in pressure



(a) Δp-form; Used for ψ and χ

(b) $^{\Delta p}/_{p}$-form; Used for p^A

(c) Inner product used for μ

Figure 3.2: Schematic diagrams representing the general form of the diagonal of the inner product matrices used to weight the vertical levels in the analysis

space, and using eigenvectors of $\mathbf{B}_v$ would give unsatisfactory weight to thicker layers. Instead the inner product is applied as follows:

$$\mathbf{B}'_v = \mathbf{P}^{1/2}\mathbf{B}_v\mathbf{P}^{T/2} \tag{3.34}$$

The vertical modes are defined by $\mathbf{F}$, the matrix of eigenvectors of $\mathbf{B}_v$, and their associated eigenvalues $\mathbf{\Lambda}$, within the inner product space of $\mathbf{P}$:

$$\mathbf{B}'_v = \mathbf{F}\mathbf{\Lambda}\mathbf{F}^{\mathrm{T}} \tag{3.35}$$

The form of the inner product matrix varies depending on the control variable. Figure 3.2(a) illustrates the form of the inner product used for ψ and χ, which is related to Δp; figure 3.2(b) illustrates the inner product for p^A, related to $^{\Delta p}/_{p}$; and figure 3.2(c) for μ.

The inner products and vertical modes are defined globally. In other words, the same structures are represented whether at the pole or at the equator. However, the variance associated with each mode (i.e. $\mathbf{\Lambda}$ in equation 3.35) varies latitudinally in 5° boxes, which allows for some variability in the vertical correlations of $\mathbf{B}_v$.

The horizontal transform, $\mathbf{T}_h$

The horizontal transform for each vertical mode of each state vector variable maps from model grid point space into horizontal spectral functions. The spectral functions are orthogonal, much as the decomposition used in the vertical transform. The horizontal transform removes correlations between grid points for each vertical error mode. The spectral functions also allow the increments to be filtered, to remove very small scales in the horizontal at the poles where model grid points are very closely spaced. This is analogous to removing unbalanced variables that are poorly modelled and unobserved via the parameter transform.

3.6 Some assumptions that are made in variational analysis

3.6.1 The error covariance terms

Finding the optimal solution to 4D-Var depends upon the background and observation errors being correctly defined. These are important assumptions, the impact of which are explored in this thesis in the context of IASI assimilation.

The background error term

The background error term is extremely important in the above equations. Although the optimum weights are dependent on both observation and background error, it is the background error term which acts as a filter on the analysis increment. The analysis increment can only occur in the model subspace spanned by $\mathbf{B}$. Daley (1991) describes $\mathbf{B}$ as "the most important element in the interpolation algorithm", adding that "to a large extent the form of this matrix governs the resulting objective analysis". This is often explained by looking at the right-hand formula for the analysis gain matrix defined in equation 3.10 and pointing out that, in the matrix multiplication, $\mathbf{B}$ is the last term to act as an operator on the observation increment, projecting it onto the directions spanned by the background error covariance. The importance of $\mathbf{B}$ in defining the form of the analysis is evident when comparing the results of different retrieval schemes (Hilton et al., 2009a).

Where the model is insensitive to perturbations in the atmospheric structure, this is known as null space (the observations may have their own null space if they are insensitive, or there is ambiguity in their sensitivity, to atmospheric perturbations).

Components in the near-null space have assumed background errors so low that it is not possible to make increments in those directions directly, although in strong-constraint 4D-Var, the behaviour in one variable can be forced to follow another (e.g. the model can generate wind increments from a temperature observation). The subspace spanned by **B** may not be optimal for the use of high-vertical resolution observations. For example, if there is a mismatch in the scales of the observation and of the background error, information from the observation may be unable to influence the analysis.

As described in sections 3.5.2 and 3.5.3, the background error term, although represented by its matrix form **B**, is in fact a set of coefficients for a series of operations to project the atmospheric state onto a control vector of variables uncorrelated in state, horizontal and vertical space. Operational data assimilation is a challenging area of work, because the stable running of the forecast model is of paramount importance and so any modifications made to its initial state must be smooth and realistic within the bounds of the model parameterisations. The performance of the DA system itself is also fundamental, with computational time constraints and practical considerations often playing a part in specifying aspects of the system. One of the greatest challenges in DA is to constrain the form of the transforming operations to represent the true errors in the background state, whilst maintaining stable performance of the forecast model.

One of the major difficulties is to incorporate flow-dependence into the **B** term. Until 2010, most NWP centres with 4D-Var schemes used a climatological set of statistics to represent the error covariance, calculated using the NMC method or an ensemble formulation, as described in section 3.5.3. Since then there has been a gradual shift towards the use of ensemble forecasting methods, which allow incorporation of uncertainty in the present atmospheric state into the covariance model (see section 3.7). Even so, with current computing power, ensembles are small and these dynamically produced error terms are undersampled. Thus, the process of estimation of the background error, and our inability to properly model its synoptic evolution mean that the **B** term will inevitably be misspecified. If the errors are misspecified, the analysis will no longer be optimal, since it can only follow the estimated **B**. Chapter 10 investigates the possible effects of an incorrect **B** on the impact of IASI data (whilst assuming that the observation errors, **R**, are correct, and that the forward model Jacobian **H** is accurate).

The observation error term

The general question of well-characterised observation errors has received attention in the past from Watts and McNally (1988) and more recently, in the context of IASI assimilation, from Bormann et al. (2010) and Stewart (2010). This thesis does not explore the general question of misspecified observation errors.

However, it is concerned with how to select IASI data from the full spectrum to maximise impact whilst maintaining a robust and stable solution. The two methods of increasing the information content explored in this thesis are principal component analysis and reconstructed radiances (chapter 6). In chapters 8 and 9, it becomes apparent that the correct definition of the $\mathbf{R}$ matrix for IASI is particularly important to exploit these assimilation techniques because they rely on interchannel correlations to provide enhanced information content.

3.6.2 Other assumptions

The equations of 3D- and 4D-Var also rely on other assumptions, the most important of which are summarised below.

- **The background and observations have Gaussian errors:** Although this assumption is usually stated, it is not essential for the variational equations (and in fact is not used in the assimilation of scatterometer data at the Met Office). If the background and observation pdfs are Gaussian, and with linear $\underline{M}$ and $H(\mathbf{x})$, the posterior pdf will also be Gaussian so the most likely solution will also be the minimum variance solution. With non-Gaussian errors and non-linear models, the most likely solution can still be found, but it may not be the best one.
- **The background and observation errors are uncorrelated:** Typically this assumption is valid, since the observations are made independently of the NWP model. However, this assumption may not hold if the observation is converted into model variables prior to assimilation with the help of the model itself. An example of this would be the assimilation of a one-dimensional retrieval from a satellite observation using the NWP model as a priori.
- **The background and observations are unbiased relative to the truth:** This assumption is generally not valid, and is the source of some not inconsiderable problems in data assimiltion. Background biases are certainly not zero – an example of diagnosis of upper tropospheric humidity bias in the Met Office and ECMWF models is given by Hilton et al. (2012b) – but creating a

large analysis increment can often cause numerical problems with the forecast model, which may reach an unstable state or simply not run. Observation biases may result from the instrument design, and from systematic errors in the forward model, such as errors in the spectroscopic parameters used in radiative transfer models. It is usually assumed that it is straightforward to remove such biases (see, for example Daley, 1991), but it is often far from easy! In practice, efforts are made, through regression modelling, to minimise the bias between the background and the observation, but this may result in model biases being corrected out of the observations. Furthermore, regional residual biases may remain even if the global bias is close to zero. Dee (2005) discusses handling of bias in data assimilation.

- **The evolution of the full model error pdfs is close to linear:** It is assumed that the cost function is close enough to linear locally that the optimal solution (the mean) is in fact the most likely (the mode; to be found by Var). The variational analysis system can deal with some non-linearity, as the algorithm iterates towards the optimal solution, but if the problem is too non-linear the system may struggle to reach convergence, and may find instead a local minimum. The most likely solution may also cease to be the best solution as the mean and mode diverge.

Practical application of the variational assimilation equations is often made considerably easier by some further assumptions, such as:

- **The observations have errors that are uncorrelated spatially:** This is achieved in practice by thinning the observation data so that there is typically no more than one observation per model grid box. In vertical coordinates, the assumption that all channels of a given satellite observation are uncorrelated is not valid, as there may be significant correlations arising from the instrument design, incorrect assumptions in the forward model which affect all or a subset of channels, or because multiple channels are sensitive to the same features which do not form part of the NWP model such as trace gas concentrations, or sub-grid scale features – the latter are known as errors of representation. Most NWP centres, including the Met Office, have moved, or are moving toward the use of error correlations in the vertical (e.g. Weston et al., 2014).
- **The background errors may be assumed to be vertically and horizontally separable:** Some formulations of background error covariance assume that correlations in the vertical coordinate are independent of correlations in the horizontal coordinate. This separability assumption is neither

valid, nor desirable. This assumption does not form part of the Met Office global covariance formulation.

- **The background errors are homogeneous and isotropic:** It is also usually assumed that errors in the horizontal domain are isotropic, and take no account of variations in meteorological conditions, such as the presence of weather fronts. This is addressed by ensemble-based and hybrid DA approaches, which are discussed in section 3.7.

There are also some aspects of the assimilation process that lead to problems for the use of indirect measurements such as satellite data. For instance, not all of the required variables to which satellite measurements are sensitive are part of the state vector, $\underline{\mathbf{x}}$. Important examples of this are skin temperature and surface emissivity, which affect the determination of lower atmospheric temperature and humidity from satellite radiances and are not known to a sufficient degree of accuracy from the NWP model itself or ancillary data. Also, at the present time, many NWP centres are using cloud-affected radiances in their assimilation systems, and required cloud properties may not be available to a sufficient degree of accuracy from the model, and may not even form part of $\underline{\mathbf{x}}$. At the Met Office, the practical measure used in these cases is to run a 1D-Var retrieval for each observation prior to assimilation, and to hold the retrieved variables (skin temperature, surface emissivity, cloud fraction, cloud top pressure, etc) as a fixed constraint within 4D-Var.

3.7 The use of ensembles in data assimilation

The most well-known of the family of ensemble DA methods is the EnKF, a Monte-carlo approximation to the Kalman Filter. The Kalman filter uses the same analysis equations as 4D-Var, but rather than using the cost function and a descent algorithm to approach the optimum solution, the Kalman Filter calculates the weights (equation 3.10) directly, and in order to generate $\mathbf{B}$ for the next cycle, the analysis error, $\mathbf{A}$ is propagated forward in time. The Kalman filter is a series of matrix multiplications, and therefore assumes a linear system. The Extended Kalman Filter does allow a non-linear model, but it is impossible to propagate $\mathbf{A}$ in this way for a global NWP model because of the size of the problem. If the outer loop of incremental 4D-Var converged, it would give a solution equivalent to that of the Extended Kalman Filter (although in practice this does not happen; in particular the Met Office does not run an outer loop). The Ensemble Kalman Filter removes the need for the explicit propagation of $\mathbf{A}$, by using an ensemble to estimate the uncertainty in the

forecast, and hence $\mathbf{B}$, at succesive timesteps.

Many centres already run ensemble-based forecast systems, such as the Ensemble Transform Kalman Filter (ETKF) known as MOGREPS that is run at the Met Office (Bowler et al., 2008). Ensembles of forecasts provide a measure of forecast uncertainty that is both useful to weather forecasters and to the data assimilation system, where it provides an estimate of the background error that reflects the atmospheric state of the particular forecast cycle.

Although the ETKF uses observations to control the ensemble spread, it is not optimised for data assimilation: for example, the localisation and filtering may need to be improved (Bowler et al., 2013). However, several techniques are possible to combine this flow-dependent background error into the data assimilation system so as to make the best possible use of observations. The main options for ensemble DA methods are summarised in this section. A comparison between 4D-Var and ensemble techniques is provided by Fairbairn et al. (2014).

The ensemble generates a measure of error that is valid for the atmospheric state at that particular cycle, rather than being a climatological average. The error term also differs from that used in 4D-Var in that, where 4D-Var explicitly evolves the 3D background pdf $\mathbf{B}$ using the PF model $\underline{\tilde{\mathbf{M}}}$, the ensemble 4D $\underline{\mathbf{B}}$ is generated by sampling the variance of the forecast ensemble at timesteps through the assimilation window. Note that a background error matrix generated by an ensemble is usually denoted $\mathbf{P}^b$, but here $\underline{\mathbf{B}}$ is used to emphasise that it is an alternative representation of background error, and 4D in nature.

The main challenge in using an ensemble to define background errors is that the ensembles are very small (only a few tens of members) and the covariance structure remains extremely under-determined. A technique called localisation is required to improve conditioning of the problem and to suppress spurious long-range correlations by restricting the influence of increments to a local area. Many implementations will also combine the flow-dependent background errors with the climatological errors used in traditional 4D-Var to provide stability.

3.7.1 4D-Ensemble Var (4DEnVar)

Of the ensemble techniques, 4DEnVar is probably the most popular technique at the present time as it is quite affordable in terms of computing cost. The United States National Centre for Environmental Prediction (NCEP) and Environment Canada are two centres that are moving over to this scheme at the time of writing.

4D-Ensemble Var is a variational solver like 4D-Var, but rather than solving

for a control vector comprising variables related to the atmospheric state, the cost function represents the fit of the ensemble members (generated by the ETKF, for example) to the observations, given the errors in each, where the forecast errors, $\underline{\mathbf{B}}$, are defined by the ensemble spread. The control vector consists of a series of weights, often referred to as alphas, α.

$\underline{\mathbf{B}}$ contains a time dimension and because the ensemble trajectories are already defined for the whole time window, there is no need for a model like $\underline{\tilde{\mathbf{M}}}$ to propagate the atmospheric state. This makes 4DEnVar more similar to a 3D-Var problem in terms of computational cost. $\underline{\mathbf{B}}$ here is the representation of the errors as that used in the Ensemble Kalman filter; the difference between the methods is that the variational algorithm uses a background state that comes from a previous deterministic analysis, whereas the ETKF retains and propagates all ensemble members independently. The localisation method is also different between the two techniques, and is theoretically more correct for 4DEnVar (Fairbairn, 2014).

4DEnVar generally outperforms 4D-Var with a climatological $\mathbf{B}$ in toy model experiments because the flow-dependent background error allows better use of the observations (Fairbairn et al., 2014). However, if the ensemble size is small, severe localisation is required, which degrades the time correlations in the background error covariance.

4DEnVar generates a single (deterministic) analysis, to which perturbations are then applied, for example by the ETKF system, to generate a new ensemble for the next forecast step.

3.7.2 Ensemble data assimilation (EDA)

4DEnVar generates a single analysis, but it is also possible to generate an ensemble of analyses. EDA techniques calculate an analysis for each ensemble member in turn, and then use, for example, the ensemble mean in the place of a deterministic forecast. ECMWF run an ensemble of 4D-Vars, i.e. a full 4D-Var for each ensemble member, which is very expensive and only a small ensemble can be run.

Another possibility is to run an ensemble of 4DEnVars. This DA method is self-sufficient, since each ensemble member is propagated forward, and thus the DA and ensemble systems are combined.

3.7.3 Hybrid DA

The term Hybrid DA refers to combining the flow-dependent $\underline{\mathbf{B}}$ generated by an ensemble of forecasts with a climatological $\mathbf{B}$. The advantage of this is that $\mathbf{B}$ is better specified than $\underline{\mathbf{B}}$, but $\underline{\mathbf{B}}$ can provide information about the current atmospheric state. Both Hybrid 4D-Var and Hybrid 4DEnVar are possible.

In Hybrid 4D-Var, the ensemble is used to specify errors of the day (EOTD) at the start of the assimilation window, and these are combined with the static climatological $\mathbf{B}$. $\underline{\tilde{\mathbf{M}}}$ is then used to propagate the hybrid covariance. This technique is currently operational at the Met Office (Clayton et al., 2012), with the covariances used in the ratio 80:50 static:EOTD, so that the total error is increased to counteract the problems with localisation caused by small ensemble size.

In Hybrid 4DEnVar, the climatological $\mathbf{B}$ is simply assumed to be constant throughout the assmiliation window and is added to $\underline{\mathbf{B}}$. It is found that the propagation of $\mathbf{B}$ by the PF model means that Hybrid 4D-Var generally outperforms Hybrid 4DEnVar (Fairbairn, 2014) and is likely to do so until a large enough ensemble can be run to reduce dependence on $\mathbf{B}$.

3.8 A 1D-Var approximation to 4D-Var

1D-Var is often used in independent retrieval schemes, which seek to determine a separate atmospheric profile from each observation in turn. These retrievals can be used in nowcasting applications or as a priori in further retrieval schemes, for example for trace gas concentrations. The results presented in this thesis involve the use of a 1D-Var analogue to 4D-Var. New ways of using observations can be coded quickly into a 1D-Var framework to allow testing of various options. Furthermore, the Met Office operational DA scheme uses a 1D-Var preprocessor for all radiance observations prior to assimilation in 4D-Var, primarily as a method of quality control and to derive estimates of certain quantities which are not solved for by the 4D-Var scheme, such as skin temperature and emissivity. The code used in this thesis is that of the NWPSAF[2]-Met Office 1D-Var (Weston et al., 2013) version 3.4 from 2010.

The 1D-Var system has an obvious difference from the full DA scheme, in that the information is spread in the vertical but not the horizontal. Nadir satellite sounding is fundamentally a vertical problem, and the work presented here is concerned mainly with exploiting correlations between channels in the same field of view and in the

[2]NWP Satellite Applications Facility. A EUMETSAT-funded project to promote the use of satellite data in NWP

implied vertical correlations in the temperature and water vapour profiles, so the 1D-Var analogue is a useful tool for this work.

One difference between retrieval schemes and DA is that the former often have a single climatological state as the background, or at best use regression techniques to choose from amongst a set of climatological profiles. The relative skill of the prior information is a major difference between NWP and most independent retrieval schemes using satellite data, and is demonstrated in Hilton et al. (2009a). It is therefore important when using a 1D-Var retrieval as an analogue of 4D-Var to make the systems as similar as possible. This is achieved here by:

- Using atmospheric columns taken from the Met Office Unified Model (UM) operational global forecasts as background states for the 1D-Var.
- Performing the 1D-Var on height-based model levels rather than the standard 43 levels used in the operational 1D-Var preprocessor. The UM model levels are described and plotted in appendix A.
- Using a technique called covariance sampling to match the 1D $\mathbf{B}$ matrix to the static $\mathbf{B}$ used in 4D-Var. This is described in section 3.8.1. Although it is not easy to recode the 1D-Var to use the same uncorrelated control vector variables as 4D-Var, the covariance sampling method captures the behaviour in state space that is implied by the $\mathbf{U}$ transform (section 3.5.3).

3.8.1 Covariance sampling

For 1D-Var experiments, an approximation to the 4D-Var error covariance matrix is required. This is achieved by a process called covariance sampling, similar to that described in Andersson et al. (2000). A set of locations spread over a wide geographic region is used to sample the covariance structure. At each location, a vector the size of the control vector for that location is generated whose elements are determined by a Gaussian distribution with zero mean and unit variance. These vectors are then transformed using the full T-transform, which maps the vectors from control vector space to PF model variables. The average of the outer products of these vectors tends to a 1-dimensional representation of the transforms, and hence the implied covariance structure of the PF model variables. The 1D matrix can be further transformed from θ and q_t to T and $\ln(q)$ which are the standard variables used in the Met Office 1D-Var preprocessor.

The T-transform is slightly dependent on the linearisation state (for example, through the inner product matrix), so a different cycle would not give an identical representation of $\mathbf{B}$. However, the variations are slight when aggregating many

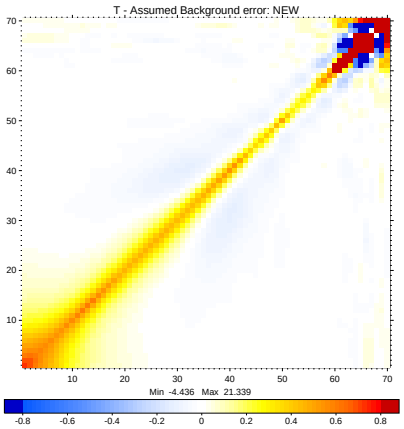
samples and not much seasonal dependency of the 1D covariance statistics has been seen.

For this thesis, covariance sampling was used to generate 1D **B** matrices using 100 000 pseudo-observations from 500 locations across the northern hemisphere from two sets of climatological 4D-Var background error statistics: those that were operational in September 2010 and those that were operational in March 2012. These will be referred to respectively as **NMC** and **MoistCov**. The **NMC** covariances were, as the name implies, derived using the NMC method, and the water vapour is represented as a total humidity, combining cloud liquid water, cloud ice water and water vapour. The **MoistCov** covariances were derived using an ensemble of model states (Piccolo, 2011), with water vapour represented via a new non-linear total humidity control variable, described in Ingleby et al. (2013), designed to take better account of the amount of water vapour present at each vertical level in the atmosphere. The **NMC** and **MoistCov** matrices for temperature and the natural logarithm of total water mass mixing ratio, $\ln(q_t)$, are shown in figure 3.3. The **NMC** matrices have slightly longer correlation length scales and higher variances. In both cases, there are no cross-correlation terms between T and q_t calculated. When these matrices are used, both here and in operational applications, it is assumed that the covariance structure for water vapour mass mixing ratio, q, is equal to that of q_t.

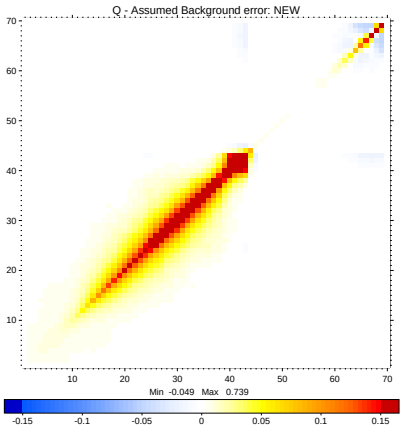
These are just two possible representations of the sampled forecast errors, and the differences between them may provide some indication of the likely level of error in the estimation of the **B** matrix used for operational assimilation. For technical reasons, it did not prove possible to sample the errors of the day relative to the static climatological covariances.

In order to understand the scales on which information is being extracted from IASI observations, it is first necessary to examine the scales to which the analysis is sensitive. One way to investigate the vertical scales is to examine the eigenvectors of the matrices. The eigenvectors have no true physical meaning, as they are sensitive to purely mathematical aspects of the representation, such as whether the humidity varibale is relative humidity or a mass mixing ratio. However, they are broadly consistent with the concept of the vertical transform applied within the Met Office 4D-Var. The orthogonality of the eigenvectors also means that one can split the system into mutually independent components for which a scalar analysis may be considered more directly applicable. An eigenvector decomposition is used to explore the vertical scales of the optimal linear analysis in following chapters.

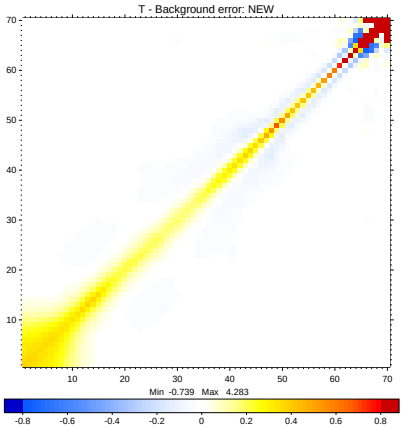
The first 20 eigenvectors for temperature and humidity ranked by eigenvalue are



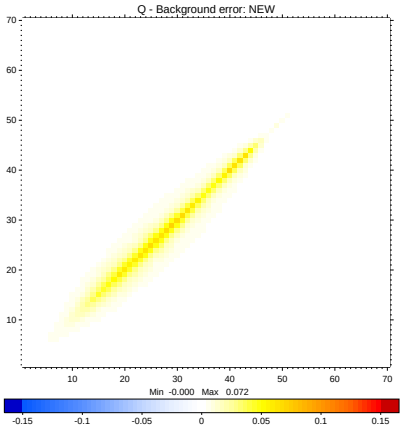
(a) NMC Temperature



(b) NMC ln(q_t)



(c) MoistCov Temperature



(d) MoistCov ln(q_t)

Figure 3.3: The background error covariance structures used in this thesis.

shown in Appendix B for the **MoistCov** matrix, and in Appendix C for the **NMC** matrix. The first 20 humidity vectors for **MoistCov** form a well-behaved family of functions, with increasing vertical structure – this is not always the case, and shows that this matrix increasingly filters out structures in the analysis increment with shorter and shorter vertical length-scales (this is equivalent to saying that they are well-known by the model). For the **NMC** matrix, this is also broadly the case, but this family is punctuated by vectors with large stratospheric components. These features are spurious artifacts, and do not apply in the 4D-Var analysis as the increments for humidity are set to zero above the tropopause. The fact that these artifacts are not present for the **MoistCov** matrix is a result of the new humidity control variable. For temperature, we do not see increasing structure with decreasing eigenvalue for either matrix, and in both cases the first few vectors seem to consist mostly of stratospheric or surface structures.

3.8.2 Minimisation

The Met Office global 4D-Var algorithm uses a conjugate gradient descent method, as described in section 5 of Fisher (1997). The NWPSAF-Met Office 1D-Var uses Newtonian or Gauss-Newton minimisation, which can be equivalently formulated in two ways, the first of which (equation 3.36) is more efficient if the observation vector is longer than the state vector (for example, a 1D analysis from a hyperspectral sounder such as IASI), and the second (equation 3.37) is more efficient if the state vector is longer (for example, a 1D analysis from a microwave sounder such as AMSU). The state for iteration $i+1$ is calculated as follows:

$$\mathbf{x}_{i+1} = \mathbf{x}^b + (\mathbf{B}^{-1} + \mathbf{H}_i^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}_i)^{-1}\mathbf{H}_i^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}^o - H(\mathbf{x}_i)) + \mathbf{H}_i(\mathbf{x}_i - \mathbf{x}^b)) \quad (3.36)$$

$$= \mathbf{x}^b + \mathbf{B}\mathbf{H}_i^{\mathrm{T}}(\mathbf{R} + \mathbf{H}_i\mathbf{B}\mathbf{H}_i^{\mathrm{T}})^{-1}(\mathbf{y}^o - H(\mathbf{x}_i)) + \mathbf{H}_i(\mathbf{x}_i - \mathbf{x}^b)) \quad (3.37)$$

Newtonian iteration is a quadratic solver and, in cases where the cost function is poorly represented by a quadratic surface, may struggle to find the optimal solution. Other iteration algorithms exist which can be useful for more non-linear problems: one such solver, also implemented in the NWPSAF-Met Office 1D-Var, is the Levenberg-Marquardt algorithm (Levenberg, 1944), which incorporates an additional factor γ into the first term in round brackets in equation 3.36:

$$\mathbf{x}_{i+1} = \mathbf{x}^b + ((1+\gamma)\mathbf{B}^{-1} + \mathbf{H}_i^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}_i)^{-1}\mathbf{H}_i^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}^o - H(\mathbf{x}_i)) + \mathbf{H}_i(\mathbf{x}_i - \mathbf{x}^b)) \quad (3.38)$$

The Levenberg-Marquardt technique is often used in 1D-Var solvers where cloud variables are analysed. Cloud analysis is inherently non-linear because the presence of cloud in the field of view has a very strong effect on the weighting function peaks of satellite observations, and because cloud liquid and ice do not vary smoothly across all values of humidity mass mixing ratio, but instead form suddenly once saturation is reached.

3.8.3 Experimental set-up for this thesis

Unless otherwise stated, in this thesis all optimal linear analysis calculations and 1D-Var experiments use the following settings:

- A control vector consisting of 70 fixed-height levels of temperature and 70 levels of $\ln(q)$ from the surface to 80km is used. See below for a note on the treatment of surface variables
- The **MoistCov** matrix is used as $\mathbf{B}$, to define the relationship between the elements of the control vector.
- The $\mathbf{R}$ matrix consists of diagonal instrument noise provided by EUMETSAT (Hultberg, 2009), being the same noise profile used in normalisation for the disseminated IASI PC Score Level 1c product.
- RTTOV-10 (Saunders et al., 2010) is used for all radiative transfer modelling, to calculate Jacobians and simulated observations. Coefficients are Version 9 LBLRTM on 101 levels. Interpolation from 70 levels is performed inside RTTOV. Sea surface emissivity is calculated by RTTOV.

In addition, the NWPSAF 1D-Var is used with the following experimental set-up:

- A set of 4348 profiles over sea were used, covering all latitudes and taken from a regular spacing of model grid-points were extracted from the Met Office global model from a forecast cycle on 15th March 2011. These profiles were used as truth. No cloud liquid or ice water was included - the analyses are all clear-sky only.
- Simulated background profiles were calculated from the true profiles by adding a random perturbation of each eigenvector of the $\mathbf{B}$ matrix taken from a normal distribution with zero mean and a standard deviation equal to the squareroot of the eigenvalue.
- Simulated true clear-sky observations were calculated from the true profiles using RTTOV-10.

- Simulated noisy observations were calculated from the true observations by adding a normally distributed random perturbation consistent with the instrument noise (or other error covariance as specified in the individual experiments).
- Any PC compression and radiance reconstruction was done after the addition of noise.
- The 1D-Var used Newtonian minimisation (equation 3.37) and was allowed up to ten iterations to reach convergence, which is determined by the cost function changing by no more than 1 % on a given iteration. If the 1D-Var is still not converged after ten iterations, the retrieved profile is rejected from subsequent analysis.

For optimal linear analysis calculations, single profile results are for the US Standard Atmosphere. Results presented for multiple profiles include seven additional atmospheric profiles from the 4348 Met Office model profile set above, distributed across different latitudes.

Surface variable analysis

Early results were produced for two control vectors: one with the 70 level profiles of temperature and $ln(q)$ only; and one including skin temperature, surface air temperature and humidity, and surface pressure. The former is more representative of the 4D-Var control vector, and the latter is the same as the control vector used in the operational 1D-Var preprocessor. In the case of the expanded control vector, the error variances of the extra variables, and the correlations between the surface variables and the atmospheric profiles were the same for both **MoistCov** and **NMC** matrices.

It was found that the inclusion of the surface variables had little effect on the performance of the analysis for the rest of the profile, as long as the errors in the quantities were correctly defined. In other words, the analysis for the temperature and water vapour profiles was very similar in either of the following two situations:

- The surface variables in the background profile were given errors consistent with the extended **B** matrix, and the control vector contained surface variables.
- The surface variables in the background profile were the same as in the true profile, and the control vector omitted the surface variables.

Where background error was added to the surface variables but they were not analysed, the results for the rest of the atmospheric profile were found to be substantially in error. In section 10.3 the 1D-Var experiment did analyse surface variables, but

in the rest of the thesis they were omitted from the control vector because this is more consistent with the 4D-Var system that the 1D-Var experiments are designed to represent.

Chapter 4

Measuring the information provided to the assimilation system by IASI

The aim of this thesis is to increase the amount of information extracted from the IASI observations: to do this it is necessary to understand the impact the observations have within the current assimilation system, and to define measures by which changes to the observation processing can be assessed to see whether they improve access to the information contained in the spectra.

When IASI was launched, pre-operational trials showed that it gives very good impact on the NWP forecasts as a whole (Hilton et al., 2009a; Collard and McNally, 2009). These trials measure the impact over broad forecast measures, and at the Met Office the quantity used is the "NWP index" which is a weighted average of skill scores measured against persistence for variables that are of interest to its main customers. The trials showed that IASI gave impact at least as great as that of any other new observing system (Hilton et al., 2012a). However, the broad-brush forecast impact scores do not provide detailed information and cannot tell us whether there are, for example, parts of the spectrum that are not being fully exploited.

The next section introduces a method, Forecast Sensitivity to Observations (FSO) that has become popular in the last few years to examine the impact of observations within the variational analysis system. FSO is then used in section 4.1.1 to examine the impact of individual IASI channels within the Met Office 4D-Var.

Section 4.2 then introduces some traditional measures of information content that can be used within the context of a 1D-Var system to compare different schemes and predict the likely impact of changes to IASI usage on the full 4D-Var system.

4.1 Forecast sensitivity to observations

The FSO measure is designed to demonstrate the impact of individual observations within the Met Office 4D-Var assimilation system on global forecast errors out to 24 hours. It uses an integrated quantity, known as an energy norm, which combines information on potential temperature, pressure, wind and humidity errors, to meas-

ure the impact of each observation. Based on a system developed by Baker and Daley (2000), the Met Office FSO system is described in full by Lorenc and Marriott (2013). FSO calculates the change in energy norm resulting from the forecast difference at 24 hours from a prior state with and without assimilated observations (i.e. the analysis and background respectively). This is propagated back to analysis time, resulting in the sensitivity of the forecast error to the analysis increment from the combined set of observations. The adjoint of the Perturbation Forecast model is then used to partition this change in energy norm between observations, based on the weights given to, and the magnitude of each observation increment. As such, the FSO is integral to 4D-Var, and relies on the use of the linear PF model – thus it does not represent the true impact of the observations on the forecast produced by the non-linear model. It also assesses the impact of an observing system on one individual forecast, rather than the cumulative effect of assimilating observations over successive forecast cycles (Gelaro and Zhu, 2009). However, it is a useful tool to examine in some detail the performance of different observations relative to each other.

FSO is powerful because the individual impacts can be aggregated in numerous different ways, such that the impact of total observing system (e.g. ATOVS, radiosondes), or groups of observing systems (e.g. satellite sounders) can be measured, along with the impact of satellite platforms (e.g. Metop-A), separate satellite instruments, or even observations over different surface types. Joo et al. (2013) used the FSO system to examine the impact of satellite observations within the Met Office Global 4D-Var system, and showed that satellite observations accounted for 64 % of the total impact of all observations; that Metop-A had the largest impact of any satellite platform; and that IASI provided 49 % of the impact of Metop-A. FSO can also be used to separate the impact from each individual channel, and it is this aspect that is of interest here.

New FSO results are shown in section 4.1.1 that present the impact of each IASI channel over a range of energy norms, to diagnose how the pattern of impact relates to the principal sensitivity of each channel. The standard energy norm used at the Met Office includes humidity fields (a moist energy norm) and combines model levels from the surface up to 150 hPa. Impact can also be examined using a dry energy norm, where the impact on humidity fields is removed from the total. Both energy norms are examined here. In addition to the standard vertical range, three additional ranges have been added:

- Lower Troposphere: from 1000–4000 m, comprising model levels 12–23

- Mid Troposphere: from 4000–8000 m, comprising model levels 24–33
- Upper Troposphere/Lower Stratosphere (UTLS): from 8000–20 000 m, comprising model levels 34–52

4.1.1 IASI FSO by channel

The main interest in the FSO metric within this thesis is to examine the relative impacts of the IASI channels, and to see how the pattern of impact changes with height in the atmosphere. In this section, the impact of each of the assimilated channels of IASI (the *VAR* channel selection, see appendix E) on Metop-A is measured over the period January 31st 2012 to March 18th 2012. The channel numbering is based on the position of each channel within the 314 channels defined by Collard (2007, see appendix E), referred to in this thesis as the *Collard* set. The results presented are the mean impact per sounding, which is the change in the energy norm on assimilation of a given IASI channel over the full set of observations used in 4D-Var. A negative value is a reduction in the energy norm, and means that the observations are beneficial.

Figures 4.1(a) and 4.1(b)) present the results over the standard pressure range for moist and dry norms respectively, figures 4.2(a) and 4.2(b) for the lower troposphere, figures 4.3(a) and 4.3(b) for the mid-troposphere, and figures 4.4(a) and 4.4(b) for the upper-troposphere/lower-stratosphere (UTLS). All of the plots have the same range on the x-axis. The plots are arranged such that the channels are in order of weighting function peak, with those high in the atmosphere first. Channels labelled in blue are temperature sounding channels, those in green are window channels, and those in red are water vapour channels. The first column in the figures therefore shows temperature sounding channels, starting at the top in the stratosphere and moving down in peak height. The middle column shows mid tropospheric channels at the top, starting with those that peak around 400 hPa, and moving down the atmosphere to the window channels. The third column has more window channels at the top, and then water vapour channels, starting with the highest peaking and moving down through the atmosphere. The black lines at the end of the bars indicate the standard error. This is small, because the sample size is very large. However, it takes no account of any systematic bias in the calculation of the energy norm, or any errors correlated between observations, so should be considered an underestimate (Richard Marriott, pers. comm.).

Comparing figure 4.1(a) and (b), the impact of the higher-peaking temperature sounding channels is unaffected by the choice of energy norm, but the window chan-

nels and water vapour channels have much more impact when measured by a moist norm. This demonstrates that the water vapour channels' impact is mostly on the humidity fields; this is the expected result but was not a given as the assimilation of AIRS water vapour channels did little to improve humidity forecasts at ECMWF (Collard and McNally, 2009). The window channels also show a much greater impact on the moist energy norm. Skin and surface air temperature do not form part of the 4D-Var analysis, and the impact of these channels is measured via their contribution to the analysed column as a whole: most of these channels are sensitive to the water vapour continuum, and some are near the frequencies of weak water vapour lines. They are thus sensitive to water vapour in the lower atmosphere, and so have greater impact on a moist energy norm. However, the relative impact scores of observations sensitive to lower atmospheric water vapour is believed to be less reliable than those of observations sensitive higher in the atmosphere, because the PF model is used to measure the impact, and the behaviour of water vapour in the boundary layer is not well captured in the linear model (Lorenc and Marriott, 2013).

This pattern of impact is even clearer in the plots for the lower troposphere (figures 4.2(a) and 4.2(b)). Comparing these plots with the other height ranges, it seems that most of the impact of IASI is on the lower troposphere, and in particular the window and water vapour channels seem to have a large effect on the moist energy norm. A few channels even show a tiny negative impact when measured on the dry energy norm.

In the mid troposphere (figures 4.3(a) and 4.3(b)) and UTLS (figures 4.4(a) and 4.4(b)), the impact of the window channels is much smaller. The water vapour channels again show more impact with the moist energy norm, and the higher peaking channels have more impact in the mid troposphere than the lower troposphere. They have very little impact in the UTLS, as expected given the low concentrations of atmospheric water vapour. The exception are channels 215, 221, 251, 259, 261, and 263, (corresponding to instrument channel numbers 3168, 3248, 3506, 3577, 3582, 3589) which have more impact in the UTLS than lower in the atmosphere. These are curious channels with strangely shaped Jacobians without well-defined peaks, and with very long tails into the stratosphere. These channels have just as much impact when measured by the dry energy norm.

For temperature sounding channels, the choice of norm makes little difference to the impact of the higher peaking channels (in the first column of the plots) but the lower tropospheric channels show more impact with a moist norm. The pattern of impact is as expected: moving higher through the atmosphere, the impact of the

lower-peaking channels reduces and that of the higher-peaking channels increases.

The results presented here indicate that IASI observations are mostly having the expected pattern of impact based on the main sensitivity of the channels: the water vapour channels have an impact on the moisture fields, the temperature sounding channels have an impact on the dry fields, and moving up through the atmosphere the impact of the lower-peaking channels decreases and that of the higher-peaking channels increases. This suggests that there is no major deficiency in the IASI assimilation that could be addressed leading to quick gains in the observation impact. Instead, this thesis will focus on the addition of extra information by increasing the spectral coverage presented to the assimilation system.

4.2 Traditional information content measures

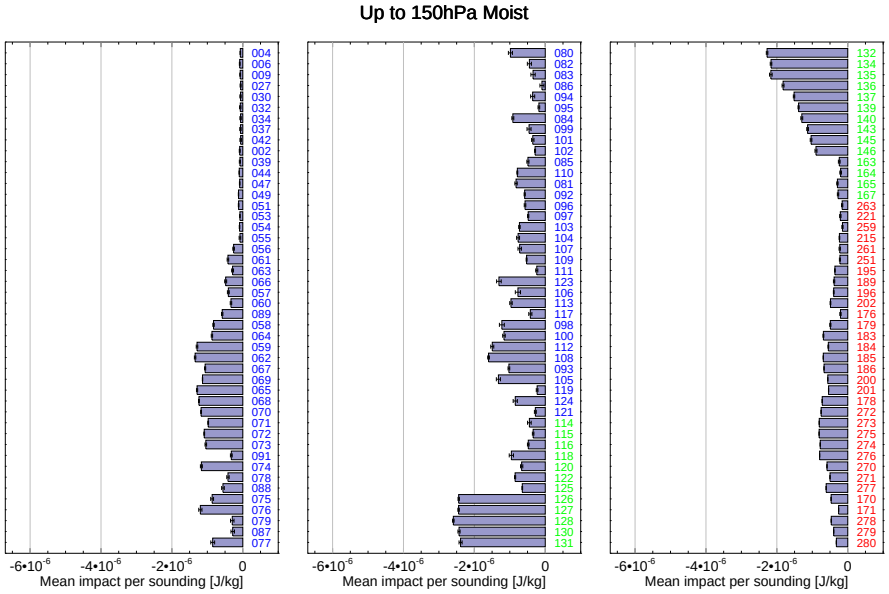
In this thesis, different aspects of the assimilation system are explored using a 1D-Var framework. Before running relatively expensive full 1D-Var experiments, it is useful to predict the performance of a fully linear system. In order to assess the relative performance of these systems, and the amount of information provided to the analysis by the IASI spectra, it is necessary to define some quantitative measure by which they can be compared.

Rodgers (2000) provides a chapter on error characterisation for the optimal estimation problem. It is customary to assess different retrieval schemes against each other using information content measures, generally by comparing the retrieval errors with the background errors. In operational DA, the term analysis is used instead of retrieval, and will also be adopted in this thesis. The quantity that is most generally used is DFS (degrees of freedom for signal). This can be calculated in various ways, but one of its simple forms is:

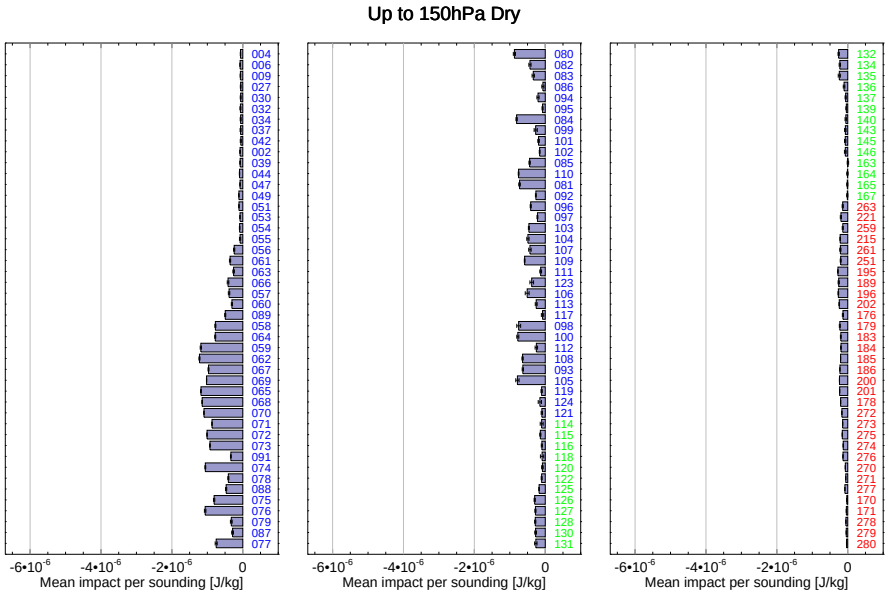
$$\mathrm{DFS} = \mathrm{Tr}(\mathbf{I} - \mathbf{A}\mathbf{B}^{-1}) \tag{4.1}$$

Another common measure is the Shannon information content (SIC), a measure of the entropy of the system. In this type of application, it is usual to calculate the SIC difference between the analysis and background states. Again, this can be calculated via several different equations, but one of the simplest is:

$$\mathrm{SIC} = -\frac{1}{2}\ln|\mathbf{A}\mathbf{B}^{-1}| \tag{4.2}$$

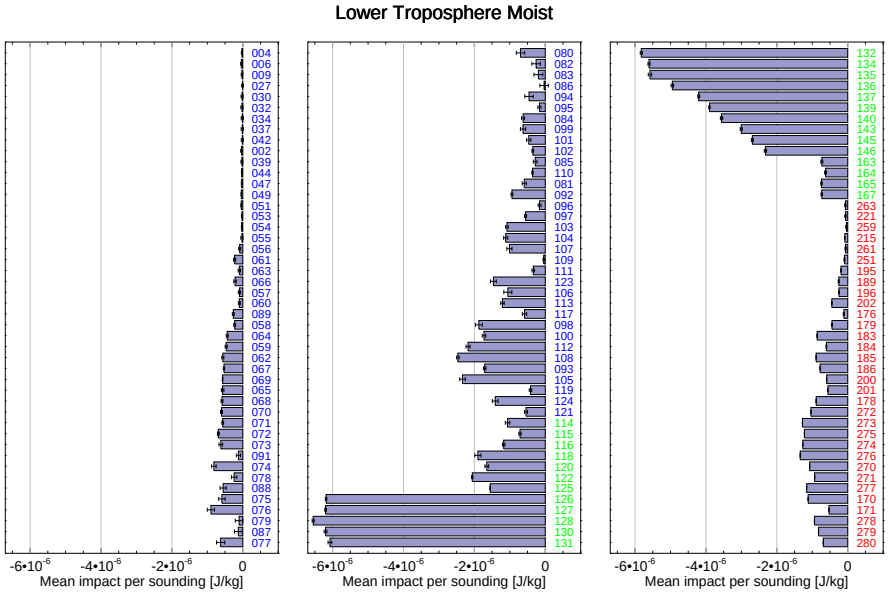


(a) Moist Energy Norm

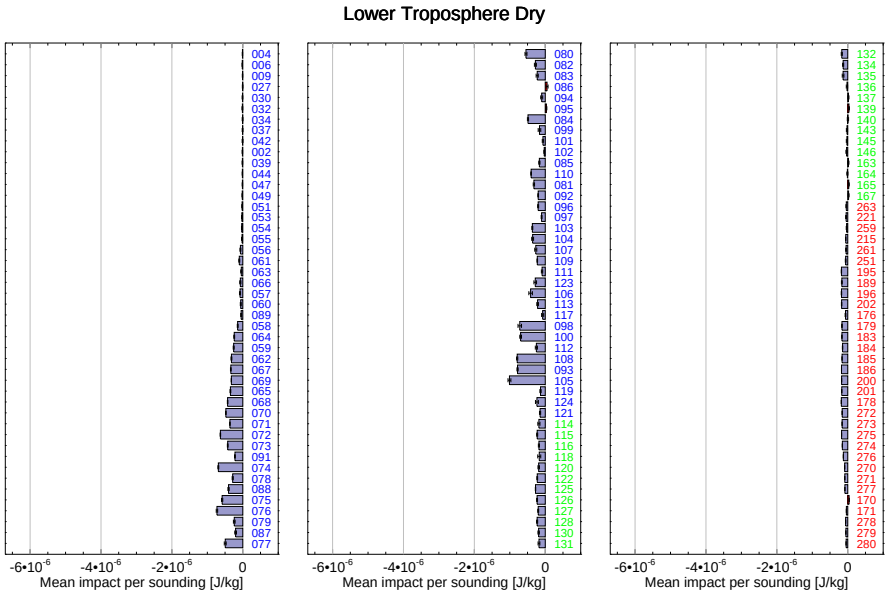


(b) Dry Energy Norm

Figure 4.1: Impact of IASI channels from the *VAR* selection, surface to $150\,\mathrm{hPa}$. Channels are in order of weighting function peak, with those high in the atmosphere first. Channels labelled in blue are temperature sounding channels, those in green are window channels, and those in red are water vapour channels. The black lines indicate the standard error.

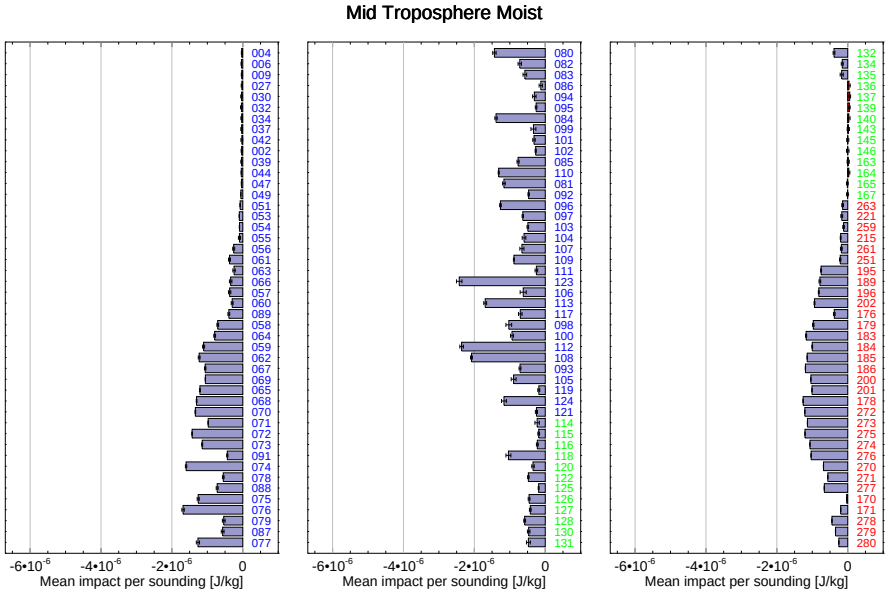


(a) Moist Energy Norm

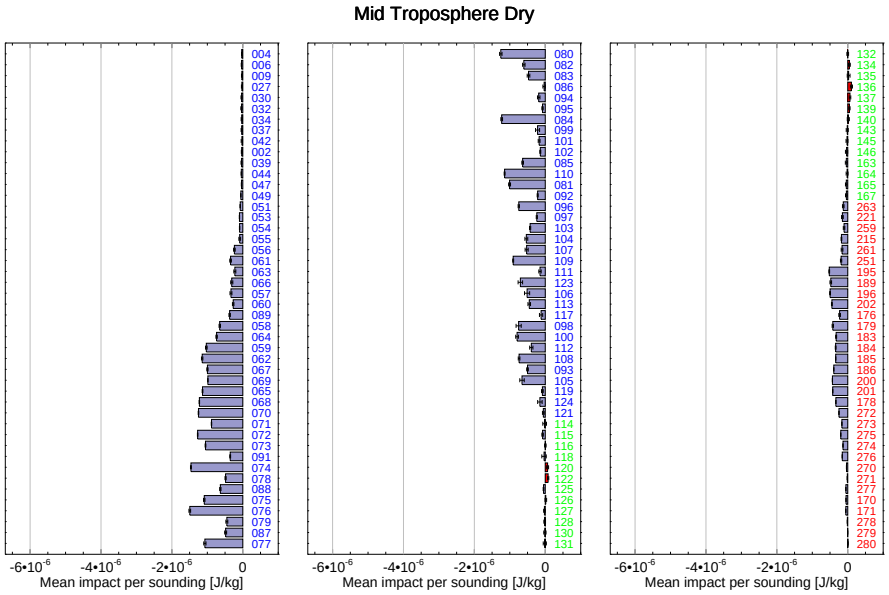


(b) Dry Energy Norm

Figure 4.2: Impact of IASI channels from the *VAR* selection, lower troposphere. Channels are in order of weighting function peak, with those high in the atmosphere first. Channels labelled in blue are temperature sounding channels, those in green are window channels, and those in red are water vapour channels. The black lines indicate the standard error.

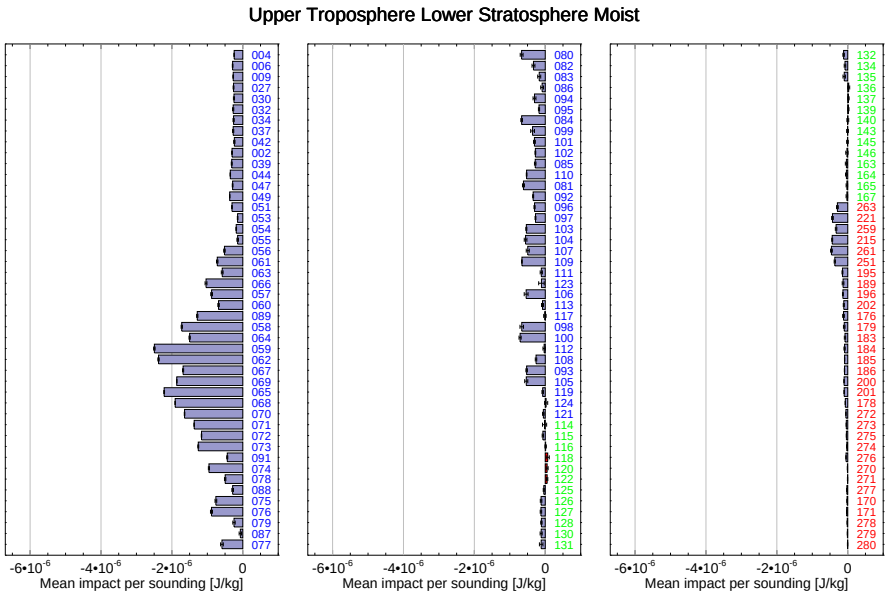


(a) Moist Energy Norm

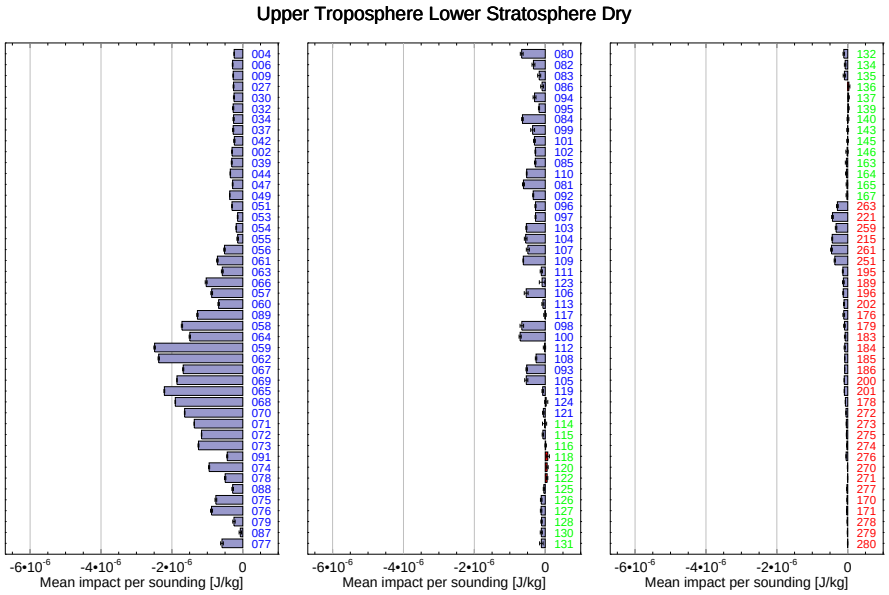


(b) Dry Energy Norm

Figure 4.3: Impact of IASI channels from the *VAR* selection, mid troposphere. Channels are in order of weighting function peak, with those high in the atmosphere first. Channels labelled in blue are temperature sounding channels, those in green are window channels, and those in red are water vapour channels. The black lines indicate the standard error.



(a) Moist Energy Norm



(b) Dry Energy Norm

Figure 4.4: Impact of IASI channels from the *VAR* selection, upper troposphere/lower stratosphere. Channels are in order of weighting function peak, with those high in the atmosphere first. Channels labelled in blue are temperature sounding channels, those in green are window channels, and those in red are water vapour channels. The black lines indicate the standard error.

It is customary to rescale SIC to Base 2, as in information theory this then gives the number of pieces of information in bits. This rescaling has not been done here as it does not affect the comparison of one system relative to another – it is purely a division by $\ln(2)$. Another quantity used by Rodgers (2000) is the number of measurements of the analysed components made to better than observation error. This quantity is estimated from a singular value decomposition (SVD) of the normalised forward model Jacobian matrix, $\tilde{\mathbf{H}}$:

$$\tilde{\mathbf{H}} = \mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{B}^{\frac{1}{2}} \tag{4.3}$$

Singular vectors with magnitudes greater than about 1 represent measurements in analysed space made to better than observation error.

These techniques all assess the number of independent pieces of information contained within the analysis. They can never represent all aspects of the behaviour of even an optimal linear retrieval, as they condense a multi-dimensional process down into just one number. It can be useful, therefore, for a column analysis, to examine also the averaging kernels, a set of vectors given by $\mathbf{KH}$, which represent the response of each level of the analysis to the observation.

Furthermore, it is important to note that it is not uncommon to see these information content measures presented as fundamental properties of the retrieval as a whole, whereas they really tell us about the amount by which the observations have improved the information content over the background. Thus it can be seen that DFS is often quite low for a retrieval starting from a forecast from a state-of-the-art NWP model, whereas a retrieval using the same observation but with a climatological background profile would seem to contain more information. This is problematic when comparing completely different schemes, but when comparing different settings within the same assimilation scheme, we are indeed interested in how much information the observation is providing to the analysis, and not how much information the analysis holds overall.

Further analysis of averaging kernels can be similarly misleading: Rodgers (2000) states that where the sum under the curve of the averaging kernel is close to 1, an "accurate retrieval" is performed for those atmospheric levels. However, during the detailed case study analysis of several retrieval schemes performed by Hilton et al. (2009b), it was seen that this was usually the case for smooth retrievals from climatological backgrounds, and almost never the case for retrievals based on good quality forecast backgrounds. This quantity is also really examining how much information comes from the observation rather than the background, although the shape and

spread of the averaging kernels is to a large extent governed by **B**. The width of averaging kernels is another quantity that may be examined, but this is regarded as somewhat unhelpful for the analysis system under examination: many of the IASI channels have broad weighting functions and each averaging kernel vector rarely has a single peak, often crossing the zero line. Whilst well-defined peaks would be a desirable property of a perfect analysis, it must be recognised that these long tails are a feature of the system, and so instead the aim is for evenness of spacing of information in the vertical.

Other measures of information content exist, such as the relative entropy of Xu et al. (2009). They state that although an observation that agrees with the background exactly still provides information (it affects the spread of the system by reducing uncertainty), it does not tell us something new. The portion of an observation we are really interested in is that which cannot be predicted by the background. They discuss the measurement of surprise in an analysis, and show how this can be measured using relative entropy. In fact, this idea is also reflected in FSO, because the impact of each observation is dependent on both the weight given to it, and on the observation increment: observations that agree perfectly with the background and thus have an increment of zero have no impact by the FSO measure (Gelaro and Zhu, 2009). However, relative entropy is a combination of SIC and DFS, and so is closely related to the more traditional measures and will not be considered further.

The three information content measures proposed by Rodgers (2000), i.e. DFS, SIC and number of significant measurements, are examined in the next section for an optimal linear 1D system representative of the Met Office analysis. These results provide a measure of the information content currently extracted from IASI, and the expected gains possible with more spectral coverage. They provide a starting point against which the experiments in subsequent chapters can be compared.

4.3 Optimal linear analysis for the 1D assimilation of IASI

Prunet et al. (1998) and Collard (1998) both estimated the information content of an optimal linear IASI retrieval on 40 vertical levels using the full spectrum assuming no forward model error, and with a 1D **B** matrix from the ECMWF model at the time, to be about 20 (note that Collard uses DFS while Prunet uses a significant eigenvalue method to assess information content).

The results presented here assess the information content using the 1D **B** matrices, **NMC** and **MoistCov**, presented in section 3.8.1, which are representative of the current Met Office 4D-Var formulation. The analysis is performed on the 70 unified model levels, using RTTOV-10 (Saunders et al., 2010) as the forward model, and assuming diagonal instrument noise only, as provided by EUMETSAT (Hultberg, 2009). Different channel sets are investigated, all of which are detailed in appendix E (by convention throughout this thesis, italics will be used to identify a channel set):

A The full IASI spectrum

B *Collard*: 314 channels from Collard (2007)

C *OPS*: 183 channel subset of the 314 that are used in the Met Office 1D-Var preprocessor

D *VAR*: 138 channel subset of the 314 that are assimilated in the Met Office 4D-Var

E *Band1*: 1997 channels that represent the whole of IASI Band 1

F *PCs*: 290 principal component scores consistent with the EUMETSAT PC score Level 1c IASI product that represent the whole of the spectrum

PC compression is described in chapter 6 and was mentioned briefly in chapter 2. This is the current method of choice for dissemination of compressed IASI spectra and is planned for MTG-IRS in the future. It is the premise of this thesis that PC compression is a useful tool to recreate the information content of the full spectrum using a compressed representation that is easier to handle in the assimilation system because the **R** and **H** matrices are of a manageable size.

Figure 4.5 shows the values of the information content measures described above for this set of optimal linear analyses. Column A shows the information content for the use of the full spectrum, and the other columns are for the subsets of the spectrum as defined above and in the figure caption. The three information content measures are numerically different, as expected, but all show the same pattern of variation with **B** matrix and channels used, and so there is little to choose between them in terms of a definitive measure to apply to test systems.

For each of the channel sets, slightly more information is extracted from the observations using the **NMC** matrix, because of its higher variances. Increasing the number of channels assimilated in general increases the information content (though see below on the use of Band 1). These plots demonstrate four important points for the aims of this thesis:

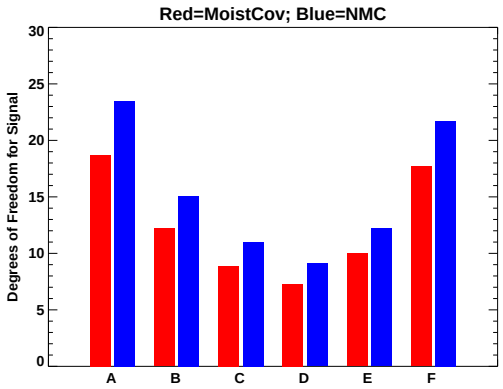
1. Column A: The full spectrum still gives a DFS of approximately 20, as pre-

dicted by Collard (1998) and Prunet et al. (1998). This is surprising, since the finer vertical structure of the current analysis system might be expected to allow the extraction of a greater level of information from the data. The model errors have also reduced over the last fifteen years.

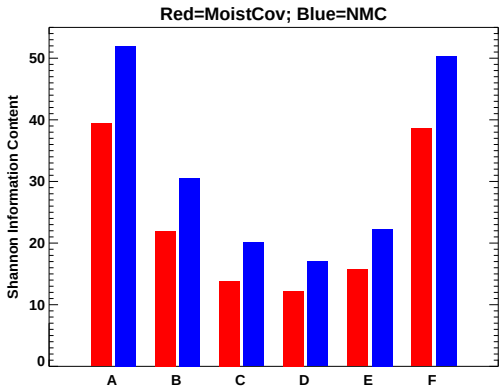
2. Column D: The *VAR* channel selection that is currently assimilated has an information content of less than 50% of that of the full spectrum, whichever measure is chosen (DFS ~8). There is significant opportunity for increasing the information content of the IASI assimilation system.
3. Column E: *Band1* has a lower information content than the *Collard* selection (Column B), despite containing more channels. This is because *Band1* only covers the longwave CO_2 band and therefore has little water vapour information, whereas the *Collard* set was chosen to maximise information content for NWP applications. Comparing *Band1* with the full spectrum (Column A), we can estimate that approximately 50% of the information content comes from the longwave CO_2 band. This shows that there is significant potential for the extraction of information from the water vapour band, which is only lightly used in the *OPS* and *VAR* channel selections.
4. Column F: The 290 PCs from the full spectrum contain the vast majority of the information content of the full spectrum, indicating that the use of PCs might be a way to condense the full spectral information into a more manageable quantity of measurements. This technique, and the related radiance reconstruction, is explored in this thesis in chapters 6 to 9.

Chapter 10 will discuss the fact that the analysis is in fact not optimal, because the **B** matrix is misspecified. The linear equations are expanded to calculate a suboptimal analysis error. Since the 'number of measurements' metric does not use the analysis error in its calculation, it is not suitable for assessing the performance of a suboptimal assimilation system. For this reason, it will be disregarded from here on since it gives similar results to the other measures for the optimal system. There is little to choose between DFS and SIC - they both show similar effects of increasing the spectral coverage. DFS will be used in following chapters of this thesis because of its heritage in channel selection methods, such as that of Collard (2007).

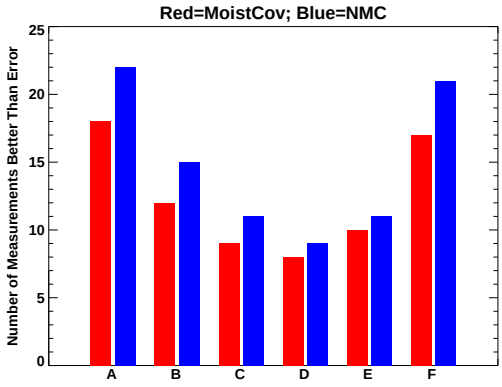
Figure 4.6 displays the rows of the averaging kernel matrices for the full spectrum and the *VAR* channel selection plotted against model level. It is clear from the distribution of averaging kernel peaks that the full spectrum provides a better vertical coverage and that more information is extracted from the observations. Of particular note is the improvements in both vertical resolution and sensitivity from



(a) Degrees of Freedom for Signal (DFS)



(b) Shannon Information Content (SIC)



(c) Number of pieces of information measured to better than measurement error

Figure 4.5: Information content measures for analyses with different number of channels assimilated: A=Full Spectrum; B=314 *Collard* Channel Set; C=183 *OPS* Channel Set; D=138 *VAR* Channel Set; E=1997 *Band1* Channels, F=290 PCs from full spectrum. All channel sets are detailed in appendix E. The red columns calculate the information content measure assuming $\mathbf{B} = \mathbf{MoistCov}$ and the blue columns assuming $\mathbf{B} = \mathbf{NMC}$.

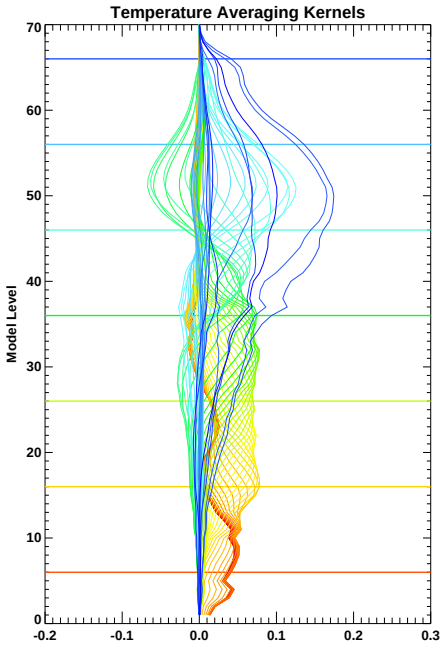
model level 30 upwards. There is also improved sensitivity to both temperature and water vapour in the lowest five model levels (although there is no corresponding improvement in vertical resolution). In fact the *VAR* channel selection is completley insensitive to water vapour in the lowest levels of the atmosphere. Figure 4.6 therefore implies that increasing the spectral coverage should improve the analysis throughout the model column, but that the greatest gains should be expected in the stratospheric temperature, and in the lowest model levels. The potential for impact on the stratospheric temperature is important, because there are relatively few observations assimilated that provide information at the top of the model column.

4.4 Summary

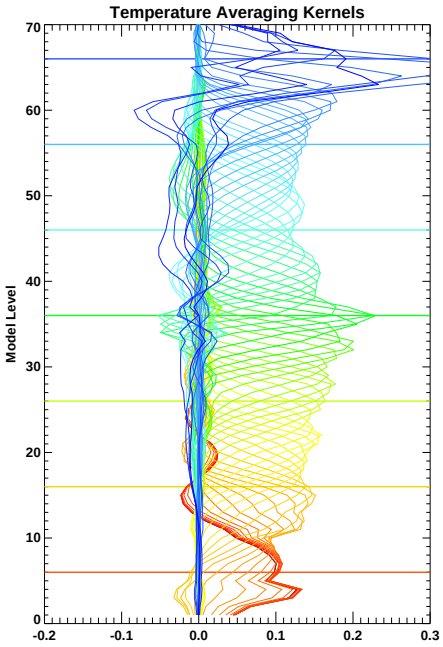
This chapter has explored the information content of IASI in the Met Office assimilation system. New FSO results that split the forecast impact into different regions in the vertical suggest that IASI is generally having the expected effect within the assimilation system. Water vapour and window channels provide information to the humidity fields, while temperature channels contribute mostly to potential temperature and wind fields. Higher peaking channels have impact high in the atmosphere, while lower peaking channels contribute more information to the lower and mid troposphere. This suggests that there are no obvious deficits that need correcting, and that to improve the amount of information extracted from IASI, other than by reducing observation errors through a more accurate forward operator, the best approach is to increase the spectral coverage presented to the 4D-Var analysis.

Information content measures have been investigated as tools to predict performance within a 1D-Var system. The DFS and SIC measures behave similarly, and DFS is chosen as the measure by which to examine the behaviour of improved 1D-Var systems in later chapters. The optimal linear equations have been used to assess the performance of an IASI analysis of temperature and water vapour profiles with two different $\mathbf{B}$ matrices and several different combinations of channel selection.

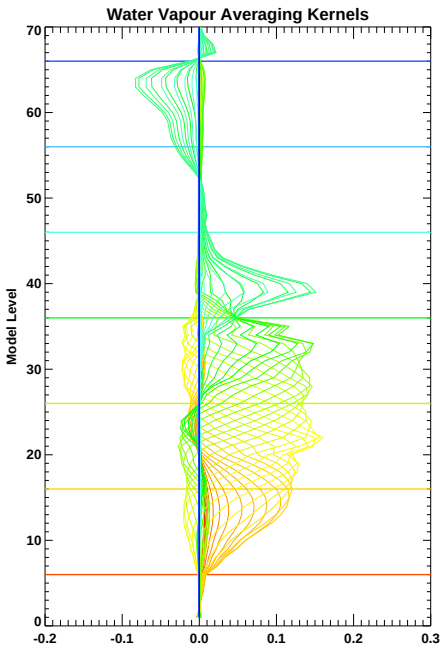
Increasing the number of channels assimilated improves the information content of the optimal analysis. The assimilation of the full spectrum from IASI, with no forward model or representativeness errors, would be expected to yield a DFS of approximately 20 (in line with previous estimates), of which approximately 50% comes from the longwave CO_2 band. The currently assimilated *VAR* channel selection provides less than half that information content (DFS ~8). There is significant scope to extract more information from the observations if the full spectrum can be



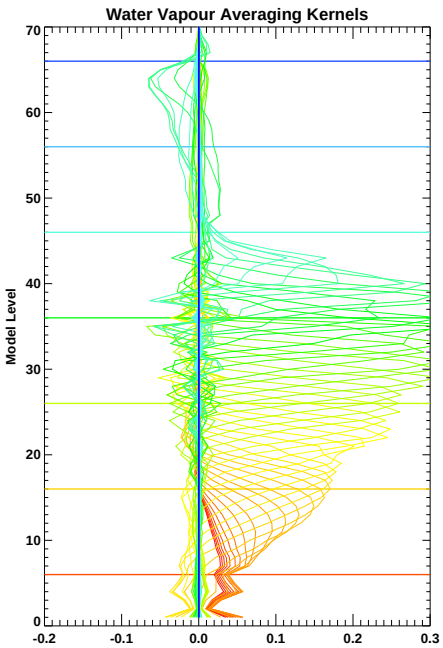
(a) Temperature, *VAR* channels



(b) Temperature, Full spectrum



(c) Water Vapour, *VAR* channels



(d) Water Vapour, Full spectrum

Figure 4.6: Comparison of averaging kernels for analysis using the *VAR* channel selection and using the full spectrum. $\mathbf{R}$ contains instrument noise only. Optimal analysis where $\mathbf{B}$ =**MoistCov**. Plotted against model level.

exploited.

The averaging kernels show that using the full spectrum should allow the assimilation system to access information about temperature and water vapour in the boundary layer, that is not present in the averaging kernels of the analysis from the current channel selection. The use of the full spectrum improves sensitivity throughout the atmospheric column, and increases the vertical resolution in the stratosphere. This increase in resolution and sensitivity in the stratosphere is of particular intrerest to NWP because there are relatively few observations sensitive to stratospheric temperature assimilated.

The results presented in this chapter justify the motivation for this thesis: IASI can provide much more information to the assimilation system if the full spectrum is used. Since it is not possible to assimilate 8461 channels, to make the most of IASI, a way must be found to compress this information into a form that can be assimilated.

Chapter 5

The behaviour of an optimal 1D-Var analysis

This chapter describes the behaviour of an optimal 1D system in line with the experimental set-up described in section 3.8.3. These results, which use the *VAR* 4D-Var channel selection and climatological **B**, are presented as a baseline against which results in subsequent chapters can be compared.

The results presented here are for the somewhat unrealistic scenario of an **R** matrix consisting of instrument noise only. This simulated environment allows us to see the maximum benefit that IASI could contribute to an atmospheric analysis if a perfect forward model existed and there were no errors of representation. It is particularly useful for the methods explored in this thesis, because calculations in principal component and reconstructed radiance space are more straightforward without other error terms (see chapter 6). Also, the covariance structure of the other major sources of error – radiative transfer and representation – is not well known for the full spectrum, as will be discussed further in chapter 7.

5.1 Optimal linear analysis

This section presents more detailed results of linear optimal estimation analysis using the *VAR* channels for the 1D-Var set-up presented in section 3.8.3 that is used throughout this thesis. The results are presented in a new way, proposed by Eyre and Hilton (2013). The analysis error covariance, **A**, is projected onto the eigenvectors of the **B** matrix allowing the exploration of which scales in the vertical are being influenced by the assimilation of IASI data. The quantity plotted is the square root of the diagonal of $\mathbf{VAV}^{\mathrm{T}}$ where **V** comes from an eigenvector decomposition of the background error term, i.e. $\mathbf{B} = \mathbf{V\Lambda V}^{\mathrm{T}}$. The value plotted for **B** is the square root of the eigenvalue ($\mathbf{\Lambda}$).

The analysis error is calculated using equation 3.11, and assuming instrument noise only, using the **B** matrix **MoistCov**. The profile used to generate the forward model Jacobian matrix, **H**, is the US Standard Atmosphere. As well as the eigenvector projection, the results are also presented in vertical profile form, with the standard deviation analysis error (the square root of the diagonal of the **A** matrix)

compared against the equivalent background error for temperature and humidity.

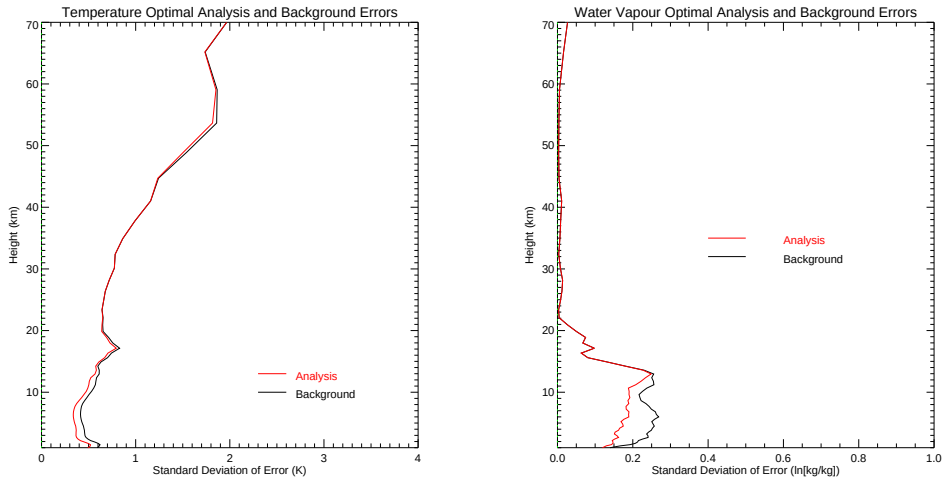
Note that although $\mathbf{B}$ is diagonal when projected onto its own eigenvectors, $\mathbf{A}$ is not because the observations are not localised in either profile or eigenvector space, and 'see' a combination of all the vectors. This is demonstrated by projecting all components of the optimal inverse analysis equation (equation 3.11) onto $\mathbf{V}$, the matrix of eigenvectors of $\mathbf{B}$.

$$\begin{aligned} \mathbf{V}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{V} &= \mathbf{V}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{V} + \mathbf{V}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{V} \\ &= \quad \mathbf{\Lambda}^{-1} \; + \mathbf{V}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{V} \end{aligned} \tag{5.1}$$

The term in $\mathbf{B}^{-1}$ collapses to diagonal matrix $\mathbf{\Lambda}^{-1}$, but because the term in $\mathbf{R}^{-1}$ does not, the matrix $\mathbf{V}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{V}$ has non-zero correlations. By examining the diagonal of this matrix, the reduction in the variances of each mode caused by the assimilation of IASI observations can be shown.

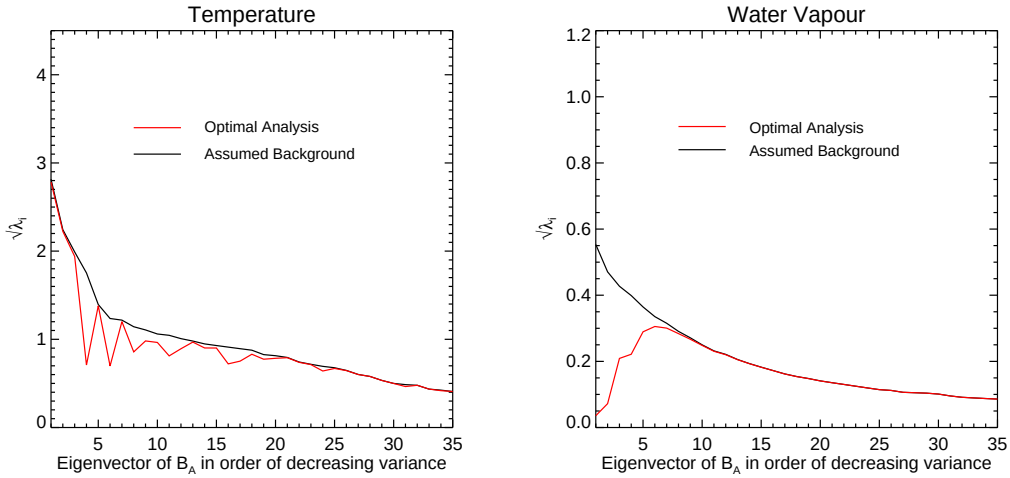
Figure 5.1 compares the analysis using the *VAR* channel selection and background errors in profile space. There is a small reduction of around 0.1 K in the tropospheric temperature errors but up to $0.1\ln(q)$ units through the tropospheric water vapour profile. Figure 5.2 shows the same errors in the eigenvector space of the **MoistCov** matrix. The errors in the directions of the first five to six humidity eigenvectors are improved with lower errors in the analysis. Examination of the eigenvectors shown in figure B.3, indicates that these vectors represent wavelengths of up to 7 km. Approximately seven temperature vectors show improved errors, and figure B.1 shows that these vectors have mostly tropospheric features; those with large stratospheric contributions are not greatly affected. The DFS for the whole analysis is 7.3, split almost equally between temperature and humidity. These improvements fall well short of the impact predicted by pre-launch studies (Collard, 1998; Prunet et al., 1998).

Figures 5.3 and 5.4 are equivalent to figures 5.1 and 5.2 but present the results for the use of the full spectrum. The improvements are much more substantial. In particular, the temperature vectors that have strong stratospheric components are now improved as well, and humidity vectors are improved out to 10–12 vectors, now influencing features with wavelengths of up to 3 km. The DFS for the analysis is 18.7 (this is plotted as the red bar in Column A of figure 4.5(a), of which 10.4 comes from the temperature analysis. These results show that there is still much more information to be captured from IASI than is possible with the *VAR* channel selection.



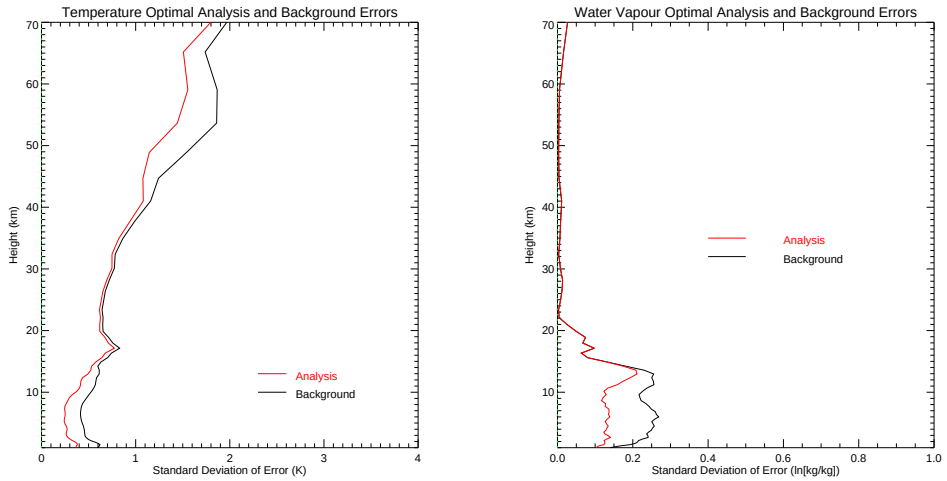
(a) Temperature **(b)** Water vapour

Figure 5.1: Comparison of linear optimal estimation analysis and background errors in profile space, for the assimilation of *VAR* channels with instrument noise only, where $\mathbf{B} =$ **MoistCov**. The analysis error is shown in red and the background error in black.



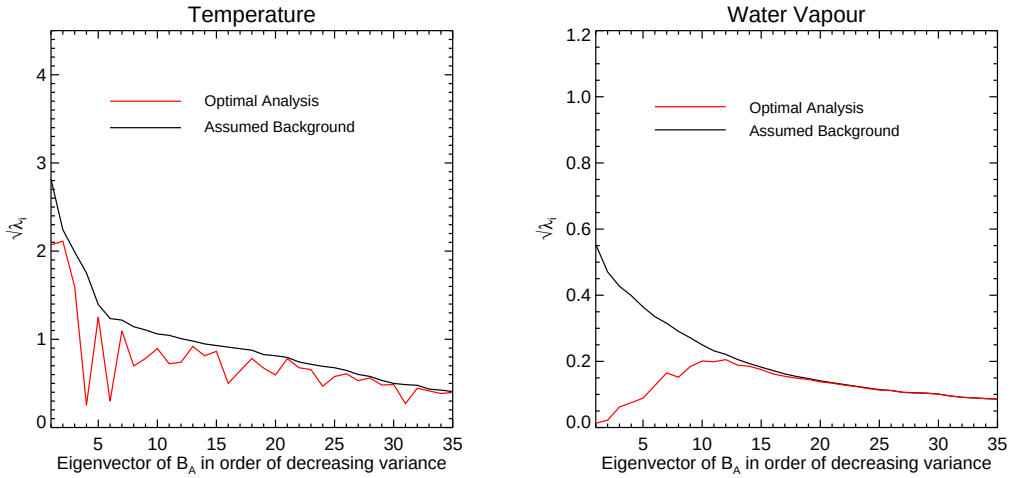
(a) Temperature **(b)** Water vapour

Figure 5.2: Comparison of linear optimal estimation analysis and background errors in the eigenvector space of $\mathbf{B}$, for the assimilation of *VAR* channels with instrument noise only, where $\mathbf{B} =$ **MoistCov**. The analysis error is shown in red and the background error in black.



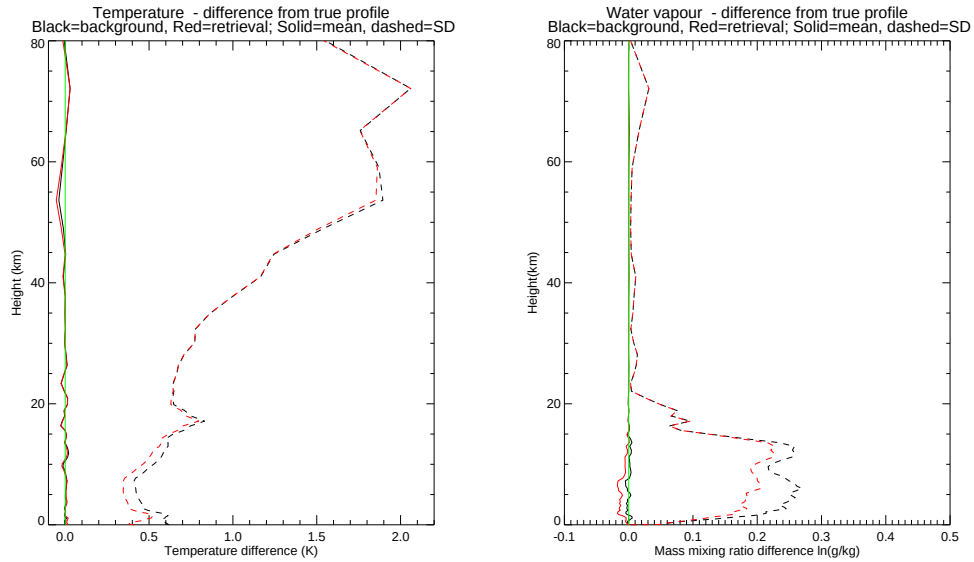
(a) Temperature **(b)** Water vapour

Figure 5.3: Comparison of linear optimal estimation analysis and background errors in profile space, for the assimilation of the full spectrum with instrument noise only, where $\mathbf{B} =$ **MoistCov**. The analysis error is shown in red and the background error in black.



(a) Temperature **(b)** Water vapour

Figure 5.4: Comparison of linear optimal estimation analysis and background errors in the eigenvector space of $\mathbf{B}$, for the assimilation of the full spectrum with instrument noise only, where $\mathbf{B} =$ **MoistCov**. The analysis error is shown in red and the background error in black.



(a) Temperature **(b)** Water vapour

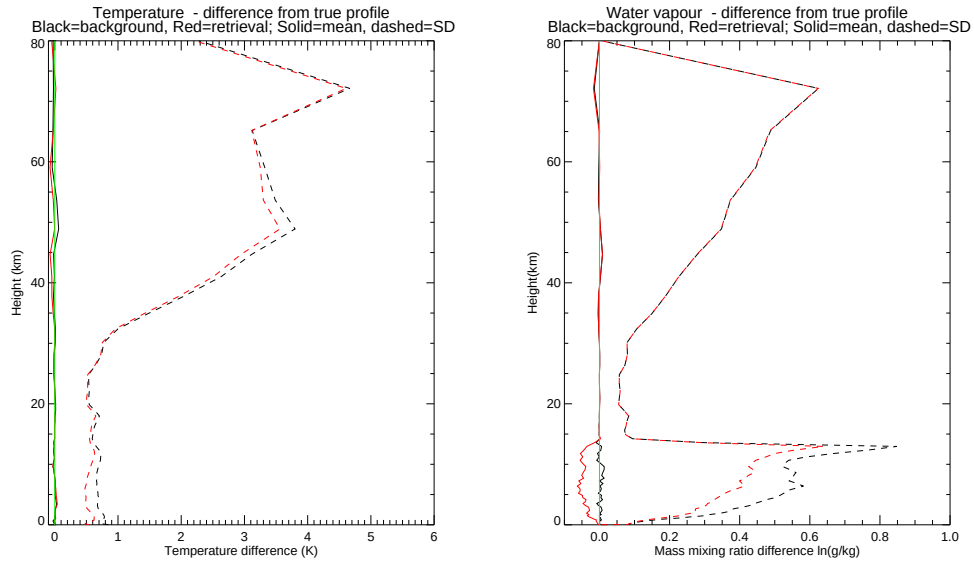
Figure 5.5: Mean and standard deviation of (Background–Truth) and (Analysis–Truth) for optimal analysis where background errors are calculated using the $\mathbf{B}$ matrix **MoistCov**. *VAR* channels. $\mathbf{R}$ contains instrument noise only.

5.2 Optimal 1D-Var

Section 5.1 has established, using the optimal estimation equations, how we might expect a 1D-Var linear analysis to perform if allowed to iterate to convergence. In practice, there are some non-linearities that affect the results, so this section examines the behaviour of a full 1D-Var. The set-up, as described in section 3.8.3, is equivalent to the optimal linear system used in the previous section. Again, the *VAR* channel selection is used for observations with only instrument noise added, and with background errors from **MoistCov**. In addition, results are presented for **NMC** background errors because the non-linearity effect is clearer.

Figure 5.5 shows the mean and standard deviation of retrievals and backgrounds against the true profiles. In terms of standard deviation, the optimal analysis performs similarly to the predictions of the linear equations, with up to 0.1 K improvements in the retrieval error over the **MoistCov** background error in temperature and 0.08 $\ln(q)$ units.

Figure 5.6 shows the equivalent for the **NMC** matrix (note the different scale on the x-axis). Here, because the background error is larger, the gains over the background are somewhat larger: up to 0.2 K in the troposphere and in the 40 to 60 km region of the temperature profile, and approximately 0.2 $\ln(q)$ units in the tropo-



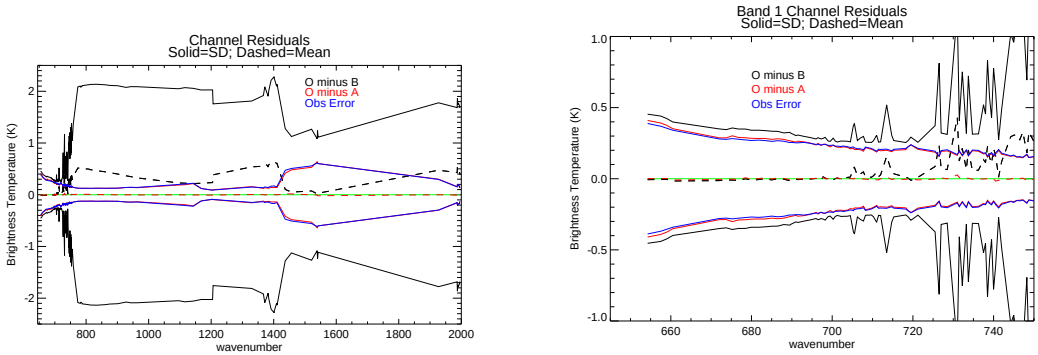
(a) Temperature **(b)** Water vapour

Figure 5.6: Standard deviation of (Background–Truth) and (Analysis–Truth) for optimal analysis where background errors are calculated using the $\mathbf{B}$ matrix **NMC**. *VAR* channels. $\mathbf{R}$ contains instrument noise only. The statistics are derived from 4348 observations, thus the standard error for the water vapour profile between 5 and $15\,\mathrm{km}$ is approximately 0.006.

spheric water vapour profile. This is also in line with the linear equations (not shown).

What linear theory does not predict is that the retrieved water vapour profile is biased relative to the background profile by approximately $0.05\ln(q)$ units with the **NMC** background errors. This is a consequence of non-linearity bias, described in Eyre and Collard (1999): although neither the observations nor the background errors in $\ln(q)$ are biased relative to the truth, the radiance response to the errors in $\ln(q)$ is non-linear, resulting in a bias in radiance space between the observations and the background. Figure 5.7 shows the mean and standard deviation of the observation departures from the background and analysis for the **NMC** background errors. The background bias is the black dashed line, which is approximately zero from 645–700 cm^{-1}, but in the water vapour band can be nearly $0.5\,\mathrm{K}$, much larger than the assumed observation errors. The analysis bias is the red dashed line: the analysed profiles are not biased with respect to the observations – the red dashed line is close to zero across the spectrum – but the non-linearity bias has resulted in a bias between the analysis and the truth in profile space.

Clearly, the introduction of a bias in the analysis is undesirable in an operational assimilation system. In practice, however, observations are bias-corrected before as-



(a) Full Spectrum **(b)** Band 1

Figure 5.7: Mean (dashed) and ±standard deviation (solid) of (Observation–Background) in black and (Observation–Analysis) in red, expressed as radiances for 1D-Var with $\mathbf{B} = \mathbf{NMC}$. Also shown in blue is the assumed observation error. Only the channels used in the retrieval (*VAR* selection) are plotted.

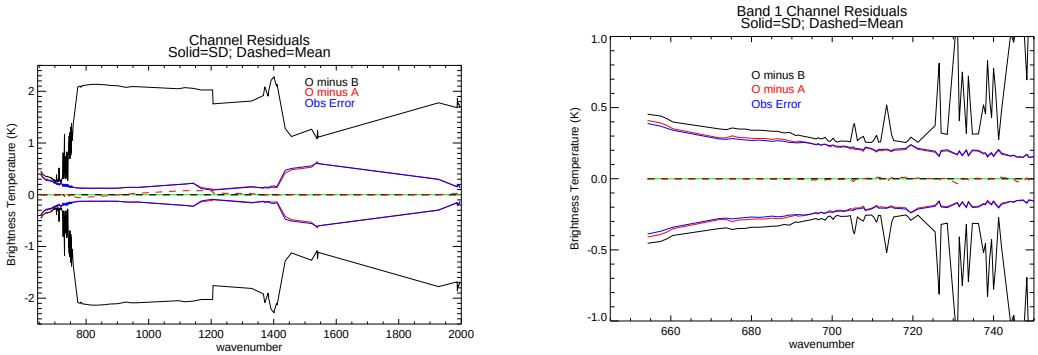
similation, and Section 5.2.1 explores what happens in a 1D-Var when the non-linearity bias is corrected before assimilation. It has not been investigated whether an attempt to estimate the covariance of the non-linearity error and combine it into the $\mathbf{R}$ matrix would help to mitigate this problem.

5.2.1 Bias correction of the non-linearity bias

In the 4D-Var assimilation system, Observation–Background (O–B) statistics are monitored and are used to generate bias correction coefficients to try to ensure that the observations and background remain unbiased. The scheme of Harris and Kelly (2001) that is used operationally uses model-based predictors (e.g. 850–300 hPa thickness) to which the calculated coefficients are applied to generate a bias correction for each channel of each observation. In the Met Office 4D-Var, there are no model bias terms. Thus, biases are assumed to be part of the observation, regardless of whether they are instrumental in origin. The non-linearity bias demonstrated above is one such example.

The effect of bias correcting away the non-linearity bias is investigated with an optimal 1D-Var experiment using $\mathbf{B} = \mathbf{NMC}$, with observations that have been bias-corrected by the subtraction of the average non-linearity bias (i.e. the dashed black line in figure 5.7). This represents a simpler bias correction than is applied in practice but the principles are similar, in that the applied correction renders the average O–B statistics unbiased, as shown in figure 5.8.

Figure 5.9 shows the statistics of the analysed profiles. The water vapour analysis



(a) Full Spectrum **(b)** Band 1

Figure 5.8: Mean (dashed) and ±standard deviation (solid) of (Observation–Background) in black and (Observation–Analysis) in red, expressed as radiances for 1D-Var with $\mathbf{B} = \mathbf{NMC}$. Observations are bias corrected using the values of the Observation–Background bias shown as the black dashed line in figure 5.7. Also shown in blue is the assumed observation error. Only the channels used in the retrieval (*VAR* selection) are plotted

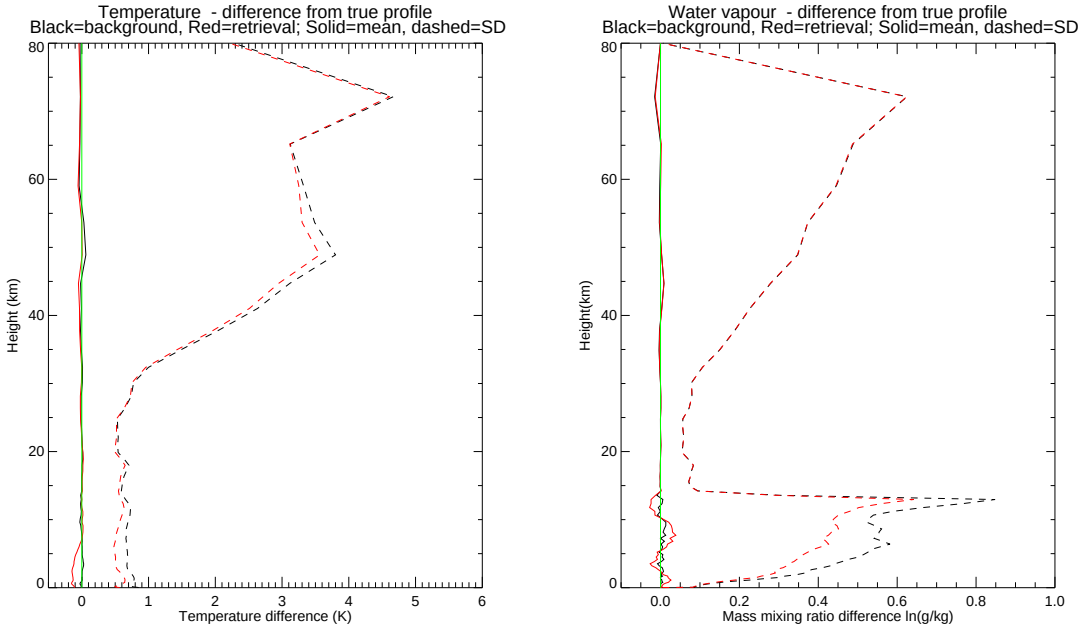
bias is now slightly reduced, although it oscillates from positive to negative. However, the temperature analysis is now biased with respect to the true profiles by about 0.15 K from the ground up to about 7 km. This is also undesirable. The analysis remains unbiased in radiance space.

It is not clear how this behaviour could be addressed in the operational system, but the observation errors used in 4D-Var are very different from those used here, being diagnosed quantities including forward model and representation errors, and so it is not clear whether this issue is important in the real system.

5.3 Summary

This chapter provides a baseline against which the results in subsequent chapters can be compared. Linear optimal estimation results are shown in profile space and in the eigenvector space of the background error covariance matrix. The results of the linear optimal estimation are then verified using a simulated 1D-Var experiment. The 1D-Var analysis is improved over the background in terms of standard deviation by amounts consistent with the optimal prediction, but a bias is introduced into the water vapour analysis because of non-linearity errors, demonstrating the importance of the assumption of linearity in the optimal analysis equations of section 3.2.

Assuming $\mathbf{B} = \mathbf{MoistCov}$, with the *VAR* channel selection the analysis is improved by around 0.1 K in the troposphere for temperature and 0.08 $\ln(q)$ units (a relative improvement of approximately 25%) in humidity. About five or six eigen-



(a) Temperature **(b)** Water vapour

Figure 5.9: Mean and standard deviation of (Background–Truth) and (Analysis–Truth) for optimal analysis where background errors are calculated using the $\mathbf{B}$ matrix **NMC**. Observations are bias corrected using Figure 5.7 before the 1D-Var. *VAR* channels. $\mathbf{R}$ contains instrument noise only. The statistics are derived from 4348 observations, thus the standard error for the water vapour profile between 5 and $15\,\mathrm{km}$ is approximately 0.006 and for temperature in the lower troposphere is approximately 0.0075.

vectors of $\mathbf{B}$ in each case show improvement, representing vectors with wavelengths of up to 7 km in humidity, much lower than the vertical resolution predicted from the instrument specification, and vectors with mostly tropospheric features for temperature. The DFS for the whole analysis is 7.3, split almost equally between temperature and humidity, which falls well short of the impact predicted by pre-launch studies. The optimal estimation equations predict that much more impact could be gained from IASI if the full spectrum could be assimilated, in particular smaller-scale features in the vertical would be improved for the water vapour analysis, and the stratospheric temperature analysis could also be improved.

Chapters 6 and 7 introduce principal component compression and reconstructed radiances, two methods that might allow the practical assimilation of more spectral information from IASI, without having to resort to using 8461 channels, which is too expensive computationally. Chapters 8 and 9 then present linear optimal analysis and 1D-Var results for these two methods that should be compared with the performance of the *VAR* channel selection as presented in this chapter.

The results presented in this chapter and in chapters 8 and 9 assume that the $\mathbf{B}$ matrix is well known, but it was explained in section 3.6.1 that this term is necessarily misspecified because of estimation and modelling. Chapter 10 investigates the effect of a misspecified $\mathbf{B}$ matrix on the predicted behaviour of the IASI assimilation, again building on the optimal results presented in this chapter.

Chapter 6

Principal Component Compression and Reconstructed Radiances: Theoretical Development

Hyperspectral sounder spectra can be distributed and processed more efficiently than in their raw form, by the use of principal component (PC) compression. Eigenvectors are taken from a covariance matrix of radiance climatology, and PC scores for each observation are then computed for only the leading eigenvectors. Thus the full spectrum of several thousand channels can be compressed into just a few hundred PC scores. The eigenvectors that are discarded mostly contain random measurement noise. The term reconstructed radiances refers to the truncated PC score representation of the spectrum that has been converted back into radiances. These radiances are similar to the raw spectrum, but the truncation reduces the random noise in the measurements. However, errors that are spectrally correlated in the original data, such as those arising from inadequate correction of detector non-linearity, will be preserved by PC compression and the remaining signal has errors that are significantly correlated between channels.

Figure 4.5 demonstrated that 290 PC scores were able to capture most of the information in the 8461 channels of IASI, despite condensing each observation into a much smaller number of observed quantities. This theoretical result makes the technique worthy of further consideration. PC Scores are being investigated for assimilation by several groups at the present time, with promising early results: for example, ECMWF have demonstrated that 165 channels could be represented instead by 20 PC Scores without significant differences in forecast accuracy (Matricardi and McNally, 2014).

The main drawback of PC scores is that each eigenvector contains information from throughout the atmospheric column. This makes it very difficult to use the parts of the observation that are unaffected by cloud or surface temperature and emissivity, limiting the use of the data to only those pixels where every part of the atmospheric column can be well modelled, and every PC can be well forward-modelled. This severely restricts usage of the data in cloudy conditions. Month-long trials at the Met

Office in 2007 demonstrated that the inclusion of AIRS channels above cloud doubled the impact of the data; we would not want to lose these important observations.

An alternative would be to assimilate reconstructed radiances. Collard et al. (2010) describe reconstructed radiances in some detail, and present some early examples of their use in NWP systems. The main difference between reconstructed radiances and the raw spectrum is the highly non-diagonal observation error covariance. The lack of evidence for any benefit in the assimilation of reconstructed radiances in the early studies, coupled with the knowledge that error properties differ, suggests that the channel selections in common use for raw radiances (e.g. Collard, 2007), which were constructed under an assumption of uncorrelated errors, are not appropriate for reconstructed radiances.

This chapter begins by outlining the calculation of PC scores and reconstructed radiances and demonstrates their theoretical equivalence. There follows a discussion about the structure of the reconstructed radiance error covariance matrix, which is required for optimal retrieval or assimilation, and also influences the choice of reconstructed radiance channels assimilated. The effect of using a raw radiance forward model rather than a reconstructed radiance forward model is discussed, and it is demonstrated that this leads to a scenario referred to as suboptimal-in-**H**, for which the assimilation equations are presented and the error terms derived.

6.1 PC compression and reconstructed radiance calculations

The mathematics of PC compression and reconstruction are well known, and have been presented in papers such as Antonelli et al. (2004) and Collard et al. (2010), but the practical consequences and treatment of error terms are not well described in the literature. Chapter 7 introduces a method for making the best use of the reconstructed radiance spectrum for assimilation or retrieval purposes. To this end, the effects of the PC compression on the observation error covariance term must be understood.

The method presented here for the calculation of PC scores are reconstructed is a standard method, and it is the same as that used by EUMETSAT in the production of their IASI Level 1c PC Score product (Hultberg, 2009).

Let us first define:

n**:** number of channels in raw radiance observation (for the full IASI spectrum 8461; for bands 1 and 2, 5116)

p**:** number of retained PCs (for IASI bands 1, 2 and 3, respectively, 90, 120 and 80)

$\mathbf{y}_{obs}$**:** Raw observed radiance spectrum

N: Assumed instrument noise covariance matrix (size $n \times n$)

Before compression, it is usual to subtract a climatological mean spectrum $\mathbf{y}_m$, and to normalise the spectrum with an estimate of the instrument noise $\mathbf{N}^{1/2}$, to give observation $\mathbf{y}$:

$$\mathbf{y} = \mathbf{N}^{-1/2}(\mathbf{y}_{obs} - \mathbf{y}_m) \tag{6.1}$$

with assumed instrument noise:

$$\mathbf{N}^{-1/2}\mathbf{N}\mathbf{N}^{-1/2} = \mathbf{I} \tag{6.2}$$

The PC basis vectors are derived from $\mathbf{C}$ (size $n \times n$), an observation climatology consisting of many tens of thousands of normalised observed atmospheric spectra:

$$\mathbf{C} = Cov(\mathbf{y}) \tag{6.3}$$

The PC vectors are the eigenvectors, $\mathbf{L}$, of $\mathbf{C}$:

$$\mathbf{C} = \mathbf{L}\mathbf{\Lambda}\mathbf{L}^{\mathrm{T}} \tag{6.4}$$

where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues. Figure 6.1 shows the first three principal components for IASI band 1, as used by EUMETSAT and in this thesis.

The PC vector basis is then truncated to compress the data and remove components consisting mostly of random measurement noise. The truncation is based on retaining only the first p vectors with the largest associated eigenvalues. Typically, p would be of the order of 300 for an IASI spectrum; this is the number of vectors used by EUMETSAT. The truncation, effected by the PC selection matrix $\mathbf{P}$ (size $n \times p$), results in a PC compression matrix, $\mathbf{L}_p$ (size $n \times p$):

$$\mathbf{L}_p = \mathbf{L}\mathbf{P} \tag{6.5}$$

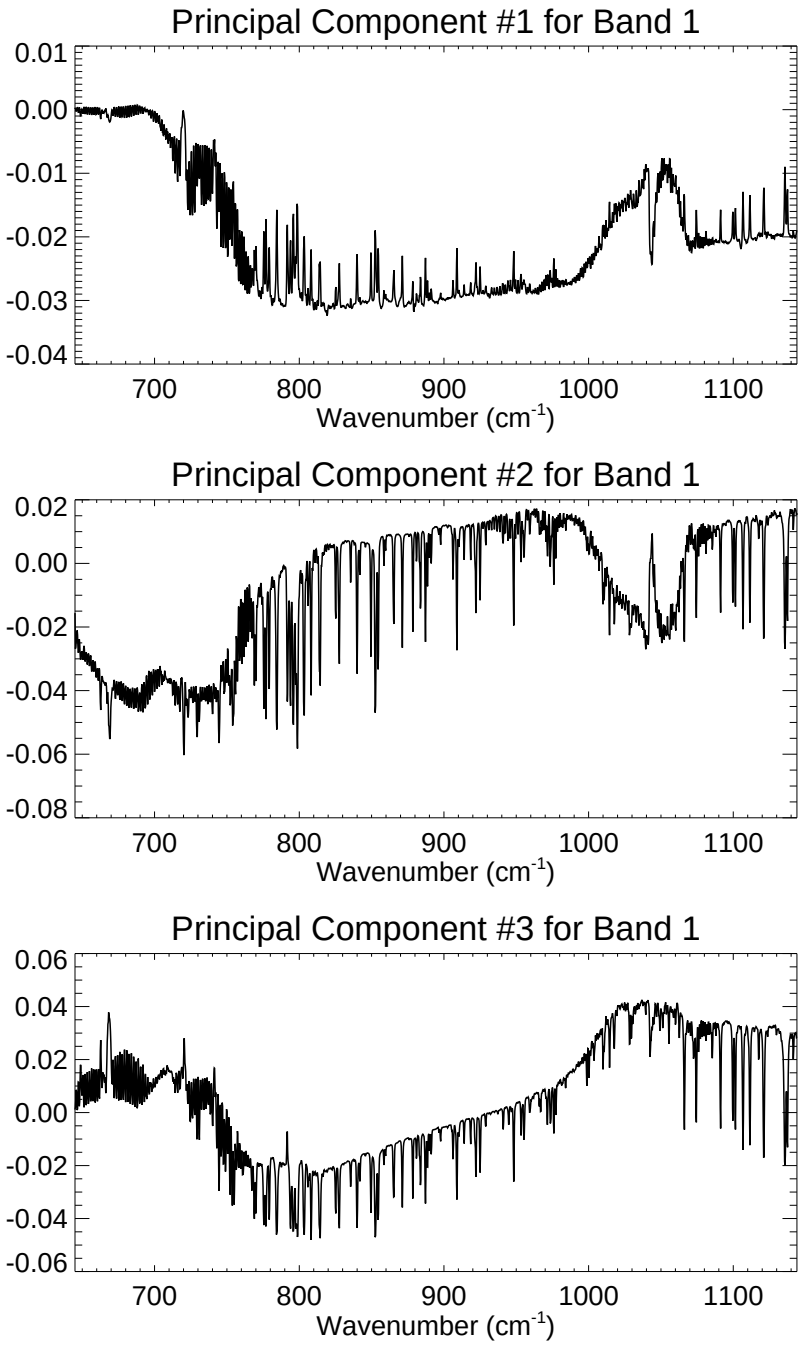


Figure 6.1: The first three EUMETSAT principal components for Band 1

from which we can generate $\mathbf{p}$, the vector of PC scores of $\mathbf{y}$:

$$\mathbf{p} = \mathbf{L}_p^{\mathrm{T}}\mathbf{y} \tag{6.6}$$

The assumed instrument noise covariance of the PC scores is thus

$$\mathbf{L}_p^{\mathrm{T}}\mathbf{N}^{-1/2}\mathbf{N}\mathbf{N}^{-1/2}\mathbf{L}_p = \mathbf{I} \tag{6.7}$$

The noise-normalised reconstructed radiance spectrum deviation from the mean, $\tilde{\mathbf{y}}$, is then calculated by pre-multiplying $\mathbf{p}$ by $\mathbf{L}_p$:

$$\tilde{\mathbf{y}} = \mathbf{L}_p\mathbf{p} = \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{y} \tag{6.8}$$

Thus, the full matrix transform that is used to convert from raw to reconstructed normalised radiances is given by $\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}$, which is also the assumed instrument noise covariance for the full reconstructed radiance spectrum:

$$\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{N}^{-1/2}\mathbf{N}\mathbf{N}^{-1/2}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}} = \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}} \tag{6.9}$$

This matrix product is of size $n \times n$, but its rank is at most p. In other words, if the full spectrum is reconstructed (or, even, just more channels than there are PC scores), the resulting matrix is rank deficient. This has consequences for the use of reconstructed radiances in assimilation systems, as the instrument noise covariance matrix, or indeed any observation error covariance matrix, projected into reconstructed radiance space will also be rank deficient, non-positive definite, and therefore not invertible. In other words, only a maximum of p reconstructed channels may be independent: any more are simply linear combinations of the first p reconstructed channels. In practice, the rank is often lower, as will be explained in section 7.1.

In order to assimilate reconstructed radiances, we therefore need a channel selection for assimilation of n_{ass} channels, where $n_{ass} \leq p$. This channel selection can be represented by a sparse matrix $\mathbf{S}$ (size $n_{ass} \times n$). The normalised reconstructed radiances are now defined:

$$\tilde{\mathbf{y}}_{ass} = \mathbf{L}_s\mathbf{p} = \mathbf{S}\mathbf{L}_p\mathbf{p} = \mathbf{S}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{y} \tag{6.10}$$

$$\mathbf{L}_s\mathbf{L}_p^{\mathrm{T}}\mathbf{N}^{-1/2}\mathbf{N}\mathbf{N}^{-1/2}\mathbf{L}_p\mathbf{L}_s^{\mathrm{T}} = \mathbf{L}_s\mathbf{L}_s^{\mathrm{T}} \tag{6.11}$$

Such a channel selection matrix can equally be applied to raw radiances, and usually is, for reasons of computational efficiency and mathematical stability. It is

worth noting, however, that a channel selection of 300 raw radiance channels will not embed as much information as the first 300 PC scores (compare, for example, columns B and F in figure 4.5). To simplify the equations hereafter, the use of $\mathbf{S}$ will be implicitly assumed and $\tilde{\mathbf{y}}$ will be used as an abbreviation for $\tilde{\mathbf{y}}_{ass}$, and $\mathbf{y}$ for $\mathbf{S}\mathbf{y}$. Chapter 7 describes work to construct this matrix $\mathbf{S}$ in such a way as to yield an invertible error covariance matrix for a channel selection of reconstructed radiances for use in NWP.

In practice, a complete reconstructed radiance vector, rather than a normalised deviation from the mean, is assimilated:

$$\tilde{\mathbf{y}}_{obs} = \mathbf{N}^{1/2}(\tilde{\mathbf{y}} + \mathbf{y}_m) \tag{6.12}$$

with assumed instrument error covariance

$$\tilde{\mathbf{N}} = \mathbf{N}^{1/2}\mathbf{L}_s\mathbf{L}_p^{\mathrm{T}}\mathbf{N}^{-1/2}\mathbf{N}\mathbf{N}^{-1/2}\mathbf{L}_p\mathbf{L}_s^{\mathrm{T}}\mathbf{N}^{1/2} = \mathbf{N}^{1/2}\mathbf{L}_s\mathbf{L}_s^{\mathrm{T}}\mathbf{N}^{1/2} \tag{6.13}$$

However, for simplification of the equations, it is assumed in the rest of this chapter that normalised measurement vectors $\mathbf{y}$ or $\tilde{\mathbf{y}}$ are assimilated.

6.2 Retrieval in an optimal estimation framework

Optimal estimation theory is covered in chapter 3, but a comparison between raw and reconstructed radiances, of the calculation of increments and the error terms involved is a useful illustration of the application of the PC matrix transforms. In addition to the vectors and matrices defined above, the assimilation system requires the use of a forward model H_{obs} to compute a radiance, $H_{obs}(\mathbf{x})$, from the atmospheric state vector $\mathbf{x}$. The forward modelled deviation from the mean climatological radiance, $H(\mathbf{x})$, is calculated straightforwardly as:

$$H(\mathbf{x}) = H_{obs}(\mathbf{x}) - \mathbf{y}_m \tag{6.14}$$

Although it is possible to use a dedicated PC-based forward model for assimilation of PCs or reconstructed radiances, it is also possible to construct the forward-modelled reconstructed radiance $\tilde{H}(\mathbf{x})$ using the same matrix transforms used to project the

radiances themselves. Thus:

$$H_{pc}(\mathbf{x}) = \mathbf{L}_p^{\mathrm{T}} H(\mathbf{x}) \tag{6.15}$$

$$\tilde{H}(\mathbf{x}) = \mathbf{L}_s \mathbf{L}_p^{\mathrm{T}} H(\mathbf{x}) \tag{6.16}$$

The forward model carries an error, which can mostly be removed by bias correction, but the residual is usually treated as a random error about the true radiance with covariance $\mathbf{F}$. It is also helpful to introduce an instrument error term about the true radiance, $\mathbf{E}$. In the case that the true instrument noise is used for noise normalization of the raw radiance spectra prior to PC compression, $\mathbf{E} = \mathbf{I}$, and the reconstructed radiance instrument noise simplifies to $\mathbf{L}_s\ \mathbf{L}_s^{\mathrm{T}}$ as shown in equation 6.11. The total error covariance matrix, taking account of instrument and forward model errors is, as elsewhere, represented here by $\mathbf{R}$.

Let $\mathbf{y}_t$ be the normalised deviation of the true raw radiance vector from the climatological observed mean spectrum. For optimal estimation with raw radiances, the observation increments, $\delta\mathbf{y}$, calculated during minimization, and their respective error covariance, $\mathbf{R}$, are given by:

$$\delta\mathbf{y} = (\mathbf{y} - \mathbf{y}_t) - (H(\mathbf{x}) - \mathbf{y}_t) \tag{6.17}$$

$$\mathbf{R} = \mathbf{E} + \mathbf{F} \tag{6.18}$$

The part of the cost function representing the fit of the current guess state to the raw observation, J_o, can be written as:

$$J_o = (\mathbf{y} - H(\mathbf{x}))^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x})) \tag{6.19}$$

For optimal estimation with reconstructed radiances, the equivalent terms, $\delta\tilde{\mathbf{y}}$ and $\tilde{\mathbf{R}}$, are given by:

$$\begin{aligned} \delta\tilde{\mathbf{y}} &= (\tilde{\mathbf{y}} - \tilde{\mathbf{y}}_t) - (\tilde{H}(\mathbf{x}) - \tilde{\mathbf{y}}_t) \\ &= \mathbf{L}_s \mathbf{L}_p^{\mathrm{T}} ((\mathbf{y} - \mathbf{y}_t) - (H(\mathbf{x}) - \mathbf{y}_t)) \end{aligned} \tag{6.20}$$

and

$$\begin{aligned} \tilde{\mathbf{R}} &= \mathbf{L}_s \mathbf{L}_p^{\mathrm{T}} \mathbf{E} \mathbf{L}_p \mathbf{L}_s^{\mathrm{T}} + \mathbf{L}_s \mathbf{L}_p^{\mathrm{T}} \mathbf{F} \mathbf{L}_p \mathbf{L}_s^{\mathrm{T}} \\ &= \mathbf{L}_s \mathbf{L}_p^{\mathrm{T}} \mathbf{R} \mathbf{L}_p \mathbf{L}_s^{\mathrm{T}} \end{aligned} \tag{6.21}$$

The part of the cost function representing the fit of the current guess state to the

reconstructed radiance observation, $\tilde{J}_o$, can be written as:

$$\tilde{J}_o = \quad (\,\mathbf{L}_s\mathbf{L}_p^{\mathrm{T}}(\mathbf{y} - H(\mathbf{x}))\,)^{\mathrm{T}}(\,\mathbf{L}_s\mathbf{L}_p^{\mathrm{T}}\mathbf{R}\mathbf{L}_p\mathbf{L}_s^{\mathrm{T}}\,)^{-1}(\,\mathbf{L}_s\mathbf{L}_p^{\mathrm{T}}(\mathbf{y} - H(\mathbf{x}))\,) \tag{6.22}$$

6.2.1 Equivalence between PC assimilation and reconstructed radiance assimilation

It can be demonstrated that the cost function for assimilation of reconstructed radiances is identical to that of the PC assimilation under certain assumptions, following Hultberg and August (2013). The only difference in the assimilation of reconstructed radiances is that an extra rotation through $\mathbf{L}$ is performed. The assumptions required for this to hold are:

1. The forward models should be the same, i.e. $\tilde{H}(\mathbf{x}) = \mathbf{L}_s H_{pc}(\mathbf{x})$.
2. The matrix $\mathbf{L}_s\mathbf{L}_p^{\mathrm{T}}\mathbf{R}\mathbf{L}_p\mathbf{L}_s^{\mathrm{T}}$ should be invertible and well-conditioned.
3. In order for the PC and reconstructed radiance systems to be equivalent, $\mathbf{L}_s$ should be invertible in order that the central inverse matrix term in equation 6.22 can be split into the product of inverse matrices. $\mathbf{L}_s$ is only invertible when square, hence the number of channels reconstructed should be exactly p.

Here, we take $H_{pc}(\mathbf{x}) = \mathbf{L}_p^{\mathrm{T}} H(\mathbf{x})$. The J_o term for PC scores is then given by:

$$J_{o\_pc} = \quad (\,\mathbf{L}_p^{\mathrm{T}}(\mathbf{y} - H(\mathbf{x}))\,)^{\mathrm{T}}(\,\mathbf{L}_p^{\mathrm{T}}\mathbf{R}\mathbf{L}_p\,)^{-1}(\,\mathbf{L}_p^{\mathrm{T}}(\mathbf{y} - H(\mathbf{x}))\,) \tag{6.23}$$

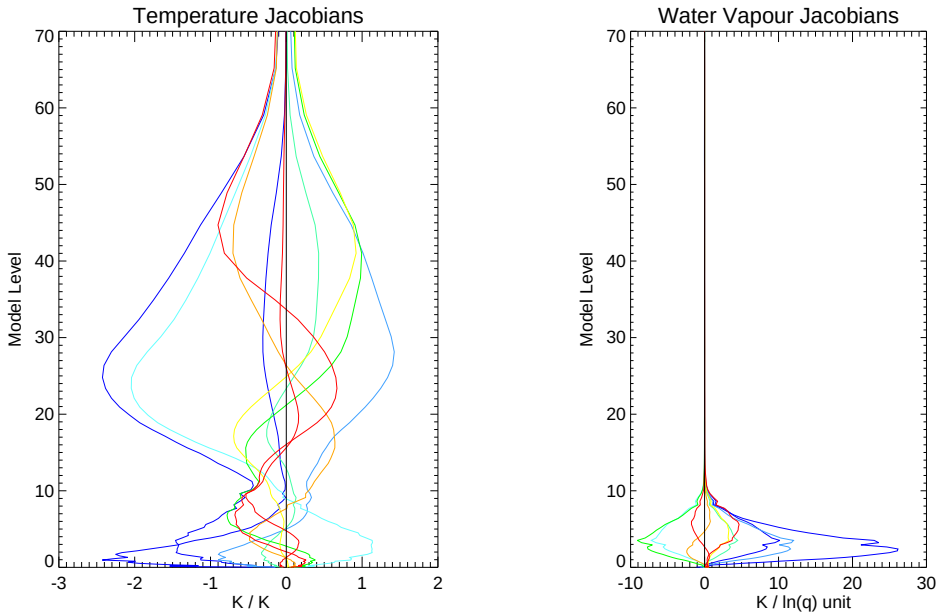
Equation 6.22 can be rearranged to show the equivalence between $\tilde{J}_o$ and J_{o_pc}:

$$\begin{aligned}
\tilde{J}_o &= (\,\mathbf{L}_s\mathbf{L}_p^{\mathrm{T}}(\mathbf{y} - H(\mathbf{x}))\,)^{\mathrm{T}} \quad (\,\mathbf{L}_s\mathbf{L}_p^{\mathrm{T}}\mathbf{R}\mathbf{L}_p\mathbf{L}_s^{\mathrm{T}}\,)^{-1} \quad (\,\mathbf{L}_s\mathbf{L}_p^{\mathrm{T}}(\mathbf{y} - H(\mathbf{x}))\,) \\
&= (\,\mathbf{L}_p^{\mathrm{T}}(\mathbf{y} - H(\mathbf{x}))\,)^{\mathrm{T}}\mathbf{L}_s^{\mathrm{T}} \quad (\,\mathbf{L}_s\mathbf{L}_p^{\mathrm{T}}\mathbf{R}\mathbf{L}_p\mathbf{L}_s^{\mathrm{T}}\,)^{-1} \quad \mathbf{L}_s(\,\mathbf{L}_p^{\mathrm{T}}(\mathbf{y} - H(\mathbf{x}))\,) \\
&= (\,\mathbf{L}_p^{\mathrm{T}}(\mathbf{y} - H(\mathbf{x}))\,)^{\mathrm{T}}\mathbf{L}_s^{\mathrm{T}} \quad \mathbf{L}_s^{-\mathrm{T}}(\,\mathbf{L}_p^{\mathrm{T}}\mathbf{R}\mathbf{L}_p\,)^{-1}\mathbf{L}_s^{-1}\,\mathbf{L}_s(\,\mathbf{L}_p^{\mathrm{T}}(\mathbf{y} - H(\mathbf{x}))\,)
\end{aligned}$$

the pairs of matrices in the middle thus cancel, leaving

$$\begin{aligned}
&= (\,\mathbf{L}_p^{\mathrm{T}}(\mathbf{y} - H(\mathbf{x}))\,)^{\mathrm{T}} \qquad (\,\mathbf{L}_p^{\mathrm{T}}\mathbf{R}\mathbf{L}_p\,)^{-1} \qquad (\,\mathbf{L}_p^{\mathrm{T}}(\mathbf{y} - H(\mathbf{x}))\,) \\
&= J_{o_pc}
\end{aligned}$$

This demonstrates that, theoretically at least, it should be possible to extract the full information content of the PC scores, and thus approach the full content of the IASI spectrum, from p reconstructed radiances, carefully chosen to ensure that $\mathbf{L}_s$ is invertible.



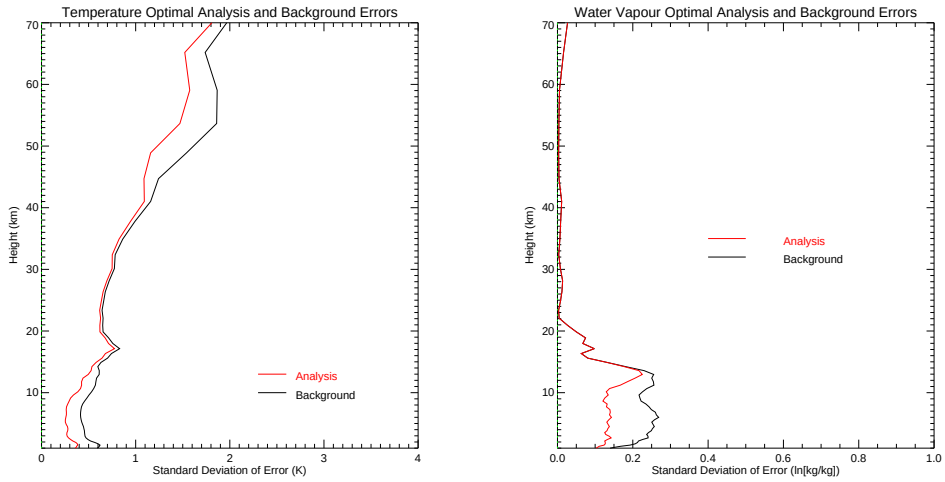
(a) Temperature **(b)** Ln(q)

Figure 6.2: Jacobians for the US Standard atmosphere for the first ten band 1 EUMETSAT PC scores

There are, however, important differences in the assimilation of PC scores and reconstructed radiances, notably in their Jacobians. Figure 6.2 shows the Jacobians of the first 10 EUMETSAT PC Scores. The forms are much less localised than those of channel radiances, which has implication for assimilation in cloudy areas because it is not possible to use methods like those of McNally and Watts (2003) or Pavelin et al. (2008) to assimilate only parts of the spectrum insensitive to cloud. Other PC compression schemes would lead to differently shaped Jacobians, but the lack of localisation in the vertical is common to all currently used schemes. The EUMETSAT PCs are constructed for each band separately: the picture is even more complex if Bands 1 and 2 are combined within the same eigenvectors, as for the PC forward model PC-RTTOV (Matricardi, 2010), as then the eigenvectors combine channels principally sensitive to atmospheric temperature, surface properties and water vapour. Reconstructed radiance Jacobians are very similar to raw radiance Jacobians (see figures 8.1 and 8.2 in chapter 8).

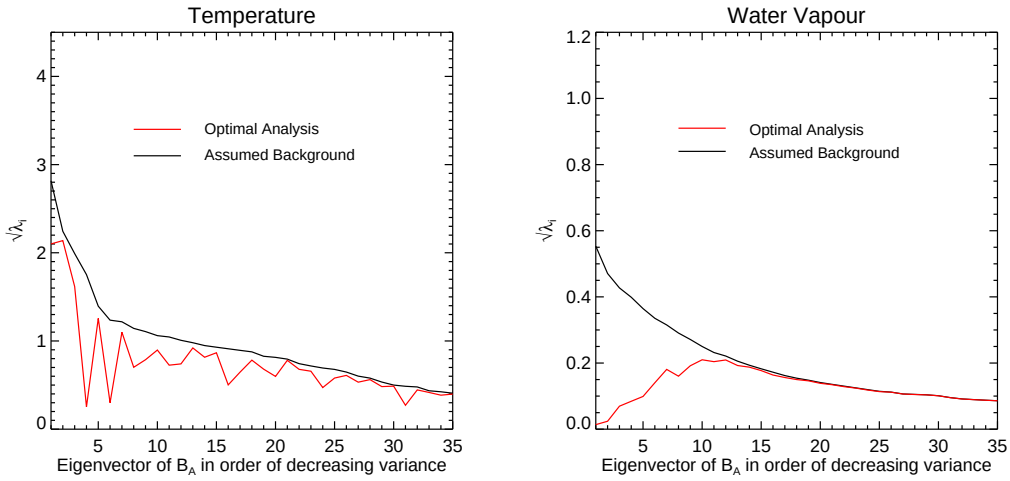
6.2.2 Optimal linear analysis for PCs

Figures 6.3 and 6.4 present the optimal linear 1D analysis results for the assimilation of PC scores from IASI, in line with the experimental set-up described in section 3.8.3. The PC eigenvectors used are those of the EUMETSAT version 102



(a) Temperature **(b)** Water Vapour

Figure 6.3: Comparison of linear optimal estimation analysis and background errors in profile space, for the assimilation of 290 PC scores with instrument noise only, where $\mathbf{B} =$ **MoistCov**. The analysis error is shown in red and the background error in black.



(a) Temperature **(b)** Water Vapour

Figure 6.4: Comparison of linear optimal estimation analysis and background errors in the eigenvector space of $\mathbf{B}$, for the assimilation of 290 PC scores with instrument noise only, where $\mathbf{B} =$ **MoistCov**. The analysis error is shown in red and the background error in black.

Level 1c IASI radiance product (Hultberg, 2009). This 1D analysis gives the value of DFS that is plotted as Column F for the **MoistCov** matrix in figure 4.5.

Figure 6.3 should be compared with figures 5.1 and 5.3, and Figure 6.4 with figures 5.2 and 5.4. The PC-compressed spectrum gives results that are virtually identical to those from the full spectrum. In other words, the 290 orthogonal vectors

contain almost exactly the same information on the vertical profile of temperature and humidity as the original spectrum.

6.3 Development of the suboptimal-in-H assimilation system for reconstructed radiances

This section introduces new equations to describe the error properties of reconstructed radiances assimilated with a raw radiance forward model, a scenario referred to as suboptimal-in-**H**. This situation is likely to arise in the practical assimilation of reconstructed radiances, because using $\tilde{H}(\mathbf{x}) = \mathbf{S}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}H(\mathbf{x})$ is an inefficient way to simulate reconstructed radiances since the full raw radiance spectrum needs to be simulated in order to forward model just a few reconstructed channels. Whilst PC-based forward models exist and are in use in several retrieval schemes at the present time, none of these are currently set up to forward model the PC scores disseminated by EUMETSAT. The relevant quantities could be calculated via another set of PCs from such a forward model and a further matrix transformation. This would still be more efficent than simulating the full raw spectrum, but those working with PC forward models tend to be interested in assimilating PC scores directly and the PC forward models available were considered insufficiently mature to attempt this approach here. Instead, most existing assimilation systems are set up to use traditional radiative transfer models simulating raw radiances. Experiments with reconstructed radiance assimilation to date (e.g. Collard et al., 2010) have therefore involved forward modelling the reconstructed radiances as though they are raw radiances.

If the approach is taken to forward model raw radiances, an additional forward model error term, $\mathbf{\Phi}$, is introduced, which, if taken into account, would give a theoretically optimal system, albeit one with a larger observation error term than would exist if the reconstructed radiances were modelled directly. The following equations show where this error term arises (equation 6.20 gives the equivalent optimal-in-**H** estimation equation). It will be assumed (for reasons that will become apparent in equation 6.30) that the full spectrum is reconstructed, even though in practice this makes the instrument error term uninvertible. The primes indicate quantities

defined for the suboptimal-in-**H** system.

$$\begin{aligned} \delta\tilde{\mathbf{y}}' &= (\tilde{\mathbf{y}} - \tilde{\mathbf{y}}_t) - (H(\mathbf{x}) - \tilde{\mathbf{y}}_t) \\ &= \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}(\mathbf{y} - \mathbf{y}_t) - (H(\mathbf{x}) - \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{y}_t)) \end{aligned} \tag{6.24}$$

and

$$\tilde{\mathbf{R}}' = \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{E}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}} + \mathbf{F}' \tag{6.25}$$

Let ϵ_{fm} be the forward model error, then:

$$H(\mathbf{x}) = \mathbf{y}_t + \epsilon_{fm} \tag{6.26}$$

Substitute this into the right-hand term in equation 6.24, which gives the forward modelled radiance departure from truth:

$$\begin{aligned} \delta\tilde{\mathbf{y}}'_{fm} &= H(\mathbf{x}) - \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{y}_t \\ &= \mathbf{y}_t + \epsilon_{fm} - \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{y}_t \\ &= (\mathbf{I} - \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}})\mathbf{y}_t + \epsilon_{fm} \end{aligned} \tag{6.27}$$

Using $< \cdots >$ to represent expectation over climatological conditions

$$\begin{aligned} \mathbf{F}' &= < \delta\tilde{\mathbf{y}}'_{fm}\, \delta\tilde{\mathbf{y}}'^{\mathrm{T}}_{fm} > \\ &= (\mathbf{I} - \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}) < \mathbf{y}_t\mathbf{y}_t^{\mathrm{T}} > (\mathbf{I} - \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}})^{\mathrm{T}} + \mathbf{F} \\ &= (\mathbf{I} - \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}) < \mathbf{y}_t\mathbf{y}_t^{\mathrm{T}} > (\mathbf{I} - \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}) + \mathbf{F} \\ &= \mathbf{\Phi} + \mathbf{F} \end{aligned} \tag{6.28}$$

The same overall error terms arise if it is assumed that reconstructed radiances are an imperfect representation of a true radiance spectrum, such that the true state is $\mathbf{y}_t$ rather than $\tilde{\mathbf{y}}_t$. This is shown in appendix D.

The important question for the assimilation system is, by how much is the error on the observation increased with the use of an incorrect forward model? The amount by which the error term differs given the use of a raw radiance forward model is given

by:

$$\begin{aligned} \Delta\tilde{\mathbf{R}} &= \tilde{\mathbf{R}}' - \tilde{\mathbf{R}} \\ &= (\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{E}\mathbf{L}_p^{\mathrm{T}}\mathbf{L}_p + \mathbf{F} + \mathbf{\Phi}) - (\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{E}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}} + \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{F}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}) \\ &= \mathbf{F} + \mathbf{\Phi} - \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{F}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}} \\ &= \mathbf{\Phi} + \Delta\mathbf{F} \end{aligned} \tag{6.29}$$

Equation 6.29 shows that the difference in error between the optimal-in-$\mathbf{H}$ and suboptimal-in-$\mathbf{H}$ cases comes from two sources:

- In the case of a proper reconstructed radiance forward model the error term $\mathbf{F}$ is actually filtered by $\mathbf{L}_p$ (equation 6.21), which has the effect of removing forward model error in directions that cannot be observed in reconstructed radiance space. That filtering does not take place if the incorrect forward model is used.
- The term $\mathbf{\Phi}$ must also be added to take account of the fact that the forward model does not correctly model reconstructed radiances.

$\mathbf{\Phi}$ arises because the artificial truncation process in $\mathbf{L}_p^{\mathrm{T}}$ means that a small portion of atmospheric signal is rejected in the discarded eigenvectors. This is done because, for the rejected components, the noise is high enough that the signal cannot be separated from the noise with any certainty and so the eigenvalues in the radiance climatology in these directions cannot be measured with low enough errors. In physical terms, measurements in these directions do not tell us anything new, and the noise on the measurement is great enough that there is a risk of adding noise into the analysis instead of information (there is a good discussion on these matters in Twomey, 1996). Effectively, therefore, it could be said that these atmospheric signals are also noise.

A particularly interesting aspect of $\Delta\tilde{\mathbf{R}}$ is that it is orthogonal to $\tilde{\mathbf{R}}$: this can be shown using the fact that the matrix $\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}$ is idempotent, which means that successive additional multiplication by $\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}$ does not alter the result. This is demonstrated by pre- and post-multiplying $\Delta\tilde{\mathbf{R}}$ by $\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}$, substituting $\mathbf{Y}$ for $< \mathbf{y}_t\mathbf{y}_t^{\mathrm{T}} >$ in equation 6.28 and combining with equation 6.21:

$$\begin{aligned}
\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\Delta\tilde{\mathbf{R}}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}} &= \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}} \quad (\mathbf{\Phi}+\mathbf{F}-\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{F}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}) \quad \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}} && (6.30)\\
&= \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}(\mathbf{I}-\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}})\mathbf{Y}(\mathbf{I}-\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}})\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\\
&\quad + \quad \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}(\mathbf{F}-\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{F}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}})\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\\
&= (\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}-\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}})\mathbf{Y}(\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}-\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}})\\
&\quad + \quad (\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{F}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}-\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{F}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}})\\
&= (\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}-\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}})\mathbf{Y}(\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}-\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}})\\
&\quad + \quad (\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{F}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}-\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{F}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}})\\
&= 0+0 && (6.31)
\end{aligned}$$

This describes mathematically what can be understood logically: $\Delta\tilde{\mathbf{R}}$ arises because of the rejected principal components, which are, by definition, orthogonal to the retained components. What is peculiar about this is that it means an increment can be made in a direction that is not observable, purely because the forward model is non-zero in that direction. This increment does not contain any information, because there is no observation of it. Essentially, the increment is just adding noise to the solution, and $\Delta\tilde{\mathbf{R}}$ defines its statistical distribution.

An assessment of the magnitude of $\mathbf{\Phi}$ would allow us to infer the quantity of atmospheric signal discarded by the truncation process that can be forward modelled (if it cannot be forward modelled, this error would be part of $\Delta\mathbf{F}$). In the case that all PC scores are retained, there will be no discarded signal, and $\mathbf{\Phi}$ will be zero, but the minor directions will be observed with high noise. In a very poorly defined set of PCs, $\mathbf{\Phi}$ will be large, and much atmospheric signal will be discarded, but those directions which are observed will be observed with low noise. In a well-constructed PC set, the vast majority of atmospheric signal is preserved with confidence, and this should yield a small $\mathbf{\Phi}$.

$\Delta\tilde{\mathbf{R}}$ is a theoretical construct. There are various sources of error that contribute to the $\mathbf{R}$ matrix that are described in section 7.1, not all of which are well known. It therefore becomes necessary to use diagnosed observation error covariance matrices (see section 7.2) rather than evaluating them explicitly. There is also no need to explicitly evaluate $\mathbf{\Phi}$ or $\Delta\mathbf{F}$: they will automatically be part of the total diagnosed observation error covariance. However, in simulation experiments where we know the error terms applied to the raw radiances exactly, neglecting $\Delta\tilde{\mathbf{R}}$ will underestimate the observation error: this system is referred to as suboptimal-in-$\mathbf{H}$-and-$\mathbf{R}$.

$\Delta\tilde{\mathbf{R}}$ may be important for channel selection. Because it arises through the neglect

of terms that are orthogonal to the retained atmospheric signal, this additional error is also orthogonal to the instrument noise covariance matrix, $\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{E}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}$. In theory, therefore, it should raise the rank of the $\tilde{\mathbf{R}}$ matrix, and paradoxically, this might allow one to reconstruct more channels than PC scores: although $\tilde{\mathbf{y}}$ for the additional channels is merely a linear combination of the first p channels, $\delta\tilde{\mathbf{y}}'$ is not a linear combination because of the involvement of the forward model. This rather strange theoretical outcome will be returned to later in section 7.2.2 where the rank of diagnosed observation error covariance matrices is examined.

6.4 Summary

The theory for PC compression and reconstruction of the IASI spectrum has been presented, and the theoretical equivalence of the optimal assimilation of PCs and reconstructed radiances demonstrated under the assumption of a mathematically well-conditioned reconstruction matrix. The importance of the rank of the reconstructed radiance observation error covariance matrix was emphasised along with the impact this has on the channel selection for reconstruced radiances.

Practical implementation of reconstructed radiances for assimilation is likely to make use of raw radiance forward model Jacobians, and an extra error term, $\Delta\tilde{\mathbf{R}}$ is demonstrated. $\Delta\tilde{\mathbf{R}}$ is orthogonal to the instrument noise covariance matrix for reconstructed radiances, and therefore has implications for the rank of the problem. Neglect of this term renders the assimilation suboptimal-in-**H**-and-**R**. In practice, however, the necessary use of a diagnosed error covariance matrix that reflects all sources of error including $\Delta\tilde{\mathbf{R}}$ reduces the requirement to have an accurate estimate of this quantity.

The next chapter presents a method for selecting channels for reconstructed radiances, and shows how a diagnosed observation error covariance matrix can be constructed for reconstructed radiances. The diagnosed observation error matrix and an instrument noise covariance matrix for reconstructed radiances are then used to generate several candidate channel selections. Chapter 8 examines the performance of reconstructed radiances via linear analysis, and chapter 9 tests reconstructed radiances and principal components in the 1D-Var.

Chapter 7

A channel selection for Reconstructed Radiances

As shown in chapter 6, reconstructed radiances have a different instrument error covariance from raw radiances, as the PC compression process introduces significant levels of correlation between channels and imposes mathematical constraints on the $\mathbf{R}$ matrix. Because of this, channel selections in common use for raw radiances such as the *Collard* selection (Collard, 2007, also see appendix E) are not optimal for reconstructed radiances. This chapter follows the convention for this thesis that italics will be used to indicate the name of a channel selection.

As stated in section 6.1, the number of channels chosen for assimilation should not be greater than the number of retained PCs, p, or the error covariance matrix for the reconstructed radiances will be non-positive definite. Even p reconstructed radiances are not guaranteed to be independent: for example, the PCs used by EUMETSAT are designed to contain as much information as possible from the given spectral range without loss. It is preferable to use a channel selection tailored to the application, and for NWP this would include avoiding channels sensitive to trace gases about which we have no prior information. Some of the PCs, however, will contain information about the variability of such species.

Say that p_{nwp} eigenvectors contain information relevant to NWP, and $p_{chem} = p - p_{nwp}$ mainly represent the signals from the many minor chemical species present in the atmosphere. If p reconstructed radiance channels were chosen avoiding parts of the spectrum with trace gas sensitivity, it would not be possible to extract the information on those gases contained within the PC scores with an acceptable degree of error. However, the p channels would contain information from only p_{nwp} PC scores, and so $p - pnwp$ of the reconstructed radiances would be linear combinations of the other channels. If p reconstructed radiances are chosen for this tailored channel selection, the resultant observation error covariance matrix will probably be singular, and at best ill-conditioned (in other words the ratio of its largest to smallest eigenvalues will be too high) and will produce mathematically unstable results in some directions.

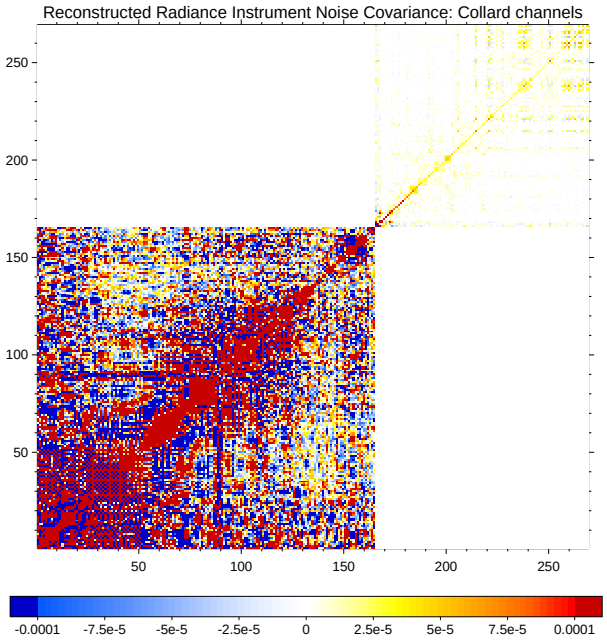


Figure 7.1: Reconstructed radiance instrument noise covariance matrix ($\mathbf{L}_s\mathbf{L}_s^{\mathsf{T}}$) for the *Collard* channel selection. The lower left block is the band 1 channels and the upper right block is the band 2 channels. Note the high degree of correlation within a band, but lack of correlation between bands because the compression is performed for each band separately.

Despite these mathematical considerations, assimilation experiments with IASI reconstructed radiances described in Collard et al. (2010) did use the *Collard* selection. The *Collard* selection was generated for raw radiances, where there are no constraints on the rank of the error covariance matrix (at least in terms of instrument noise), and it contains too many channels in Band 1 for correct use of reconstructed radiances. The channels were also chosen using a simplified diagonal R matrix, but significant off-diagonal components are present for reconstructed radiances that may affect the usefulness of the channels selected. The reconstructed radiance instrument noise covariance matrix for the *Collard* selection is shown in figure 7.1. Note the high degrees of correlation between channels in band 1 and recall that this matrix is not positive definite.

Of course, there are significant off-diagonal components of the full error covariance matrix for raw radiances also (e.g. Bormann et al., 2010; Garand et al., 2007), which continue to be ignored in many assimilation and retrieval schemes but which were implemented operationally at the Met Office in January 2013 during this thesis study (Weston et al., 2014). As shown by Hilton and Collard (2009) and Collard et al. (2010), it is quite possible to assimilate reconstructed radiances with the same

subset of the *Collard* selection that is used operationally, using the same (incorrect, diagonal) assumptions about the error covariance statistics and achieve a reasonable result. However, the benefits of the noise-reduction inherent in the reconstructed radiances will not be seen, and the observations cannot be assimilated with a realistic observation error covariance.

7.1 New channel selection methods for reconstructed radiances

The aim of this work is to propose a better way of using reconstructed radiances, enabling the assimilation scheme to extract more information from the observations. So how should the channels for reconstruction and assimilation be chosen? One important constraint on the choice is that the retrieval process will have better numerical stability if the resulting error covariance matrix has a low condition number. For this reason, methods such as those of Collard (2007); Ventress and Dudhia (2013); Migliorini et al. (2014) are unsuitable for this application because they do not restrict the final selection of the channels using the interchannel correlations, even where channel correlations form part of the information content measure. Two new approaches to performing a channel selection have been tested, that are described in the following sections.

7.1.1 Method 1

This is a purely mathematical approach developed by Tim Hultberg at EUMETSAT and described in Hultberg and August (2013). It selects channels by first computing the dot product of each row of the $\mathbf{L}_p$ matrix with itself, to give a norm for each row. The row (channel) with the largest norm is selected. The dot product of each remaining row (channel) with the selected row is then computed, and this is used to subtract the contribution of the chosen channel from each of the remaining channels before the next channel is chosen; in other words, a channel which is highly correlated with the previous channel will now have only a small extra contribution to make to the eigenvector subspace spanned by the matrix $\mathbf{L}_p$ and is unlikely to be chosen. The next channel chosen has the largest contribution to the subspace of $\mathbf{L}_p$ that remains after selection of the first channel. The process is repeated until all p channels have been selected. The channel selection is performed for each band independently, as there are no interband correlations in $\mathbf{L}_p$. The resulting channel selection from this

method is used in Version 6 of the EUMETSAT Level 2 temperature and water vapour retrieval products.

Method 1 is likely to result in a mathematically robust channel selection, with a reasonably well-conditioned instrument noise covariance: in its basic form, it is not tailored to any specific application, and we might reasonably expect the channel selection to approach the full information content of the PC scores because the process picks the most independent p channels. However, without taking into account other sources of error that are components of the $\mathbf{R}$ matrix, such as forward model error, it risks choosing channels that have little practical use for NWP. The chosen channels may, for example, have a high degree of trace gas contamination, or a large forward model error. For this thesis, the method was therefore modified to minimise this risk by taking account of the full observation error term. Rather than selecting the row of $\mathbf{L}_p$ with the largest norm, instead the row of $\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\tilde{\mathbf{R}}^{1/2}$ with the largest dot product norm was selected, and channels were chosen from the two bands simultaneously because of long-range error correlations. The modified method was used by Tim Hultberg to generate channel selections from several $\tilde{\mathbf{R}}$ matrices.

The resultant channel selections were tested to determine their sensitivity to the observation error assumptions. Each of the channel selections were very similar, indicating that the selection process is dominated by the structure of $\mathbf{L}_p$.

7.1.2 Method 2

This is a new adaptation of a physically based approach, where the impact of each channel on the assimilation/retrieval system is considered. It is similar to that employed by Collard (2007) to choose the channels for restricted bandwidth IASI dissemination, which in practice is also the basis for the channel selection assimilated at most European weather centres, but uses the full observation error covariance matrix, and additionally penalises the selection of a channel that raises the condition number of the resultant error covariance matrix by more than a chosen threshold. The details of the selection method will be given in section 7.3.2.

The second method requires the error covariance of the a priori/background profile and as such the results will be tailored to a particular assimilation or retrieval system. The resulting selection may provide more information to that system than one chosen by Method 1, but it will not approach the full information content of the PC scores (including trace gas information). It is expected that if p channels are reconstructed, the resultant observation error covariance, $\tilde{\mathbf{R}}$, will be rank deficient, and that the number of usable reconstructed channels will be somewhat less than p.

Various additions and modifications to the basic Method 2 have been tested, to determine the robustness of the solution. Several of the resulting channel selections are compared in the next chapter in terms of information content to determine which ones merit further study.

7.1.3 Incremental channel selection methods

An incremental approach to channel selection, such as both Methods 1 and 2 employ, is imposed upon us by the computational cost of testing all possible combinations of 210 channels chosen from 5116 (the total number of channels in bands 1 and 2). One issue with the incremental approach is that the exact choice of each channel determines the progress of the selection from that point on; any tiny difference in norm or DFS that may not be meaningful, or even the selection of the first of a few channels with identical norm or DFS, will potentially result in a completely different channel selection, albeit one with very similar information content overall. Therefore, the method will produce an appropriate channel selection, but cannot produce a single 'best' selection.

This chapter explores the two methods and possible modifications to the methods that may generate a more robust channel selection. Other techniques and modifications are possible, that may genereate equally valid selections as those presented here.

7.2 Estimating an observation error covariance matrix

Method 2, and the modified Method 1, require an observation error covariance term. The observation error covariance matrix should ideally contain contributions from:

- **instrument noise**: This term takes into account apodisation functions and the reconstruction process as outlined above.
- **forward model error**: Although forward model error is usually more correctly handled as a systematic error to be removed by bias correction, remaining errors can be quite complex and are usually treated as a random error.
- **reconstruction error**: This error term is defined in section 6.3, in the case that a raw radiance forward model is used.
- **errors of representation**: An additional source of mismatch between the observation and its forward-modelled counterpart results from the finite grid

of the model in question. NWP models are unable to capture detailed structures within a grid box that the observation is sensitive to. In single column retrievals this term may be considered part of the background error, but in NWP assmilation the term is generally regarded as an observation error.

It is not easy to construct such a matrix from its individual component terms as, other than the instrument noise term, they are not well known. An alternative approach is therefore to estimate the quantity by examination of observation minus forward-modelled radiance statistics. This section describes the method by which an estimated error covariance matrix for bands 1 and 2 of IASI, to be used in Method 2 and the modified Method 1, to generate a channel selection of reconstructed radiances for use in NWP. (Band 3 is not used at present as the channels are sensitive to physical processes such as solar effects that are hard to forward model and that limit the use of the channels in operational assimilation systems.)

Covariance estimation is currently used at the Met Office, ECWMF and other national meteorological centres to define the **R** matrix used during assimilation for various satellite data types. At most centres, although a full matrix may be diagnosed, only the diagonal part is retained in the assimilation at the present time, although this is an area of active research for many organisations. There are various methods for estimating observation error covariances within NWP systems. The most well known are those of Hollingsworth and Lönnberg (1986), using observation minus background (O–B) calculation statistics (background departures), and of Desroziers et al. (2005). The latter is currently more widely used (e.g. at the Met Office, ECMWF and Météo-France), but this technique requires observation minus analysis (O–A) statistics, and assumes that all channels for which errors are estimated are assimilated: we cannot assimilate 8461 IASI channels even in tests, so must rely on the background departures only and are thus restricted to the first method. Nevertheless, the two methods have been shown to give similar results by Bormann et al. (2010).

7.2.1 The observational method of Hollingsworth and Lönnberg

The Hollingsworth and Lönnberg (1986) method estimates both observation and background errors from background departure statistics. Also known as the observational method, this has been used by, among others, Garand et al. (2007), Bormann and Bauer (2010) and Bormann et al. (2010) to calculate observation error matrices for operational satellite sounders. The observational method is summarised by Bouttier and Courtier (1999) who show that background departure covariance,

$c(p1, p2)$, between two observations at locations $p1$ and $p2$ is equal to:

$$c(p1, p2) = \mathbf{R}_{p1,p2} + \mathbf{H}_{p1}\mathbf{B}\mathbf{H}_{p2}^{\mathrm{T}} \tag{7.1}$$

The observational method makes the assumptions that

- observation errors are random and uncorrelated between observations
- background errors vary only slowly with increasing separation between observation errors such that the error at zero separation can be extrapolated from the error at small separation
- observation and background errors are uncorrelated with each other

Under these assumptions, if $p1 = p2$, the error covariance is the sum of observation and background error terms in observation space. If $p1 \neq p2$, then $c(p1, p2)$ consists only of errors that are correlated between observations, which can be attributed to the background since the observation error is assumed to be spatially uncorrelated. Subtracting the matrix defined at $p1 \neq p2$ from the matrix at $p1 = p2$ therefore yields the observation error covariance matrix.

The partitioning between background and observation error is not exact. Any structure that is correlated across grid boxes would be treated as background error by this process. Errors of representation are likely to be included within the observation error covariance matrix, as the structures seen by each observation that are not observable to the NWP model are likely to be different for each footprint. However, any errors of representation or radiative transfer errors that are correlated with the atmospheric state, and therefore similar for adjacent observations, may be subtracted. This is also the case for any observation errors that are correlated between adjacent observations. Conversely, background errors that are uncorrelated between adjacent observations will be mapped into the observation error term.

The following alternative notation was used by Cameron and Eyre (2008) to describe the observational method as applied to radiance observations. Background departures are calculated for pairs of observations at $p1$ and $p2$. If $\Delta_{i,p1}$ is the background departure for channel i of an observation at $p1$, and $\Delta_{j,p2}$ is the background departure for channel j of an observation at $p2$:

$$\mathbf{R}_{i,j} = \mathrm{Cov}(\Delta_{i,p1}, \Delta_{j,p2})|_{p1=p2} - \mathrm{Cov}(\Delta_{i,p1}, \Delta_{j,p2})|_{p1\neq p2} \tag{7.2}$$

The same result can be arrived at by an alternative calculation

$$\mathbf{R}_{i,j} = \tfrac{1}{2}\,\mathrm{Cov}(\Delta_{i,p1} - \Delta_{i,p2}, \Delta_{j,p1} - \Delta_{j,p2}) \tag{7.3}$$

We shall refer to this alternative formulation as the double difference covariance. The equivalence to the classical representation of the observational method can be shown using the covariance identity:

$$\mathrm{Cov}(A-B, C-D) = \mathrm{Cov}(A,C) - \mathrm{Cov}(A,D) - \mathrm{Cov}(B,C) + \mathrm{Cov}(B,D) \quad (7.4)$$

to expand equation 7.2:

$$\begin{aligned} &\tfrac{1}{2}\,\mathrm{Cov}(\Delta_{i,p1} - \Delta_{i,p2}, \Delta_{j,p1} - \Delta_{j,p2}) \\ &\qquad = \tfrac{1}{2}\,\mathrm{Cov}(\Delta_{i,p1}, \Delta_{j,p1}) + \tfrac{1}{2}\,\mathrm{Cov}(\Delta_{i,p2}, \Delta_{j,p2}) - \\ &\qquad\quad \tfrac{1}{2}\,\mathrm{Cov}(\Delta_{i,p1}, \Delta_{j,p2}) - \tfrac{1}{2}\,\mathrm{Cov}(\Delta_{i,p2}, \Delta_{j,p1}) \qquad (7.5) \\ &\qquad = \mathrm{Cov}(\Delta_{i,p1}, \Delta_{j,p2})|_{p1=p2} - \mathrm{Cov}(\Delta_{i,p1}, \Delta_{j,p2})|_{p1\neq p2} \qquad (7.6) \end{aligned}$$

In practice, for satellite data, it is very unlikely that there would be two independent observations in the same location, i.e. at zero separation, so some method of estimating the background error at zero separation is required. The method proposed by Hollingsworth and Lönnberg (1986) and used by Garand et al. (2007) is to calculate statistics in bins of increasing separation distance, and then extrapolate the trend from larger separations (assumed to be spatially correlated background error) to zero separation to find the background error component of the covariance at zero separation. However, the choice of extrapolation function is somewhat arbitrary, and so here, in common with Bormann et al. (2010) and Cameron and Eyre (2008), the covariance statistics at the smallest possible separation distance are used as an estimate of $\mathbf{B}$. Both of these previous studies used a 25 km separation distance, which is the smallest practical distance within an NWP model where IASI observations are thinned to one footprint in four prior to processing.

The final matrix, $\mathbf{R}_{HL}$, was constructed from observations at 25 km separation, with background departures calculated within the Met Office operational 1D-Var preprocessor using RTTOV-9 (Saunders et al., 2010) as the forward model. The observations were used if deemed to be cloud-free on the basis of various cloud detection tests (Hilton et al., 2009a). Approximately 16 000 observation pairs were used, without bias correction.

As mentioned above, background errors that are uncorrelated between the paired observations could be mapped into the resulting matrix: given that the observational method relies on the subtraction of the errors at zero separation which has not been done here, this remains a distinct possibility. Figure 7.2 compares the diagonal of

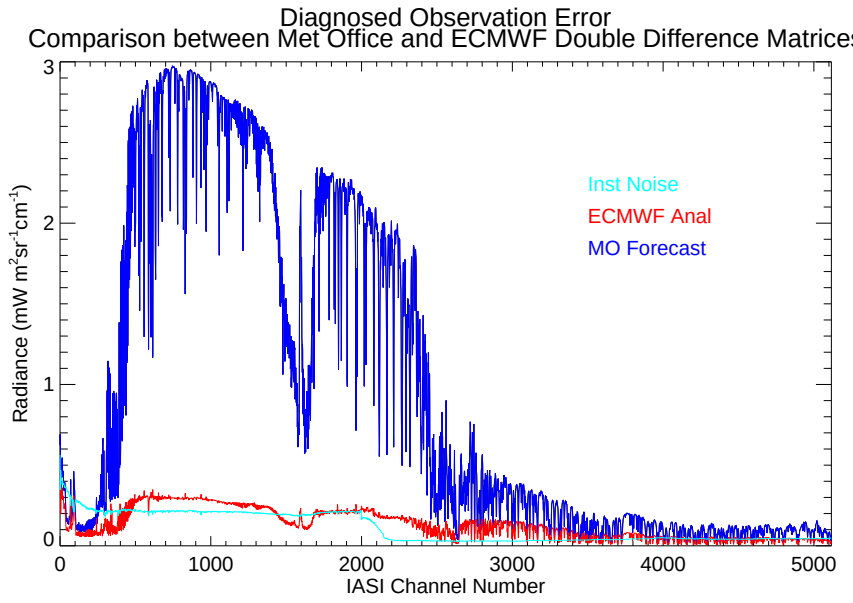


Figure 7.2: Standard deviation of observation error from the $\mathbf{R}_{HL}$ double difference matrix – comparison against a double difference matrix calculated from observations against the ECMWF analysis, $\mathbf{R}_{EC}$.

the $\mathbf{R}_{HL}$ matrix against a double difference matrix, $\mathbf{R}_{EC}$, calculated from statistics of IASI observations minus ECMWF analysis (Tim Hultberg, pers. comm.). As well as using ECMWF analyses, $\mathbf{R}_{EC}$ differs from $\mathbf{R}_{HL}$ in that it was constructed from pairs of pixels 3 and 4 (at 12.5 km separation) rather than random pairs at 25 km. Another difference is in quality control: $\mathbf{R}_{EC}$ was calculated from a dataset covering only $\pm 60°$ latitude and it is probable that the discarding of outlier observations (i.e. when background departures are too high) was more stringent for these matrices; for $\mathbf{R}_{HL}$ observations were only rejected on gross differences, and on modest departures for selected window channels as part of the cloud detection process.

$\mathbf{R}_{HL}$ has a considerably higher error in the window region in particular. Since forecast background errors are larger than analysis errors, this could suggest problems with partitioning the background departure information between observation and background error terms: it is likely that some background error is incorporated into $\mathbf{R}_{HL}$. On the other hand, the ECMWF analysis has assimilated information from the IASI observations already, so one might expect the simulated radiance to be tuned to the observations, thereby underestimating the observation error. The shape of the plotted diagonal seems to show realistic physical structures: channels with lower error are scattered through the window region, that resemble absorption lines. Close to the centre of an absorption line, a channel is less sensitive to the surface temperature than a true window channel would be. This suggests that it is

indeed the surface temperature that is causing the very high errors in the window region for $\mathbf{R}_{HL}$.

Although the two matrices are rather different, casting some doubt on the effectiveness of the method to generate a good observation error covariance matrix, the matrix $\mathbf{R}_{HL}$ is only for use in channel selection. For operational assimilation, once the channel set has been determined, the Desroziers technique would be applied to refine the $\mathbf{R}$ matrix prior to pre-operational trialling. The $\mathbf{R}_{HL}$ matrix is used to perform channel selections with Method 2 and modified Method 1 in section 7.3.

7.2.2 Mathematical stability of observation error covariances for reconstructed radiances

The eigenvalues of estimated covariance matrices are likely to have much too large a spread. There is a large body of literature in statistics – and in particular its application to finance and biostatistics – pertaining to covariance estimation (e.g. Daniels and Kass, 2001; Ledoit and Wolf, 2004, 2012; Wang et al., 2012). It is usual to apply what are known as shrinkage operations on estimated covariance eigenvalues, usually combining the estimated matrix with a diagonal target matrix to reduce the largest eigenvalues, and increase the smallest. The resultant matrix will have more stable mathematical properties, and gives better results. This performance advantage of the shrunk matrix stems from work by Charles Stein on the estimation of sample means. A very nice summary of the effect of shrinkage is provided by Efron and Morris (1977).

Twomey (1996) describes the effect of very small eigenvalues of a covariance matrix: in general, these matrices are used in their inverse form, and the very small eigenvalues become very large values in the inverse matrix: the analysis in these directions is very unstable to small numerical differences. In fact, it may be better to estimate $\mathbf{R}^{-1}$ directly, and perform the shrinkage operations on this instead (Wang et al., 2012). No shrinkage has been performed on the estimated matrices used for channel selection. However, such processes would be required on the $\mathbf{R}$ matrix for the selected channels before operational assimilation, and a form of shrinkage is investigated in chapter 9 in the context of suboptimal-in-$\mathbf{H}$-and-$\mathbf{R}$ assimilation.

As discussed in section 6.1, the instrument noise covariance matrix for reconstructed radiances will have a rank equal to p. Section 6.3 discussed the extra forward model error term, $\Delta\tilde{\mathbf{R}}$ that would result from using a raw radiance forward model (as has been used here), which is orthogonal to the instrument noise covariance matrix and may therefore increase the rank of the resultant $\tilde{\mathbf{R}}$ matrix. It is therefore

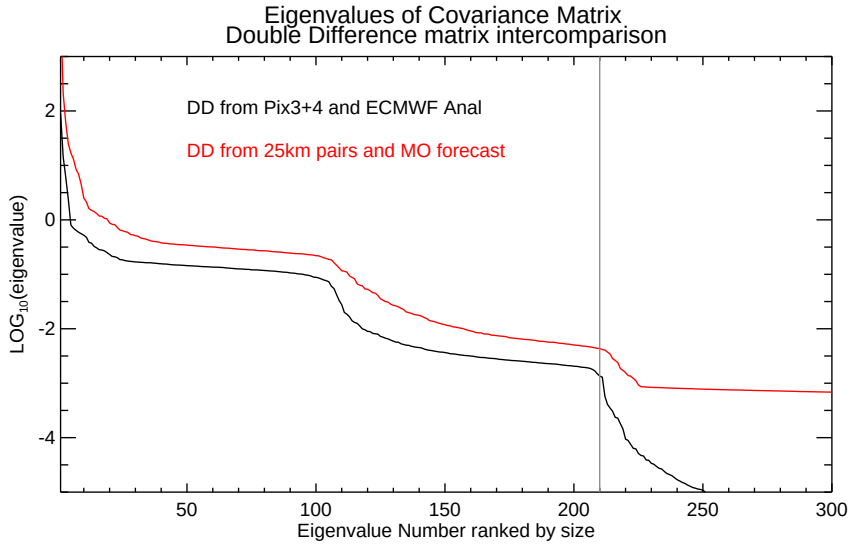


Figure 7.3: Eigenvalues of the covariance matrices calculated from different background profiles as detailed on the plots. Marked in light grey is the rank of the instrument noise covariance matrix.

interesting to look at the eigenvalues of the diagnosed observation error matrices.

Figure 7.3 shows the eigenvalues of the $\mathbf{R}_{HL}$ and $\mathbf{R}_{EC}$ double difference matrices. For an instrument noise only covariance matrix, we would expect the eigenvalues to drop to zero after the rank of the matrix had been reached, in this case after 210 values. In the case of the diagnosed matrices, the eigenvalues do drop off beyond this point, but not so sharply. The matrix constructed using the ECMWF analysis, $\mathbf{R}_{EC}$, has slightly lower eigenvalues for each ranked position for the higher order eigenvectors. This is a possible consequence of mapping more background error into $\mathbf{R}_{HL}$, and the increased error across the window region.

There is a step in the profile of eigenvalues around 100–110. This is interesting because the PC compression of bands 1 and 2 is done separately, so the instrument noise is independent between the two bands. 90 PCs are retained for Band 1 and 120 for Band 2. The step is likely to be related to this separation of the compression between the two bands, but its precise location corresponds precisely to neither of the numbers of retained PCs in Band 1 or 2.

As stated earlier in this section, what really matters in terms of mathematical stability is the condition number – the ratio of the largest to smallest eigenvalues. A matrix with too large a condition number is known as ill-conditioned and will be numerically unstable: the inverse matrix, which is the quantity required by most retrieval or assimilation algorithms, will be poorly defined leading to a solution that is also poorly defined. A maximum condition number of approximately 1×10^6 is a

rule of thumb often used to determine whether the matrix is invertible (although for operational applications higher stability may be required). Figure 7.4 shows the eigenvalues of the double difference matrices from the different background profiles as a fraction of the leading eigenvalue in the respective matrix (this is the inverse of the condition number). The grey horizontal line corresponds to a condition number of 1×10^6. If more channels are reconstructed than the rank indicated where the plotted logarithmic eigenvalue ratio falls below -5, the inversion of the matrix will likely be unstable. Figure 7.4(a) shows the ratios for the full matrix combining bands 1 and 2. The $\mathbf{R}_{HL}$ matrix is much more poorly conditioned for any given number of eigenvalues than the $\mathbf{R}_{EC}$ matrix, and the eigenvalue curve drops below the grey line by only 110 channels. Both matrices behave similarly in terms of the matrix size at which the eigenvalues drop successively. The $\mathbf{R}_{HL}$ eigenvalues asymptote sharply to a ratio of 1×10^7 at about 225 values: this is likely to be related to the low precision with which the values were written to file prior to plotting.

Figure 7.4(b) shows the eigenvalue ratios for the band 1 only covariance matrix. The drop beyond eigenvalue 90 is sharp for all the matrices, although there is some lag for $\mathbf{R}_{HL}$. This gives confidence that there is little error orthogonal to the instrument error contained within the matrices, and that the extra reconstruction error, $\mathbf{\Phi}$, is indeed small, and thus a raw radiance forward model is adequate for simulation of reconstructed radiances for Band 1. Figure 7.4(c) shows the eigenvalue ratios for a band 2 only covariance matrix. For $\mathbf{R}_{EC}$, the results are similar to those for band 1, except that the drop off after 120 (the number of retained PCs for band 2) is not quite so sharp. Forward model errors tend to be larger relative to the instrument noise for water vapour sensitive regions of the spectrum and it is possible that this is why there is a larger amount of error orthogonal to the instrument noise in the band 2 matrix in comparison with the band 1 matrix. The slower tailing off of the significant eigenvalues could imply that $\mathbf{\Phi}$ is non-zero for the water vapour band, but this could also be forward model error orthogonal to the instrument noise. For $\mathbf{R}_{HL}$, the eigenvalue ratio drops below the grey line at only 60 or so; this indicates significant error correlations for band 2 channels, likely arising through representativeness errors.

The band 1 matrix drops rapidly from the largest eigenvalue then flattens out before dropping rapidly again after p eigenvalues, but the band 2 matrix has an additional hump between 10 and 30 eigenvectors (this is less pronounced for $\mathbf{R}_{HL}$). The presence of this hump seems related to the step in the bands 1 and 2 combined matrix at around 100–110 (90 from band 1 plus 10–20 from Band 2), although its

Channel Selection	CO_2	Window	Ozone	Water Vapour	Band 1	Band 2	Total NChan
Lp-M1	39	66	10	95	90	120	210
RHL-M1	40	65	10	95	90	120	210
Collard	132	24	15	98	165	104	269

Table 7.1: Breakdown of the number of channels by principal sensitivity for the *Method 1* channel selections. Comparison with the *Collard* selection for bands 1 and 2.

cause is not known.

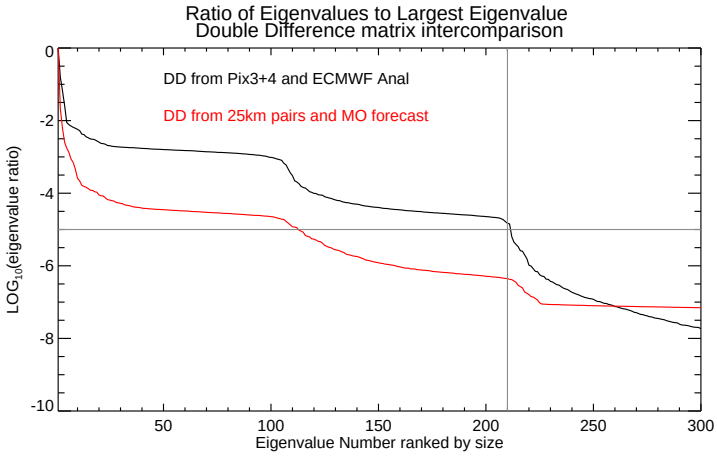
In summary, the diagnosed error covariance matrices have eigenvalue structures that reflect the fact they have been derived from reconstructed radiances. In other words, the eigenvalues drop rapidly once the number of retained PCs is reached. They do not drop immediately to zero; these matrices were estimated for a suboptimal-in-**H** system that contains additional forward model error orthogonal to the retained eigenvectors. Furthermore, they may contain background error that has not been removed by the double-difference formulation: this seems to be the case especially for the $\mathbf{R}_{HL}$ matrix in the window region. This is unfortunate, because it is likely to penalise the selection of window channels. The eigenvalue ratio plots suggest that it will not be possible to choose more than approximately 100 channels using $\mathbf{R}_{HL}$ if the resultant $\tilde{\mathbf{R}}'$ is to have a condition number of less than 1×10^6.

7.3 Channel selections

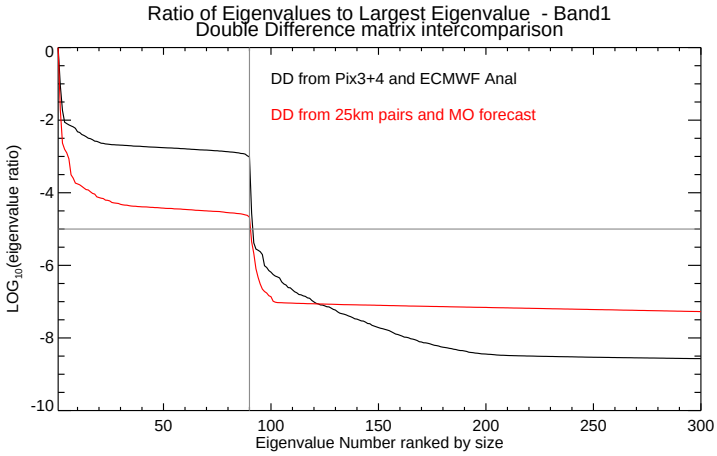
Methods 1 and 2 were used to derive channel selections of up to a maximum of 210 channels from IASI bands 1 and 2 (210 being the number of retained principal component scores and therefore the maximum number of independent reconstructed radiance channels).

7.3.1 Method 1 channel selections

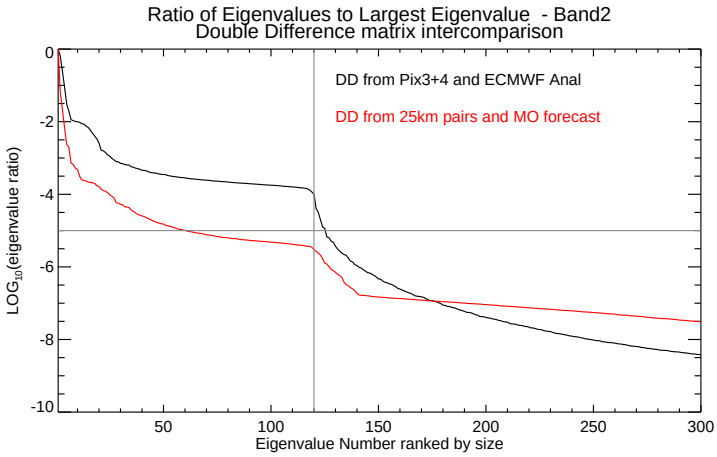
Two channel selections were generated using Method 1. The first is the EUMETSAT v6 selection, using rows of $\mathbf{L}_p$ alone to choose channels. It will be referred to as the *Lp-M1* selection. The second is the modified Method 1 selection, referred to as *RHL-M1*, which uses rows of $\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{R}_{HL}^{1/2}$. Table 7.1 shows the breakdown by principal sensitivity of the *Method 1* channel selections compared with the *Collard* selection for reference.



(a) Bands 1 and 2 combined



(b) Band 1 only



(c) Band 2 only

Figure 7.4: Ratio of eigenvalues of the covariance matrices to the leading eigenvalue. Comparison between double difference matrices with different background sources. Also indicated by a grey vertical line is the rank of the instrument noise covariance matrix. The grey horizontal line marks a condition number of 10^6.

The *Method 1* selections are very similar in terms of distribution of channels by principal sensitivity. The method generates a selection that has many fewer CO_2 channels, and many more window channels than the *Collard* selection. This is of concern for NWP applications: the main advantage of using reconstructed radiances rather than assimilating PC scores is that they should be more useful in cloudy areas in the common situation where the radiative transfer modelling of cloud effects, and underlying NWP background information that the RT model depends upon, are insufficiently accurate to allow a full cloudy assimilation. This is because, as discussed in section 6.2.1, each of the PCs is sensitive to the full atmospheric column, whereas the reconstructed radiances have more localised sensitivity, like normal radiances. The most common way to deal with cloudy scenes is to assimilate channels that peak above the cloud, and that are only weakly affected by the radiative signature of the cloud (e.g. Pavelin et al., 2008). Completely clear scenes comprise only a very small fraction of the total number of observations, say 5 %. If the selection contains too many window channels at the expense of channels that peak higher in the atmosphere, the majority of assimilated scenes will use very few channels indeed. In contrast, over half of the *Collard* selection for bands 1 and 2 is comprised of CO_2 channels used for temperature sounding of the troposphere and stratosphere.

7.3.2 Method 2 channel selections

The general principle of the method is to calculate, for each candidate channel, the change in a chosen measure of information content that would result from its inclusion, and to choose the channel with the greatest contribution. The chosen measure of information content here is DFS, according to the equation:

$$\mathrm{DFS} = \mathrm{Tr}(\mathbf{I} - \mathbf{A}\mathbf{B}^{-1}) \tag{7.7}$$

where

$$\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\tilde{\mathbf{R}}'^{-1}\mathbf{H})^{-1} \tag{7.8}$$

Method 2 differs from the Rodgers method, and from Collard's implementation of this method, in two important regards. Firstly, method 2 employs a full calculation of DFS and uses a full $\tilde{\mathbf{R}}'$ matrix. The Rodgers method was a simplification to allow efficient computation of incremental change in DFS in days when less computing power was available, and assumes a diagonal $\mathbf{R}$ matrix. It is now feasible to use a

non-diagonal matrix to perform a full calculation of DFS for each candidate channel, and this is in fact essential for the selection of reconstructed radiance channels where the interchannel error correlations are a fundamental feature even when there is no radiative transfer or representativeness error. Collard also used extensive blacklisting to rule out channels where the forward modelling was poor due to non-LTE[1] effects, known spectroscopy issues, and sensitivity to trace gases. A beneficial side-effect of using $\tilde{\mathbf{R}}'$ calculated from background departures should be that no blacklisting is required prior to channel selection, as channels that are poorly modelled will have high variances and are thus unlikely to be selected.

The second major difference is that the DFS measure is used in conjunction with the condition number of the candidate $\tilde{\mathbf{R}}'$ matrix. This method was constructed in the light of early research on the selection of channels specifically for reconstructed radiances by Collard (2012), where an attempt was made to penalise channels highly correlated with those already selected. In this work, it was found that the condition number of the resultant $\tilde{\mathbf{R}}$ matrix was a critical issue, hence Method 2 does not allow selection of channels where the condition number of the candidate $\tilde{\mathbf{R}}$ rises too high.

As with the Collard (2007) method, a collection of atmospheric profiles has been chosen over which to calculate a combined DFS. Collard used six standard atmospheres and two zenith angles. Several options have been investigated here, using atmospheric profiles chosen from the Met Office 70 level dataset used in chapter 10, but also including the US Standard Atmosphere. The baseline collection includes seven atmospheric columns from different parts of the globe plus the US Standard Atmosphere (eight profiles in total). Another collection then employed two zenith angles (0° and 40°) for each atmosphere, making a total of sixteen atmospheric columns.

Collard (2012) found that his reconstructed radiance channel selection had an uneven distribution of weighting function peaks, such that the upper tropospheric information was coming mostly from the tails of lower-peaking and higher-peaking channels. This led to concern that for cloudy situations, where channels peaking below cloud top are rejected, there would be few channels left to assimilate. Some tests were therefore carried out to try to encourage the selection of channels that would be useful under these circumstances. In addition to the baseline profiles, extra profiles were added where "clouds" were introduced at particular levels: Jacobians were zeroed below the cloud top, in order to ensure that channels were selected for

[1] local thermodynamic equilibrium

their information content above the cloud layer. Three different profile sets were used to take account of cloudy situations:

Cloudy A: In addition to the eight baseline atmospheres, three US Standard Atmosphere profiles were added where a "cloud" was introduced at 800 hPa, 600 hPa and 400 hPa. No attempt to forward model the cloud was made; a clear sky Jacobian was zeroed below each of the three given pressures. Eleven profiles in total.

Cloudy B: In addition to the eight baseline atmospheres, a cloudy US Standard Atmopsheric profile Jacobian was simulated with a grey cloud with a cloud fraction of 0.8 and a cloud top pressure of 500 hPa. The Jacobian was then zeroed below 500 hPa. Nine profiles in total.

Cloudy C: In addition to the eight baseline atmospheres, six cloudy US Standard Atmopsheric profile Jacobians were simulated with grey clouds of cloud fractions 0.2 and 0.8, and cloud top pressures of 800 hPa, 600 hPa and 400 hPa. The Jacobian was then zeroed below the cloud top pressure. Fourteen profiles in total.

For a given iteration of the channel selection, the set of all channels already chosen plus a candidate channel will be referred to as a candidate set in the explanation of the method. The channel selection proceeded as follows:

1. The $\tilde{\mathbf{R}}'$ matrix was generated for each candidate set. The eigenvalues of the candidate $\tilde{\mathbf{R}}'$ matrices were then used to calculate the condition number for each candidate set.
2. The candidate set with the lowest condition number was identified.
3. Candidate sets with an $\tilde{\mathbf{R}}'$ matrix whose condition number was greater than a factor f times the lowest condition number were rejected.
4. The combined DFS over all of the atmospheric profiles in a given collection was calculated for each remaining candidate set, using the **MoistCov B** matrix for atmospheric temperature and humidity. Note that no surface variables were included in the retrieval, in keeping with the general experimental set-up in this thesis (section 3.8.3).
5. The candidate channel set with the highest DFS was then selected to incorporate the most information content rich channel for this iteration to the final selection.
6. Channels were chosen until the resulting $\tilde{\mathbf{R}}'$ became non-positive definite (i.e. a maximum of 210 channels). Channels were chosen from within bands 1 and 2 simultaneously.

Several condition number factors f were tested: a factor of $f = 1.0$ results in the

selection of the channel giving the best condition number, in other words the DFS calculation is ignored. A factor of $f = 100.0$ means that very little weight is given to the condition number of the resulting $\tilde{\mathbf{R}}'$ matrix. A factor of $f = 1.3$ was chosen to provide a reasonable balance between the DFS and condition number.

One further option was tested: Collard (2007) found that the Rodgers selection method preferentially chose channels from band 2. These channels tend to have sharply peaking tropospheric temperature Jacobians and the channels seemed to be selected as much for the temperature information they carry as for their water vapour sensitivity. Without channels insensitive to errors in the background water vapour profile, the resulting channel selection is unable to unravel errors in temperature from errors in water vapour. Collard's solution to this was to allow selection only from band 1 in the first instance, and then to allow selection from either band 1 or band 2. A new approach was taken here. Rather than blacklisting, the selection of channels sensitive to water vapour errors in the NWP forecast was penalised by adding an extra error term to the observation error covariance matrix projecting the background forecast error for water vapour into observation space with the Jacobian:

$$\tilde{\mathbf{R}}'_{wv} = \tilde{\mathbf{R}}' + \mathbf{H}_{wv}\mathbf{B}_{wv}\mathbf{H}_{wv}^{\mathrm{T}} \tag{7.9}$$

The subscript wv indicates that only the water vapour part of the background error is added to the observation error term. Up to 100 channels (being approximately half the total available) were chosen using $\tilde{\mathbf{R}}'_{wv}$ before reverting to $\tilde{\mathbf{R}}'$ for selection of the remaining channels. The use of the water vapour channel penalisation, and of different condition number thresholds, were investigated using the baseline profile set.

Table 7.2 defines a naming convention for each of the Method 2 channel selections made, which can be cross-referenced to the channel lists themselves in appendix E, and which will be used in subsequent tables and text. Table 7.3 provides a breakdown of the channels in each selection by principal sensitivity. The first notable result is that only in the case where condition number alone was used to perform the channel selection (condition number factor of 1.0; *RHL-M2-8J-1.0*) were 210 channels selected. This result reflects what was expected: if channels are chosen on the basis of information content for a given application, the channel selection is unable to represent the full set of PC scores. This should not be considered a negative feature of the channel selections: in order to perform acceptable compression of every channel in the spectrum, some of the principal components must necessarily represent spectral features that are of no interest to NWP.

Channel Selection Name	**Jacobians**	**Condition Number Factor**	**Water Vapour Error**
RHL-M2-8J-1.0	8 Clear	1.0	
RHL-M2-8J-1.3	8 Clear	1.3	
RHL-M2-8J-100.0	8 Clear	100.0	
RHL-M2-8J2Z-1.3	8 Clear 2 Zen	1.3	
RHL-M2-11J-1.3	11 Cloudy A	1.3	
RHL-M2-9J-1.3	9 Cloudy B	1.3	
RHL-M2-14J-1.3	14 Cloudy C	1.3	
RHL-M2-8J-1.3-WVE	8 Clear	1.3	YES

Table 7.2: Naming convention for each of the Method 2 channel selections that will be used throughout the text with their corresponding settings in the channel selection code. The name takes the form *RmatrixName-Method-JacobianSet-Factor-WVE* where WVE will be added if penalisation of selection of water vapour channels was carried out.

Channel Selection	**CO_2**	**Window**	**Ozone**	**Water Vapour**	**Band 1**	**Band 2**	**Total NChan**
RHL-M2-8J-1.0	38	56	14	102	90	120	210
RHL-M2-8J-1.3	132	3	1	50	135	51	186
RHL-M2-8J-100.0	14	0	2	82	16	82	98
RHL-M2-8J2Z-1.3	109	13	5	51	127	51	178
RHL-M2-11J-1.3	38	0	2	144	40	144	184
RHL-M2-9J-1.3	44	1	4	159	49	159	208
RHL-M2-14J-1.3	32	0	1	136	33	136	169
RHL-M2-8J-1.3-WVE	143	2	2	36	147	36	183
Collard	132	24	15	98	165	104	269

Table 7.3: Breakdown of the number of channels in each selection by principal sensitivity for the *Method 2* channel selections and comparison with the *Collard* selection for bands 1 and 2.

The next result of interest is that many of the channel selections contain more than 90 channels from band 1. This indicates that errors orthogonal to the instrument noise are included in $\tilde{\mathbf{R}}'$, and that correlations between bands 1 and 2 are being exploited in the channel selection. On the other hand, the preferential selection of water vapour channels appears to occur with the cloudy atmospheric collections (*RHL-M2-11J-1.3*, *RHL-M2-9J-1.3*, *RHL-M2-14J-1.3*), unless channels sensitive to water vapour error are penalised (not shown).

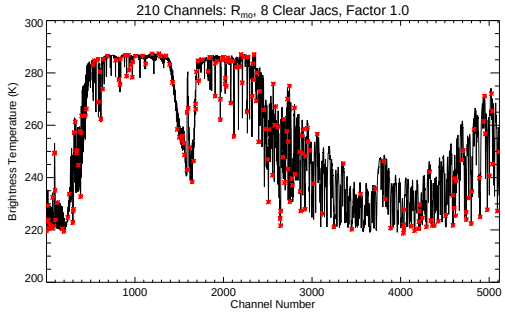
The choice of channels sensitive to ozone can be seen to be effectively penalised by the use of the DFS for NWP to select the channels: where DFS is ignored (condition number factor of 1.0; *RHL-M2-8J-1.0*) 14 ozone-sensitive channels are chosen, but when it is used (all other selections) fewer than five channels are selected in this part of the spectrum.

One of the aims of the method, particularly by incorporating the cloudy profiles, was to ensure that for cloudy conditions a reasonable number of channels remain usable. It is therefore encouraging that some of the selections contain significantly more longwave CO_2 channels than the *Method 1* channel selections. Conversely, of some concern is that, with the exception of the condition number only selection, these *Method 2* channel selections contain very few channels in the window region, and often no true window channels, which would hamper the use of the observations over land where surface emissivity retrieval is required. This is likely to be a consequence of two things: firstly the fact that emissivity was not included in the state vector for the DFS calculation; and secondly, the particularly large observation errors ascribed to the surface-sensitive channels. For an operational channel selection, it would be desirable to rerun the channel selection including emissivity in the $\mathbf{B}$ matrix, in line with the operational preprocessing step in the Met Office assimilation system.

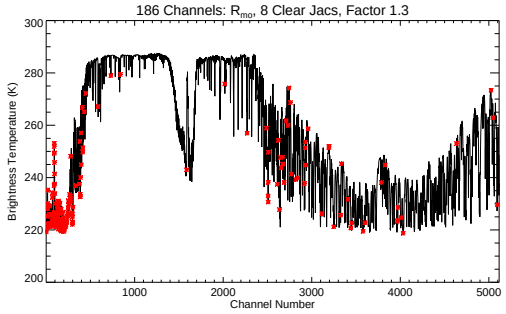
Figure 7.5 compares several of the *Method 2* channel selections in terms of their distributions within the IASI spectrum: they vary considerably in spectral coverage, but this alone is not sufficient to assess which channel selection is most useful for NWP. The next chapter assesses the information content and mathematical stability of the $\mathbf{R}$ matrix for each of the channel selections presented in this chapter.

7.4 Summary

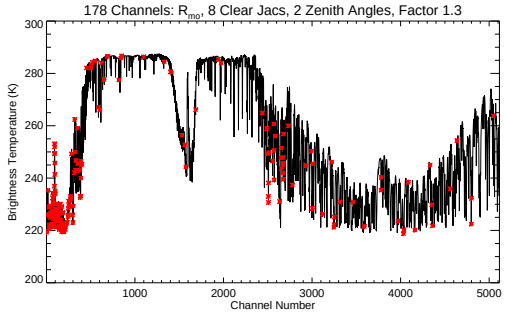
Two new channel selection methods have been developed to select reconstructed radiance channels for NWP. Method 1, which selects channels based on their contribution to the retained eigenvectors used for PC compression, was originally used by



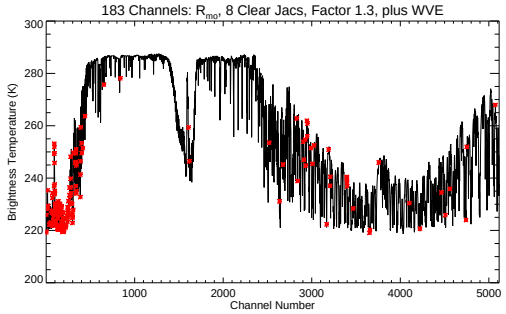
(a) *RHL-M2-8J-1.0*



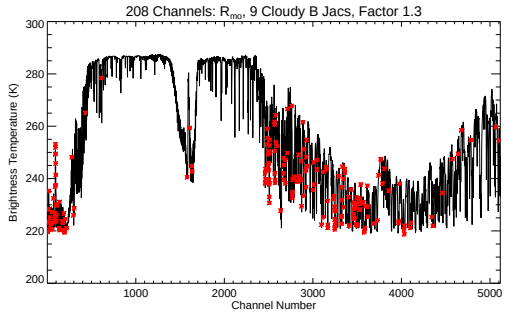
(b) *RHL-M2-8J-1.3*



(c) *RHL-M2-8J2Z-1.3*



(d) *RHL-M2-8J-1.3-WVE*



(e) *RHL-M2-9J-1.3*

Figure 7.5: *Method 2* channel selections plotted in red stars over a typical IASI brightness temperature spectrum

EUMETSAT to generate the channel selection used in their Level 2 processor. For this thesis, to tailor it to an NWP application, it was modified to take account of observation errors. However, this modification did not result in a substantial difference to the channel selection produced.

Method 2 is a new channel selection method based on information content, taking full account of observation error covariances, and penalising the selection of a channel that raises the condition number of the resultant $\mathbf{R}$ matrix. The method has been used to explore the selection of channel sets for reconstructed radiances, which have a high degree of error correlation between channels, by varying the inputs for the channel selection process.

It was found that attempting to account for the possibility that a particular channel might be cloud affected did not have a significant effect on the number of long-wave CO_2 channels chosen. Likewise, attempting to penalise the selection of water vapour channels for their temperature information did not have a large effect on the number of water vapour channels chosen. The variation of the factor by which a channel was rejected because the resulting condition number of $\mathbf{R}$ was too high did have an important effect on the resulting channel selection: in particular, if condition number was not the main selection criterion it was not possible to choose the same number of channels as PC scores for an NWP application.

The channel selections were compared with the *Collard* selection. The *Lp-M1* and *RHL-M1* selections have many more window channels and fewer long-wave CO_2 channels than the *Collard* selection. The *Collard* selection has more long-wave CO_2 channels and more window channels than the *Method 2* selections. The *Method 2* channel selections vary considerably in the spectral location of the channels chosen.

Both Methods 1 and 2 are sequential channel selection methods and are therefore not able to provide a definitive 'best' channel selection. In the next chapter, the channel selections will be compared for their information content, and for the properties of the resultant $\mathbf{R}$ matrix.

Chapter 8

Optimal linear analysis for Reconstructed Radiances

8.1 Comparison of channel selections through information content and condition number

While the distribution of channels across the spectrum is interesting, it cannot tell us which is the best channel selection: the 98 *Method 2* channels with condition number factor 100.0 (*RHL-M2-8J-100.0*) could contain more information than the 210 *Lp-M1* channels. It is therefore useful to compare the DFS over a fixed set of atmospheric profiles.

It is also crucial to examine the condition number of $\tilde{\mathbf{R}}'$ for each selection as it is an important measure of stability. In fact it is found that, for many of the channel selections, the condition number can be well over 10^6 despite it being possible to calculate an inverted matrix using the Cholesky decomposition routine used in the channel selection code.

Table 8.1 shows the DFS calculated over the eight clear sky atmospheres used as the baseline for the channel selections, with $\mathbf{R}_{HL}$ used for the observation error. The last four columns give a comparison of DFS using a restricted set of channels up to the point where the condition number rises above 10^6. In many cases, up to half the channels are not used. In some respects, this is a little misleading because a condition number of 10^6 is arbitrary. In fact the $\mathbf{R}$ matrix used for operational assimilation derived from the Desroziers method had a condition number of about 2000 in its raw form, but this was reduced to around 20 by strengthening the diagonal for operational use (Weston et al., 2014). However, it provides another basis for comparison.

Table 8.1 contains some interesting results. The most notable result is that, as expected, the *Method 1* channel selections have a much lower DFS than the *Method 2* selections, providing only around one third of the DFS of the highest information content. They also result in an $\tilde{\mathbf{R}}'$ with a lower condition number than all the *Method 2* selections apart from *RHL-M2-8J-1.0*, as expected. Even when the

Channel Selection	Full Selection				Max Condition Number 1×10^6			
	NCh	**Condition Number**	**DFS**	**DFS per chan**	**NCh**	**Condition Number**	**DFS**	**DFS per chan**
Lp-M1	210	0.496×10^7	84.3	0.40	153	0.8126×10^6	69.3	0.45
RHL-M1	210	0.942×10^7	85.5	0.41	135	0.9958×10^6	60.1	0.44
RHL-M2-8J-1.0	210	0.669×10^6	88.9	0.42	210	0.6694×10^6	88.9	0.42
RHL-M2-8J-1.3	186	0.268×10^{11}	196.3	1.05	83	0.8665×10^6	101.7	1.25
RHL-M2-8J-100.0	98	0.438×10^{14}	104.7	1.07	57	0.4315×10^6	79.8	1.40
RHL-M2-8J2Z-1.3	178	0.520×10^{11}	178.8	1.00	81	0.7961×10^6	98.2	1.21
RHL-M2-11J-1.3	184	0.849×10^{11}	174.8	0.94	97	0.9321×10^6	103.0	1.06
RHL-M2-9J-1.3	208	0.757×10^{12}	195.0	0.94	105	0.9375×10^6	106.2	1.01
RHL-M2-14J-1.3	169	0.479×10^{11}	161.1	0.95	95	0.8831×10^6	100.5	1.06
RHL-M2-8J-1.3-WVE	183	0.130×10^{11}	207.1	1.13	76	0.9469×10^6	91.5	1.20

Table 8.1: Condition number, DFS over 8 Jacobians, and DFS per channel for *Method 1* and *Method 2* channel selections using $\mathbf{R}_{HL}$ as the instrument noise covariance matrix. The left hand set of columns use the maximum possible channel selection. The right hand columns use the number of channels retained when the condition number is restricted to below 1×10^6.

channel selections are restricted to keep the condition number below 1×10^6 there are *Method 2* channel selections that give a much higher DFS.

The relative information content of each of the channel selections is similar regardless of whether the channel selections are restricted by keeping the condition number below 1×10^6 or not. *Method 2* selections give better DFS than *Method 1*, but the attempts to tailor the selection towards cloudy scenes or to avoid channels sensitive to background errors in water vapour result in a similar DFS. The use of a condition number threshold of 1.3 with Method 2 (*RHL-M2-8J-1.3*) seems to strike a reasonable balance and the channel selection has a high overall DFS and also a fairly high DFS per channel: the selection performs relatively well even when restricted to less than half the total number of selected channels. Similarly *RHL-M1* does not perform any better than *Lp-M1*, and the resultant $\tilde{\mathbf{R}}'$ is less well conditioned.

Additional tests showed that the channel selections perform better when the observation error term used to calculate the DFS subsequent to channel selection matches that used during the channel selection. If $\mathbf{R}_{EC}$ is used to calculate DFS, channel selections made using $\mathbf{R}_{EC}$ perform better than the channel selections presented here based on $\mathbf{R}_{HL}$ (not shown). Conversely, channel selections based on $\mathbf{R}_{EC}$ do not give as good results for DFS calculated using $\mathbf{R}_{HL}$ as the $\mathbf{R}_{HL}$ channel selections do (also not shown). This indicates a need for care when assimilating reconstructed radiances: the best channel selection seems to be highly tuned to the assumed errors.

Collard (2012) was concerned that, without an attempt to deal with cloudy atmospheres in the channel selection process, the resultant channel selection would be deficient when channels were restricted to those peaking above clouds. To test this theory, the DFS was calculated over the eight cloudy scenes included in the Cloudy C collection, but using the Pavelin et al. (2008) method of assimilating only those channels for which less than 10 % of the Jacobian falls below the cloud top (note that the Jacobians are not zeroed below the cloud top for this calculation). The results are shown in table 8.2. They are very similar to those presented in table 8.1: the channel selections where cloudy conditions were explicitly modelled perform no better than those based only on clear atmospheres.

Channel Selection	NCh	DFS	DFS per chan
Lp-M1	210	54.4	0.26
RHL-M1	210	55.9	0.26
RHL-M2-8J-1.0	210	60.3	0.29
RHL-M2-8J-1.3	186	136.1	0.73
RHL-M2-8J-100.0	98	71.9	0.74
RHL-M2-8J2Z-1.3	178	125.5	0.71
RHL-M2-11J-1.3	184	117.9	0.64
RHL-M2-9J-1.3	208	132.1	0.64
RHL-M2-14J-1.3	169	111.2	0.66
RHL-M2-8J-1.3-WVE	183	141.1	0.77

Table 8.2: DFS over six cloudy Jacobians (those used in Cloudy C) and DFS per channel for *Method 1* and *Method 2* channel selections using $\mathbf{R}_{HL}$ as the observation error covariance matrix.

8.2 A reconstructed radiance channel selection for instrument noise only

The main aim of this thesis was to design a channel selection for reconstructed radiances that contains almost the whole useful information content of the band 1 and 2 spectrum. In order to see whether the *Method 1* or *Method 2* selections achieve this, the channel selections should ideally be compared to the DFS for the full spectrum also. The problem is that the $\mathbf{R}_{HL}$ matrix is calculated from reconstructed radiance residuals, and is not invertible for the full spectrum. It is, therefore, not possible to do a clean intercomparison of the new *Method 2* reconstructed radiance channels with the full raw radiance spectrum using this matrix. The *Lp-M1* selection can be tested because it is equally valid for instrument noise as any other error covariance matrix.

As a general test of Method 2, channel selections were also made for the instrument noise covariance matrix projected into reconstructed radiance space, $\tilde{\mathbf{N}}$ (equation 6.13). The breakdown of the classification of the channels is shown in table 8.3. The $\tilde{\mathbf{N}}$ selections contain many more window channels than the *Collard* selection, and are fairly similar in sensitivity breakdown to the *Method 1* channel

Channel Selection	CO_2	Window	Ozone	Water Vapour	Band 1	Band 2	Total NChan
Ñ-M2-8J-1.0	21	68	19	102	90	120	210
Ñ-M2-8J-1.3	52	35	10	113	90	120	210
Collard	132	24	15	98	165	104	269

Table 8.3: Breakdown of the number of channels by principal sensitivity for the channel selection using reconstructed instrument noise, $\tilde{\mathbf{N}}$. Comparison with *Collard* selection for bands 1 and 2.

selections (table 7.1). Without the inclusion of forward model error and errors of representation, using DFS in the selection process by setting a threshold of 1.3 in selection *Ñ-M2-8J-1.3* (as opposed to 1.0, which just selects on condition number giving channel selection *Ñ-M2-8J-1.0*) does not alter the distribution of channels as much as it does when using $\mathbf{R}_{HL}$.

Table 8.4 compares the DFS of the new $\tilde{\mathbf{N}}$ channel selections and the *Lp-M1* channel selection with those of the full spectrum for bands 1 and 2, the 210 PC scores themselves, and with the *Collard* and *VAR* channel selections. The DFS is calculated over the eight atmospheric profiles used during the channel selection process. The *Ñ-M2-8J-1.3* channel selection has a similar performance to the *Collard* selection for raw radiances (the $\mathbf{H}$–$\mathbf{N}$ combination), at approximately two thirds of the DFS of the full spectrum. The *Ñ-M2-8J-1.0* selection has a much poorer information content of only 89.1.

When exploiting the reconstructed radiances using $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{N}}$, all three reconstructed radiance channel selections perform significantly better. *Ñ-M2-8J-1.0* and *Lp-M1* both yield a DFS almost identical to that of the PC scores themselves. This is the desired result: it is possible to produce the same DFS with a carefully chosen set of reconstructed radiance channels as with the PC scores. The DFS for these systems is close to, but slightly lower than the DFS for the full spectrum: it is not surprising that some small quantity of information will be lost due to compression.

The *Ñ-M2-8J-1.3* selection using $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{N}}$ gives an information content of 162.6 in comparison with 160.9 for the full spectrum: this is slightly too high. One thing to note, however, is that the condition number of $\tilde{\mathbf{N}}$ is very high in this case. This could cause mathematical instabilities when using iterative methods to arrive at a solution in variational analysis and may also be the reason for the DFS being slightly larger than expected. As described in section 7.2.2, poor conditioning is a problem,

because a covariance matrix is used in inverse form. Small eigenvalues theoretically represent observed quantities with small errors. However, if an observed quantity had an eigenvalue of exactly zero, one would say that there is no independent information about that quantity at all since it can be constructed as a linear combination of the other observed quantities, and it should be dropped from the observation vector (as discussed in Twomey, 1996). The problem arises where eigenvalues tend towards zero and change from being very well observed quantities with small errors to quantities about which there is no information at all. Because the $\mathbf{R}$ matrix is used in inverse form, and because $1/x$ is a highly non-linear function tending to infinity where $x = 0$, very small eigenvalues in the covariance matrix result in enormous eigenvalues in the inverse matrix, generating very unstable solutions. In contrast, for this idealistic scenario, the PC score covariance matrix is the identity matrix, and therefore not susceptible to such problems. However, the condition number of the matrix is not the final story: if it were, the answer to the question of whether to assimilate reconstructed radiances or PCs would be clear cut. The results presented in chapter 9 give a more complete picture of the differences and similarities of PC and reconstructed radiance assimilation.

8.3 Suboptimal-in-H-and-R assimilation

Chapter 6 showed that the use of a raw radiance Jacobian during assimilation would result in an extra error term, $\Delta\tilde{\mathbf{R}}$, that needs to be taken account of. This theoretical outcome is confirmed by a linear analysis for a suboptimal-in-$\mathbf{H}$-and-$\mathbf{R}$ system in table 8.5, identified by the combination $\mathbf{H}$–$\tilde{\mathbf{N}}$. There can never be more information in any processed form of the observation than in the full raw spectrum, therefore, as shown in table 8.4, the DFS for these instrument noise only systems should never be higher than 160.9. It has also been shown that with the correct error profile and Jacobian, a reconstructed radiance channel selection can be defined that yields a DFS very similar to that of the full spectrum. However, if the DFS for the reconstructed radiances is calculated using $\tilde{\mathbf{N}}$, but with raw radiance Jacobians $\mathbf{H}$, the apparent DFS is much higher than that for the full spectrum (table 8.5). The extra DFS seems to result from the correlations in the $\tilde{\mathbf{N}}$ matrix. Spurious correlations due to ill-conditioning and neglected error terms can result in the addition of pseudo-information, that in reality is not information, but forward model/reconstruction error that has not been accounted for. The *Ñ-M2-8J-1.3* channel selection seems to suffer the most from this problem, and it is likely that this is because the $\mathbf{R}$ matrix

Channel Selection	NChan	Jacobian	R matrix	DFS	DFS per chan	Condition Number
Raw Radiance channel selections						
VAR bands 1 and 2	127	$\mathbf{H}$	$\mathbf{N}$	65.3	0.51	157.8
Collard	269	$\mathbf{H}$	$\mathbf{N}$	105.5	0.39	179.8
Full Spectrum	5116	$\mathbf{H}$	$\mathbf{N}$	160.9	0.03	429.4
Reconstructed radiance channel selections in raw radiance space						
Lp-M1	210	$\mathbf{H}$	$\mathbf{N}$	99.0	0.47	211.7
Ñ-M2-8J-1.0	210	$\mathbf{H}$	$\mathbf{N}$	89.1	0.42	109.9
Ñ-M2-8J-1.3	210	$\mathbf{H}$	$\mathbf{N}$	102.3	0.49	283.3
Reconstructed radiance channel selections in reconstructed radiance space						
Lp-M1	210	$\tilde{\mathbf{H}}$	$\tilde{\mathbf{N}}$	152.2	0.72	5.6×10^4
Ñ-M2-8J-1.0	210	$\tilde{\mathbf{H}}$	$\tilde{\mathbf{N}}$	152.3	0.73	4.5×10^4
Ñ-M2-8J-1.3	210	$\tilde{\mathbf{H}}$	$\tilde{\mathbf{N}}$	162.6	0.77	1.4×10^8
PC Scores						
210 PC Scores	N/A	$\mathbf{H}_{pc}$	$\mathbf{I}$	152.2	N/A	1.0

Table 8.4: DFS and condition number comparison for the reconstructed radiance channel selections chosen using instrument noise. DFS is calculated over eight clear atmospheric profiles, using Jacobians and $\mathbf{R}$ as specified in the third and fourth columns.

Channel Selection	NChan	Jacobian	R matrix	DFS	DFS per chan	Condition Number
Lp-M1	210	$\tilde{\mathbf{H}}$	$\tilde{\mathbf{N}}$	152.2	0.72	5.6×10^4
Lp-M1	210	$\mathbf{H}$	$\tilde{\mathbf{N}}$	176.1	0.84	5.6×10^4
Ñ-M2-8J-1.0	210	$\tilde{\mathbf{H}}$	$\tilde{\mathbf{N}}$	152.3	0.73	4.5×10^4
Ñ-M2-8J-1.0	210	$\mathbf{H}$	$\tilde{\mathbf{N}}$	181.3	0.86	4.5×10^4
Ñ-M2-8J-1.3	210	$\tilde{\mathbf{H}}$	$\tilde{\mathbf{N}}$	162.6	0.77	1.4×10^8
Ñ-M2-8J-1.3	210	$\mathbf{H}$	$\tilde{\mathbf{N}}$	266.8	1.27	1.4×10^8

Table 8.5: DFS and condition number comparison for the reconstructed radiance channel selections using instrument noise. The results are compared for the use of the correct reconstructed radiance Jacobian, and for the incorrect raw Jacobian. DFS is calculated over eight clear atmospheric profiles, using Jacobians and $\mathbf{R}$ as specified in the third and fourth columns.

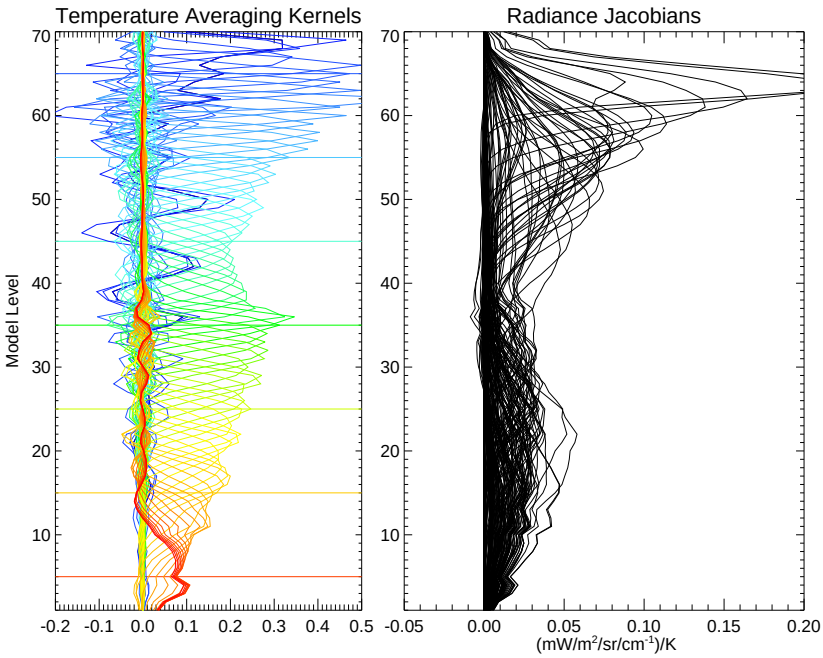
is less well conditioned for this selection. The *Lp-M1* channel selection is the least affected, but the differences in behaviour between *Lp-M1* and *Ñ-M2-8J-1.0* are not thought to be significant.

The particularly interesting thing about this is that the raw and reconstructed Jacobians appear almost identical to the naked eye and yet they result in completely different averaging kernels (compare figure 8.1 with figure 8.2). Despite this, the averaging kernels confirm that the $\mathbf{H}$–$\tilde{\mathbf{N}}$ suboptimal-in-$\mathbf{H}$-and-$\mathbf{R}$ system has far more information in the upper atmosphere than one might realistically expect for IASI (compare figure 8.1 with figure 4.6(b) and figure 4.6(d) noting difference in the horizontal scale).

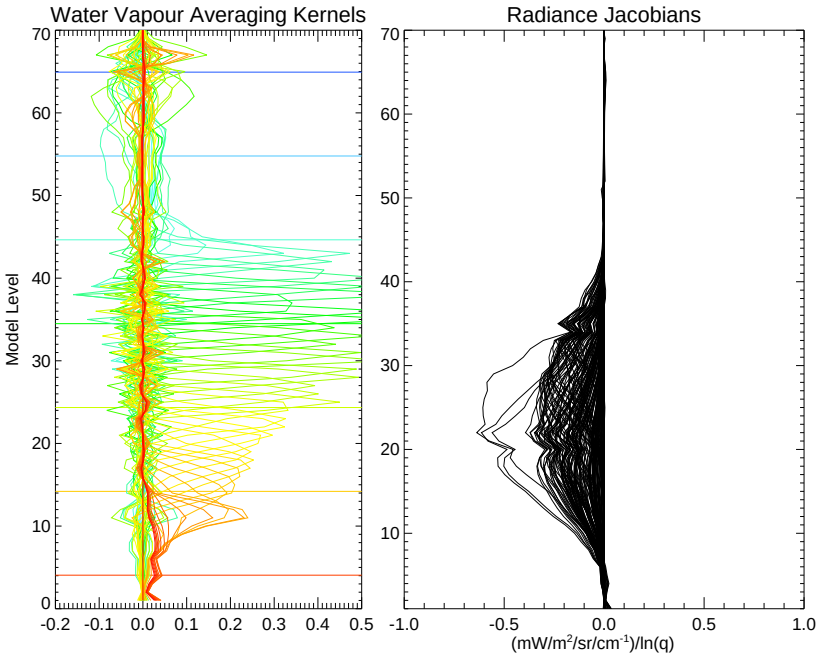
8.3.1 Matrix conditioning

This result is important, because in the real assimilation system, we may not be able to model $\tilde{\mathbf{H}}$ and must instead use $\mathbf{H}$, i.e. the system will be suboptimal-in-$\mathbf{H}$. By capturing the extra errors introduced using the observational method to estimate the total error budget empirically, it is hoped that the system is not badly suboptimal-in-$\mathbf{R}$ as well. Even so, the conditioning of the resultant $\tilde{\mathbf{R}}'$ matrix is poor. The poor conditioning is partly a result of the process of covariance estimation itself: large eigenvalues are overestimated, and small ones are underestimated. Even in the case of the idealised instrument noise only system, there is an element of estimation as the PC eigenvectors themselves are estimated from a collection of IASI spectra.

As discussed in section 7.2.2, applications relying on covariance estimation, such

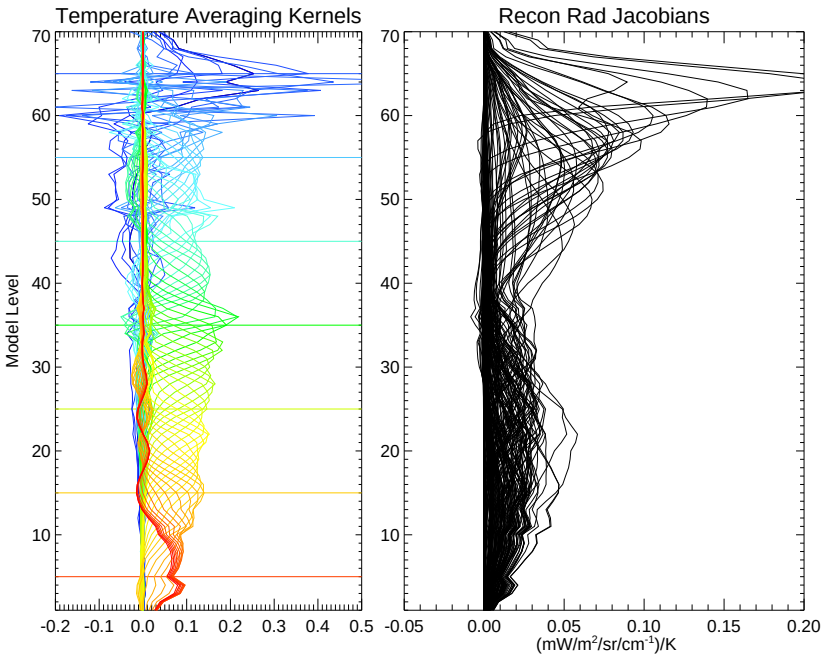


(a) Temperature

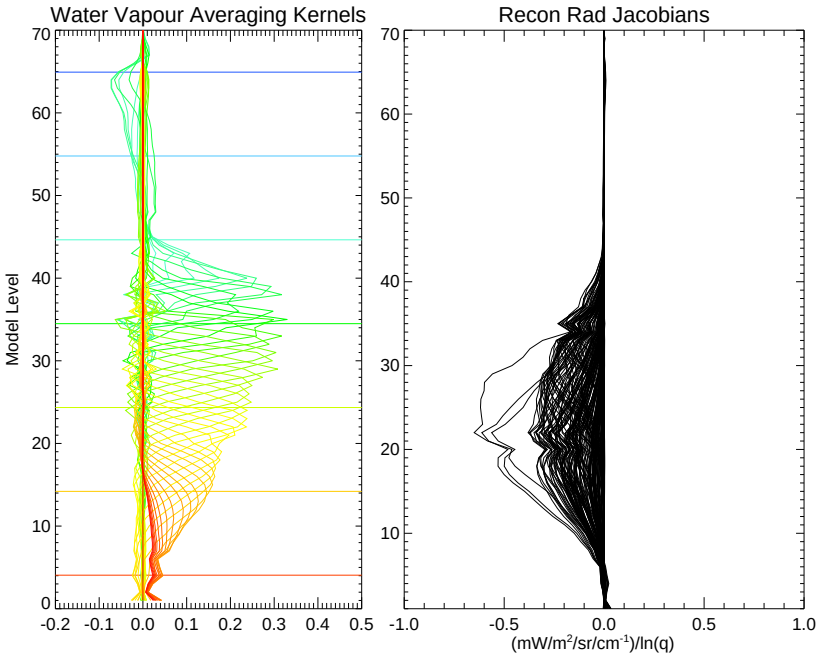


(b) Water Vapour

Figure 8.1: Raw radiance Jacobians and averaging kernels calculated using these raw radiance Jacobians and $\tilde{\mathbf{N}}$ for the *Ñ-M2-8J-1.3* channels for the US Standard Atmosphere



(a) Temperature



(b) Water Vapour

Figure 8.2: Reconstructed radiance Jacobians and averaging kernels calculated using these reconstructed Jacobians and $\tilde{\mathbf{N}}$ for the *Ñ-M2-8J-1.3* channels for the US Standard Atmosphere

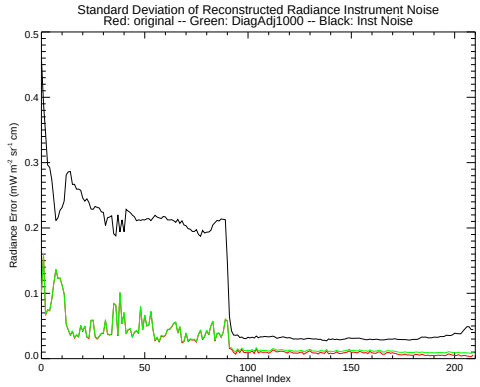
as finance and biostatistics, often employ a shrinkage methods to counteract the problems inherent in the sampling during covariance estimation. There is no exact method recommended for NWP applications, but various techniques can be explored. An extreme form of shrinkage is to strengthen the diagonal of the matrix relative to the off-diagonal elements incrementally until the condition number reaches a desired threshold. An alternative is to manipulate the eigenvalues directly, for example increasing the small ones until the threshold condition number is attained. The introduction of a correlated error matrix at the Met Office, calculated via the Desroziers et al. (2005) diagnostic, also relied on such methods to improve the stability of the minimisation in 4D-Var (Weston et al., 2014). It is thus of interest to see whether such processes are capable of improving the numerical stability of the $\tilde{\mathbf{N}}$ matrix, without reducing the information content of the system to too great an extent.

Table 8.6 shows the effects of strengthening the diagonal, either by adding increments equal to the smallest eigenvalue to the diagonal (rows marked DiagAdj), or by increasing the small eigenvalues themselves (row marked EigAdj) until the condition number is reduced to the degree shown in the final column. The effect on the squareroot of the diagonal of the matrix in each case is shown in figure 8.3. The biggest changes are in band 2, and in the DiagAdj20 case the diagonal is inflated above the level of instrument noise. The DFS can easily be brought down to a level where the pseudo-information appears to have been eliminated. This can also be seen in the averaging kernels, which are now more similar to those of the $\tilde{\mathbf{H}}$–$\tilde{\mathbf{N}}$ system (figure 8.4), albeit smoother at the top of the profile and with less sensitivity in the bottom levels. (In fact it is likely that the poor conditioning of $\tilde{\mathbf{N}}$ in the optimal scenario has resulted in some pseudo-information at the top of the profile.) However, aggressive strengthening of the diagonal reduces the DFS to a level consistent with the raw radiances (table 8.4). On the other hand, figure 8.5 compares the predicted analysis performance of the suboptimal and optimal systems, and the results with the suboptimal-in-$\mathbf{H}$-DiagAdj20 system are in fact similar to the optimal assimilation of reconstructed radiances.

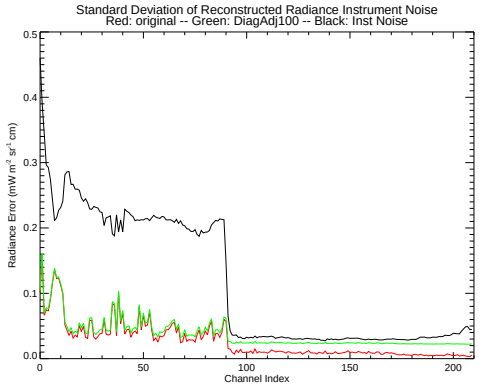
Collard et al. (2010) showed that if the diagonal of the error covariance matrix is used for reconstructed radiances rather than the full matrix, the analysis can be degraded relative to the background. The results presented here suggest that the precise nature of the covariance structure is also important: it is no good providing off-diagonal elements that are not correct, as this too will endanger the quality of the analysis. In fact, it seems better to damp down these incorrect correlations via

Jacobian	R matrix	DFS	DFS per chan	Condition Number
H	$\tilde{\mathbf{N}}$	266.8	1.27	1.4×10^8
H	$\tilde{\mathbf{N}}$ DiagAdj1000	141.5	0.67	530
H	$\tilde{\mathbf{N}}$ DiagAdj100	123.7	0.59	67
H	$\tilde{\mathbf{N}}$ DiagAdj20	108.5	0.52	18
H	$\tilde{\mathbf{N}}$ EigAdj100	132.0	0.63	91

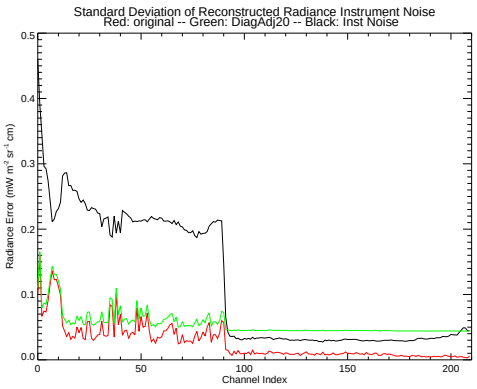
Table 8.6: DFS and condition number comparison for the suboptimal-in-$\mathbf{H}$-and$\mathbf{R}$ assimilation of the *Ñ-M2-8J-1.3* reconstructed radiance channel selection demonstrating the effect of re-conditioning the $\mathbf{R}$ matrix. DFS is caclulated over 8 clear atmospheric profiles, using Jacobians and $\mathbf{R}$ as specified in the third and fourth columns. Explanations of the $\mathbf{R}$ terms are given in the text.



(a) DiagAdj1000

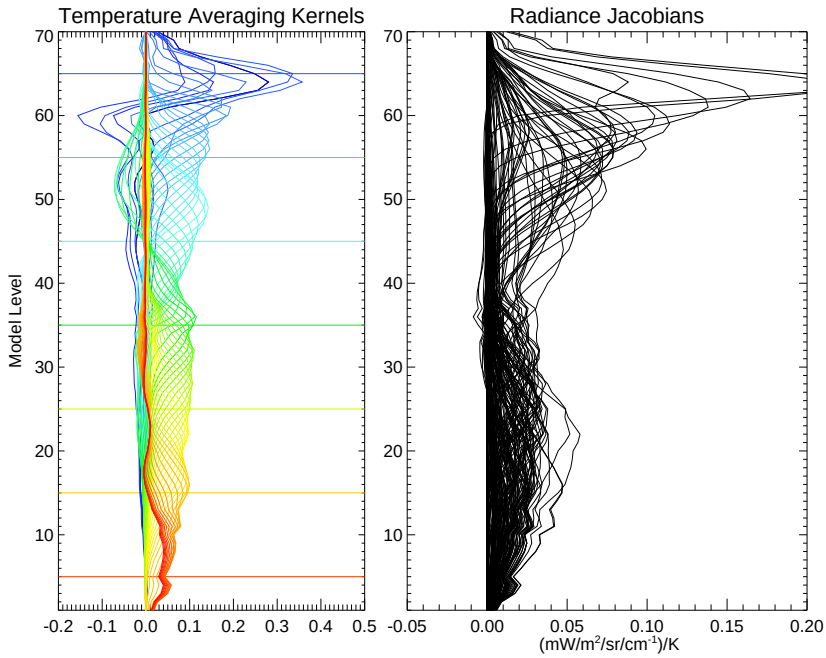


(b) DiagAdj100

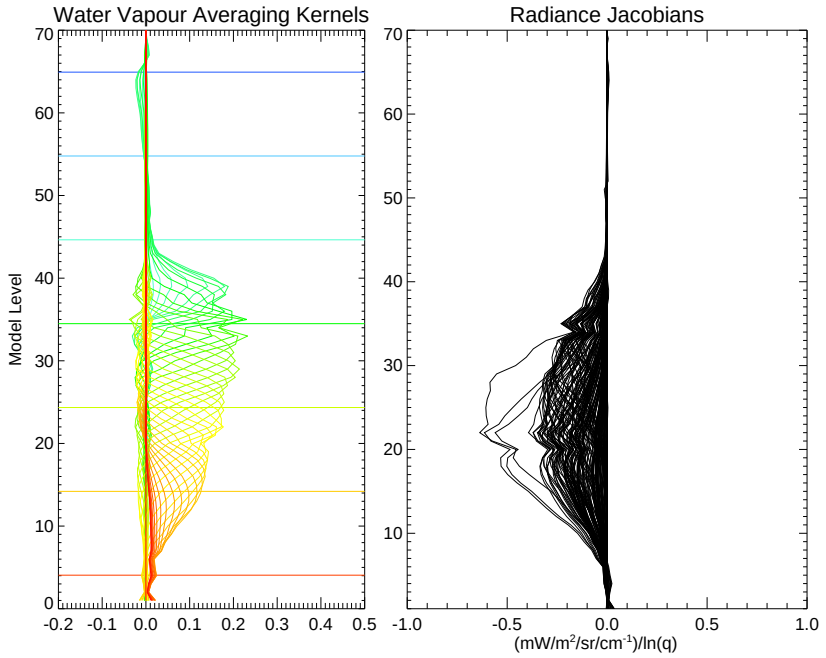


(c) DiagAdj20

Figure 8.3: The effect of DiagAdj shrinkage operations on the $\tilde{\mathbf{N}}$ matrix. Plotted in green is the shrunk matrix, in red is the $\tilde{\mathbf{N}}$ matrix and in black is the $\mathbf{N}$ raw radiance instrument noise. The plotted quantity is the square root of the diagonal of the matrix, i.e. the standard deviation of the assumed noise.

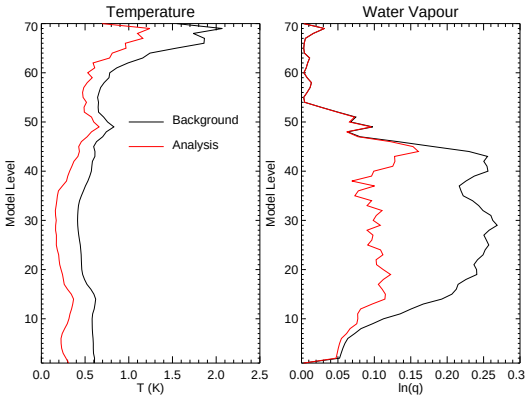


(a) Temperature

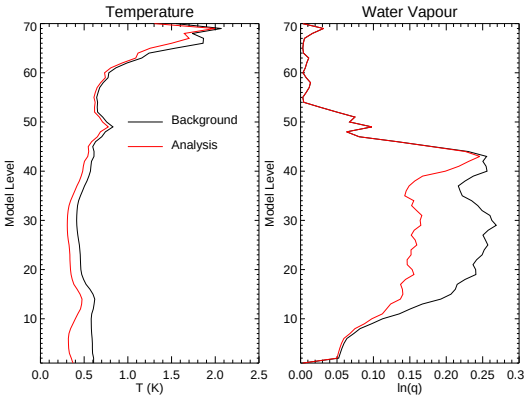


(b) Water Vapour

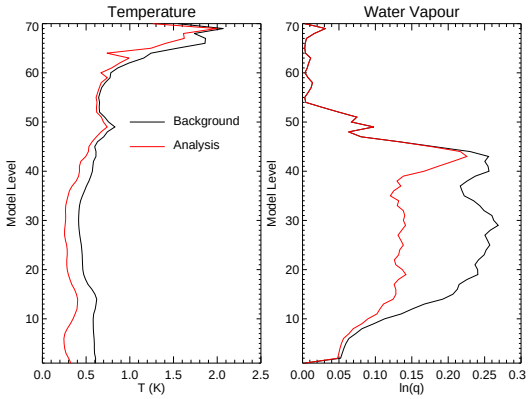
Figure 8.4: Raw radiance Jacobians and averaging kernels calculated using these raw Jacobians and $\tilde{\mathbf{N}}$ DiagAdj20 for the *Ñ-M2-8J-1.3* channels for the US Standard Atmosphere



(a) Suboptimal-in-$\mathbf{H}$-and-$\mathbf{R}$ ($\mathbf{H}$–$\tilde{\mathbf{N}}$) Before adjustment: condition number 1.4×10^8

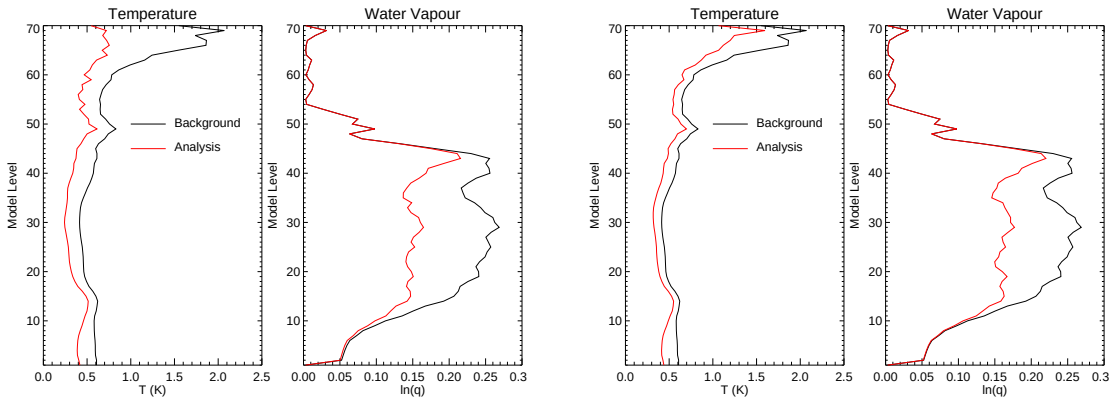


(b) Suboptimal-in-$\mathbf{H}$-and-$\mathbf{R}$ Diagadj20 ($\mathbf{H}$–$\tilde{\mathbf{N}}$Diagadj20): condition number 18



(c) Optimal ($\tilde{\mathbf{H}}$–$\tilde{\mathbf{N}}$): condition number 1.4×10^8

Figure 8.5: The effect of adjusting the diagonal of the reconstructed radiance instrument noise covariance matrix $\mathbf{L}_p\mathbf{L}_p^{\mathsf{T}}\mathbf{L}_p\mathbf{L}_p^{\mathsf{T}}$ on a supoptimal-in-$\mathbf{H}$-and-$\mathbf{R}$ retrieval using 210 channels, compared with the optimal retrieval where $\tilde{H}(\mathbf{x})$ is used for the jacobian instead of $H(\mathbf{x})$



(a) Before adjustment: condition number 2.7×10^{10}

(b) After adjustment: condition number 6.9×10^{5}

Figure 8.6: Linear analysis predicted retrieval errors relative to background error for the *RHL-M2-8J-1.3* channel selection, for R matrices before and after shrinkage.

shrinkage, and bring the matrix closer to the diagonal. This is, perhaps, counterintuitive, but the DiagAdj method used here increases the variance at the same time as making the matrix more diagonally-dominant. This increase in variance helps to prevent the overweighting of the observations that Collard demonstrated.

If the same DiagAdj shrinkage procedure is performed with the $\tilde{\mathbf{R}}'$ derived from the observational method for the 186 channel selection, DFS reduces from 196 to a more realistic 131 if the condition number is reduced from 2.7×10^{10} to 6.9×10^{5}. The effect on predicted linear retrieval performance is shown in figure 8.6. Again, the improvements of the temperature retrieval in the stratosphere appear to have been reduced to a more reasonable level by the shrinkage procedure. The next step is to test whether the shrinkage methods explored here provide a practical solution to allow assimilation in a suboptimal-in-**H**-and-**R** context. The next chapter tests this using 1D-Var.

8.4 Summary

The channel selections generated for operational implementation of reconstructed radiances at the Met Office assuming an observation error covariance matrix of $\mathbf{R}_{HL}$ were compared for information content over eight atmospheres. This is a simple test to select the most promising channel selections for further investigation, rather than a method for selecting one single 'best' channel selection. The *Method 2* selections result in an NWP system with higher overall DFS than the *Method 1* selections.

Tailoring the channel selection to NWP by choosing on the basis of DFS appears

to provide a more information-rich channel selection in fewer channels than when choosing on the basis of condition number of the **R** matrix alone. Most atmospheric columns are cloudy, resulting in a rejection of low-peaking channels in 4D-Var with the current assimilation scheme. Attempting to further tailor the selection to avoid over-selection of low-peaking channels did not result in a selection that performed better when tested on a range of cloudy atmospheres. For this reason, the channel selection with a condition number threshold of 1.3 but no further modifications, *RHL-M2-8J-1.3*, is preferred because it is a simpler method, and the resultant selection has plenty of channels in the CO_2 band.

To test whether the reconstructed radiance channel selection method is capable of finding channels that represent the full information content of the observation, further channel selections were made assuming instrument noise only. DFS calculations show that the reconstructed radiance assimilation is capable of providing the full information content of the 210 PC scores, close to that of the full spectrum, but the **R** matrix is not well-conditioned, risking the addition of pseudo-information to the retrieval and an unstable minimisation.

The suboptimal-in-**H**-and-**R** system has been explored. The misspecification of **R** leads to an apparent DFS that is higher than the DFS for the full spectrum and is therefore erroneous. Two methods of shrinkage of the **R** matrix have been explored, as a possible practical measure that can be used to correct for the missing reconstruction error. The methods of shrinkage explored here do help to remove pseudo-information that is not present in the full spectrum, but in order to be effective the DFS is reduced considerably.

The performance of reconstructed radiances will be tested against raw radiances and against the assimilation of PC scores in a simulated 1D-Var experiment in the next chapter.

Chapter 9

Comparison of PC and reconstructed radiance assimilation using a 1D-Var

In table 8.4 it was demonstrated that reconstructed radiances theoretically provide a very similar information content to PC scores in an optimal linear assimilation. This theory is now tested using simulated 1D-Var experiments, where the effects of nonlinearities in the response of the radiance to changes in atmospheric profile can be tested. First, an instrument noise only system is used to compare the assimilation of reconstructed radiances, using channel selection *Ñ-M2-8J-1.3*, with that of PC scores. Then the preferred channel selection for operational implementation derived assuming the $\mathbf{R}_{HL}$ observation error covariance matrix, *RHL-M2-8J-1.3*, is tested.

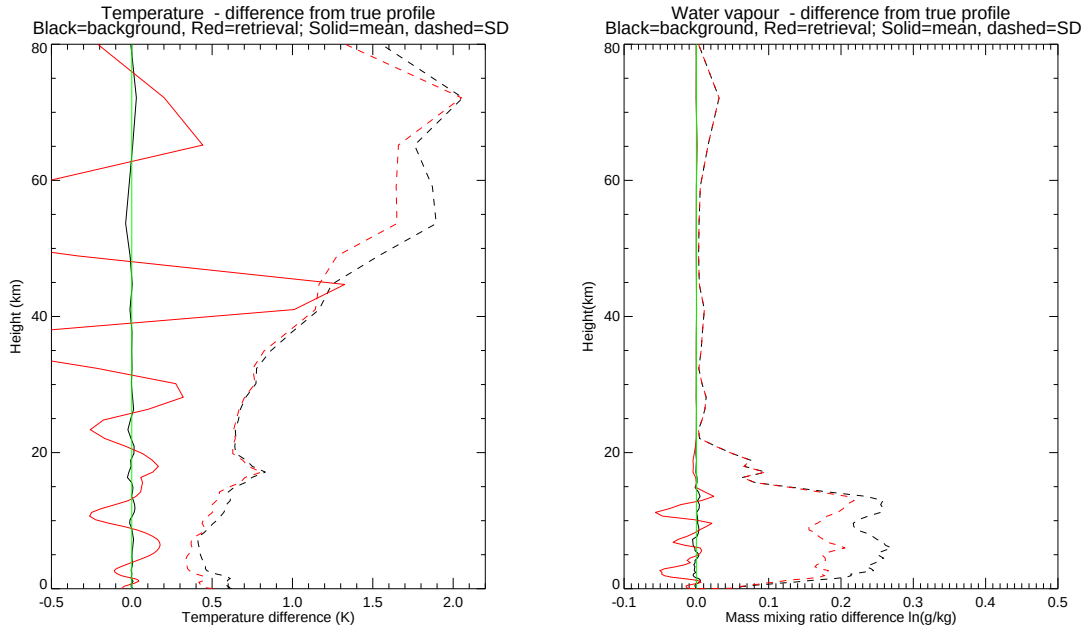
The same 4348 UM profiles on 70 levels that were described in section 3.8.3 were used to construct the experiment, and the background profiles were simulated by adding noise consistent with the **MoistCov B** matrix.

9.1 Instrument noise only 1D-Var

9.1.1 Reconstructed radiance assimilation

Simulated true radiances computed with RTTOV-10 were converted to simulated observations by the addition of noise consistent with the noise profile used for normalisation by the EUMETSAT PC compression. This choice was made because it simplifies the calculation of the $\mathbf{R}$ matrix, and because the 210 channel selection was produced from this noise profile. The observations were then converted into PC scores using the EUMETSAT PC compression process, and then converted back to reconstructed radiances. The 210 *Ñ-M2-8J-1.3* channels, selected using Method 2 on the basis of reconstructed EUMETSAT instrument noise ($\tilde{\mathbf{N}}$) and a condition number threshold of 1.3 (see table 8.3 and appendix E), were tested in a suboptimal-in-$\mathbf{H}$-and-$\mathbf{R}$ 1D-Var simulated system.

The first result from these experiments is that assimilating reconstructed radiances assuming $\tilde{\mathbf{N}}$ with raw radiance jacobians, i.e. without taking account of the additional forward model error $\Delta\tilde{\mathbf{R}}$, does not work: the 1D-Var does not con-



(a) Temperature **(b)** Water Vapour

Figure 9.1: Mean and standard deviation of (Background–Truth) and (Analysis–Truth) for 4348 observations. Reconstructed radiance Jacobians and $\tilde{\mathbf{N}}$ for the 210 reconstructed radiance channels.

verge. It is not easy to calculate $\Delta\tilde{\mathbf{R}}$, but empirically shrinking the observation error covariance matrix by strengthening the diagonal as tested with linear analysis in section 8.3.1 does work to improve the performance of the 1D-Var. With the matrices DiagAdj100 and EigAdj100, the 1D-Var converges to a solution for most observations, but the results are very poor indeed, with the analysis generally further from the truth than the background (not shown). The results with matrix DiagAdj20 are much better, with the analysis error lower than the background error in terms of standard deviation, but with large oscillating biases introduced into the analysis (figure 9.1).

It is interesting to examine the profile increments calculated during minimisation. Figure 9.2 shows the performance of the 1D-Var for three different observations. In this figure, and similar ones that follow, the plotted lines show the difference between a given profile and the true profile. The black line is the background departure from truth, and the pink line is the background departure multiplied by the averaging kernel. The pink line thus represents the best result achievable given a perfect linear system and illustrates that the analysis cannot resolve every fluctuation because of the null space of the averaging kernels. The other coloured lines are the departure

of the profile on a given numbered iteration.

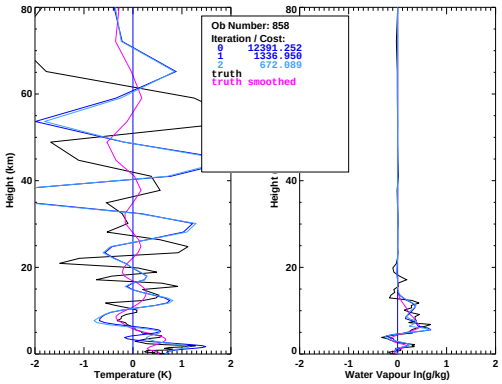
Arguably, in some cases the 1D-Var analysis actually captures some of the structures better than the linear results predict (for example between 40–60 km in figure 9.2(f)). However, it is clear that the suboptimal-in-**H**-and-**R** system allows large oscillations to result that do not match the pattern of background error. The off-diagonal elements of the $\tilde{\mathbf{R}}$ matrix allow correlated and anti-correlated increments to be generated throughout the profile, that are not properly filtered by the **B** matrix because of the overweighting of the observations that results from ignoring the $\Delta\tilde{\mathbf{R}}$ error term. This is counteracted to some extent by downweighting the observations by increasing the observation error variance.

In order to see whether there is any benefit from assimilating reconstructed radiances over raw radiances, the same 210 channel selection was also tested in a raw radiance assimilation system. The 1D-Var results in general confirm the linear analysis results presented in the DFS scores above: the same channel selection presented as raw radiances produces an analysis with nearly identical performance to the DiagAdj20 suboptimal-in-**H**-and-**R** reconstructed radiance analysis (figure 9.1) in terms of standard deviation, but without the oscillating bias (figure 9.3). In other words, in order to remove the oscillating bias completely, the reconstructed radiances would probably need to be downweighted to the extent that there would be no clear benefit to assimilating the reconstructed radiances over the raw observations.

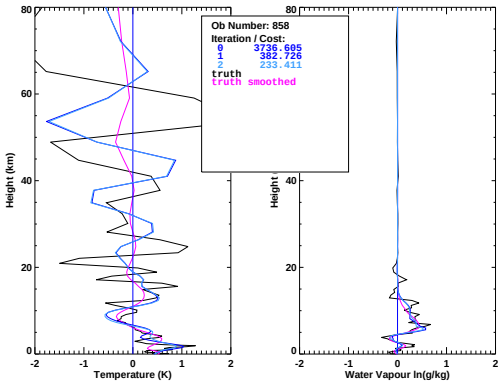
The 1D-Var results for water vapour are similar to the prediction by the optimal linear equations (compare figure 9.1 with figure 8.5), but the temperature analysis performs less well than predicted. The results with the new channel selection are much better than with the 4D-Var channel selection (figure 9.4), in agreement with the DFS calculations shown in table 8.4.

9.1.2 PC score assimilation

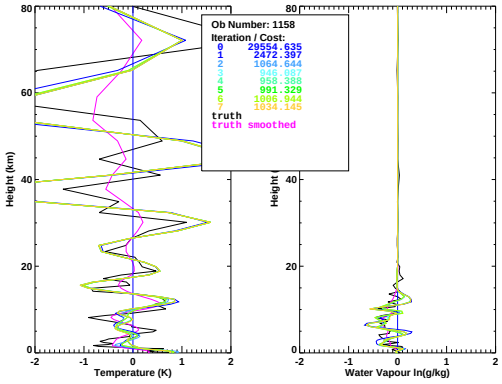
Unfortunately, no radiative transfer code to directly model the EUMETSAT PC scores currently exists, so a different set of PC eigenvectors is used instead. The chosen eigenvector set is 200 vectors provided with the radiative transfer model PC-RTTOV (Matricardi, 2010) to represent the full IASI spectrum based on 500 predictor channels (see appendix E). For the reconstructed radiance case, the noise profile used for normalisation before compression is used to construct the **R** matrix; for the PC case, the noise profile used in the PC-RTTOV compression is used. These noise profiles are different, but not significantly so (figure 9.5). This set-up differs from that used for reconstructed radiances, because information from all three bands



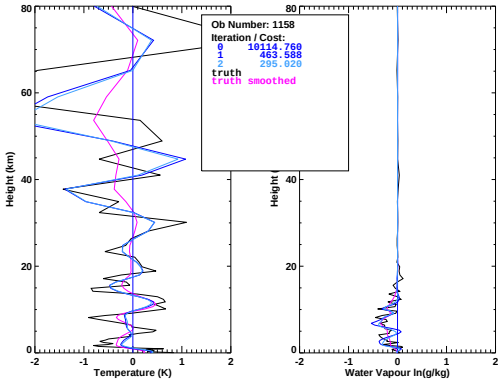
(a) Ob 858: DiagAdj100



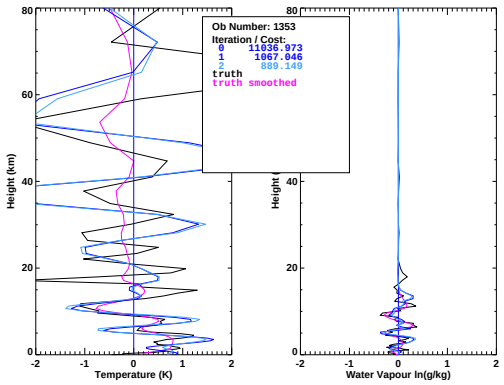
(b) Ob 858: DiagAdj20



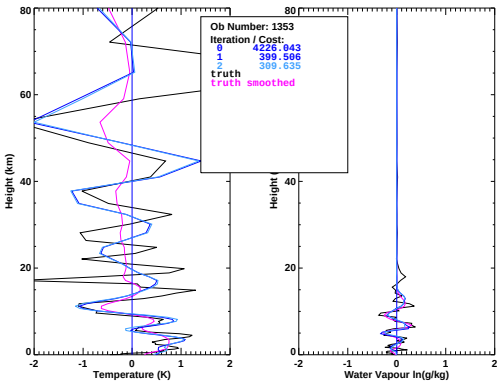
(c) Ob 1158: DiagAdj100



(d) Ob 1158: DiagAdj20

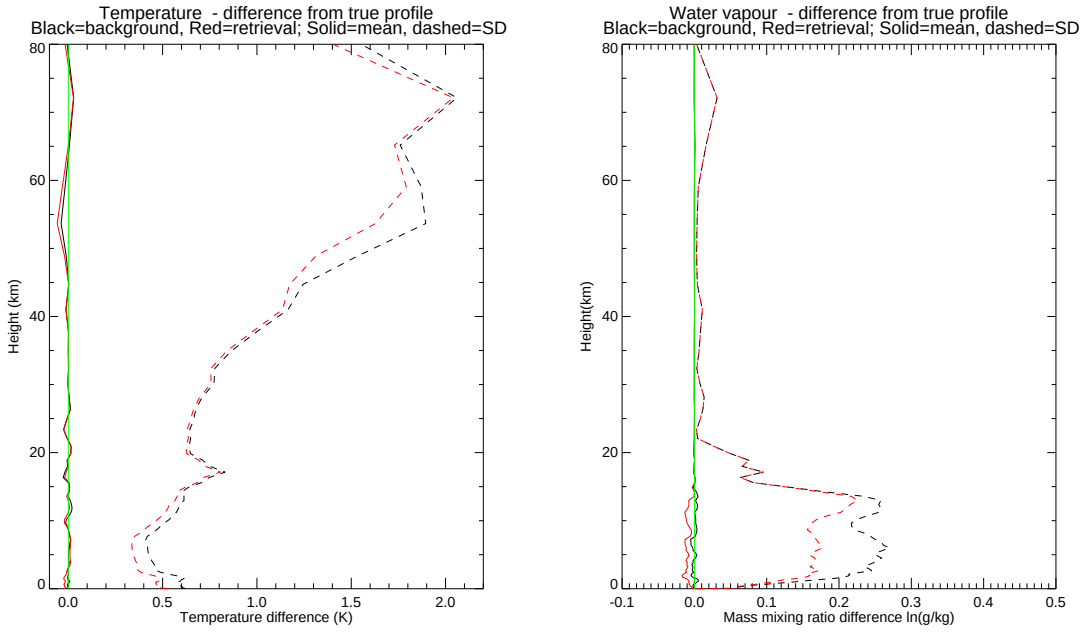


(e) Ob 1353: DiagAdj100



(f) Ob 1353: DiagAdj20

Figure 9.2: Minimisation process for observations 858, 1158 and 1353 for reconstructed radiance assimilation with instrument noise only. The left column use the $\tilde{\mathbf{N}}$ matrix modified to DiagAdj100, and the right column to DiagAdj20. The plotted lines show the difference between a given profile and the background profile. The black line is the background departure from truth, and the pink line is the background departure multiplied by the averaging kernel. The other coloured lines are the departure of the profile on a given numbered iteration.



(a) Temperature **(b)** Water Vapour

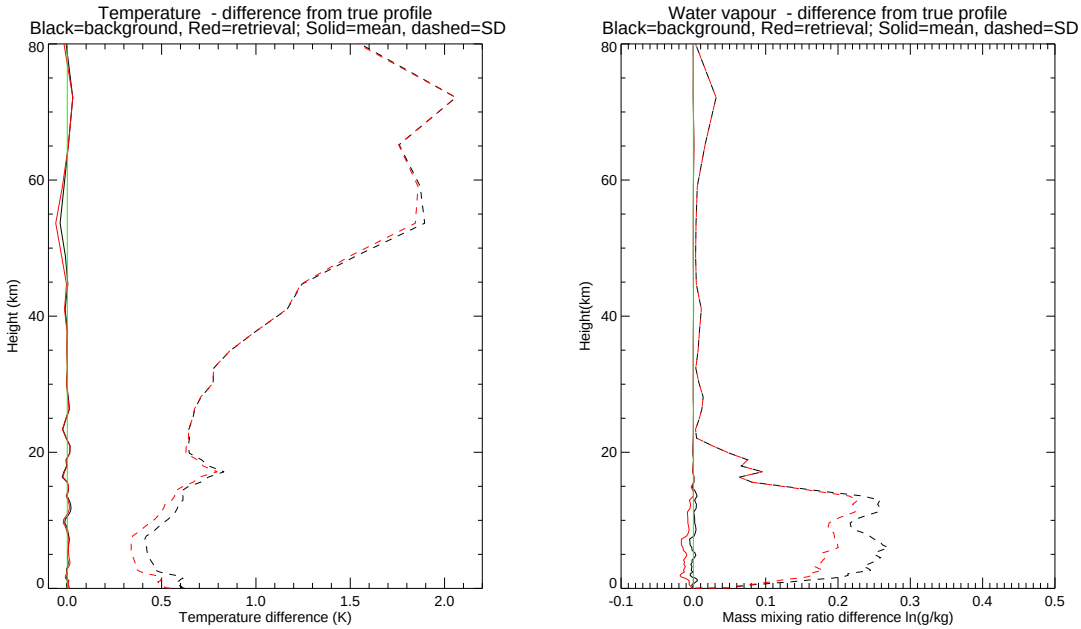
Figure 9.3: Mean and standard deviation of (Background–Truth) and (Analysis–Truth) for 4348 observations. Raw radiance Jacobians and $\mathbf{N}$ for the 210 reconstructed radiance channels (i.e. the reconstructed radiance channels, but raw radiances assimilated.)

is assimilated, rather than only bands 1 and 2.

PC-RTTOV was used in a simulated optimal 1D-Var scenario using 4348 observations simulated using RTTOV-10 converted to 200 PC scores. In this case, the instrument noise matrix used for assimilation was set to the values used for normalisation of the radiances by PC-RTTOV itself: doing this means that when the matrix is projected into PC space, it equals $\mathbf{I}$, simplifying calculations considerably. The results for the 4180 of these observations for which the 1D-Var converged are shown in figure 9.6.

Whilst the standard deviations of the errors on the analysis are mostly lower than the background errors (except at 72 km), the profiles show a pattern of oscillating bias that was not present in the background. The most likely mechanism for this is that the non-localised Jacobians generate increments on one level that are being counteracted by increments in the opposite direction on adjacent levels. Perhaps the most significant problem from an NWP perspective is a positive bias of 0.7 K in the temperature analysis in the UTLS region at 12–13 km (model levels 43–44).

These results can be compared to the assimilation of the 500 *PC-RTTOV* predictor channels: the forward modelled PC scores are linear combinations of the 500



(a) Temperature **(b)** Water Vapour

Figure 9.4: Mean and standard deviation of (Background–Truth) and (Analysis–Truth) for 4348 observations. Raw radiance Jacobians and $\mathbf{N}$ for the 138 4D-Var channels (i.e. raw radiances assimilated.)

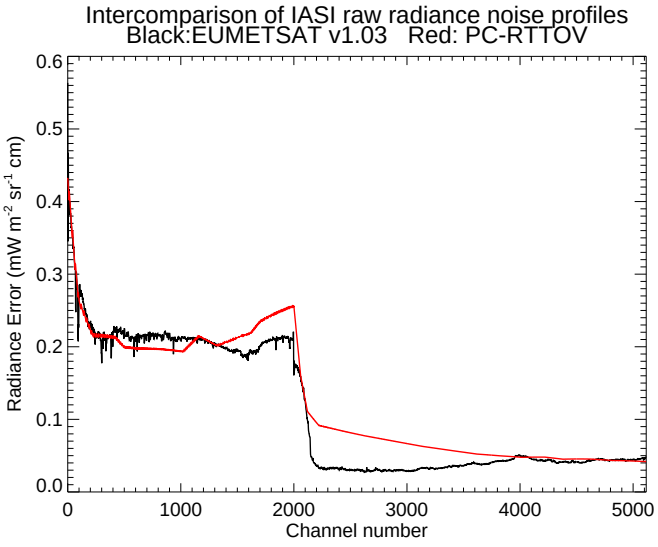
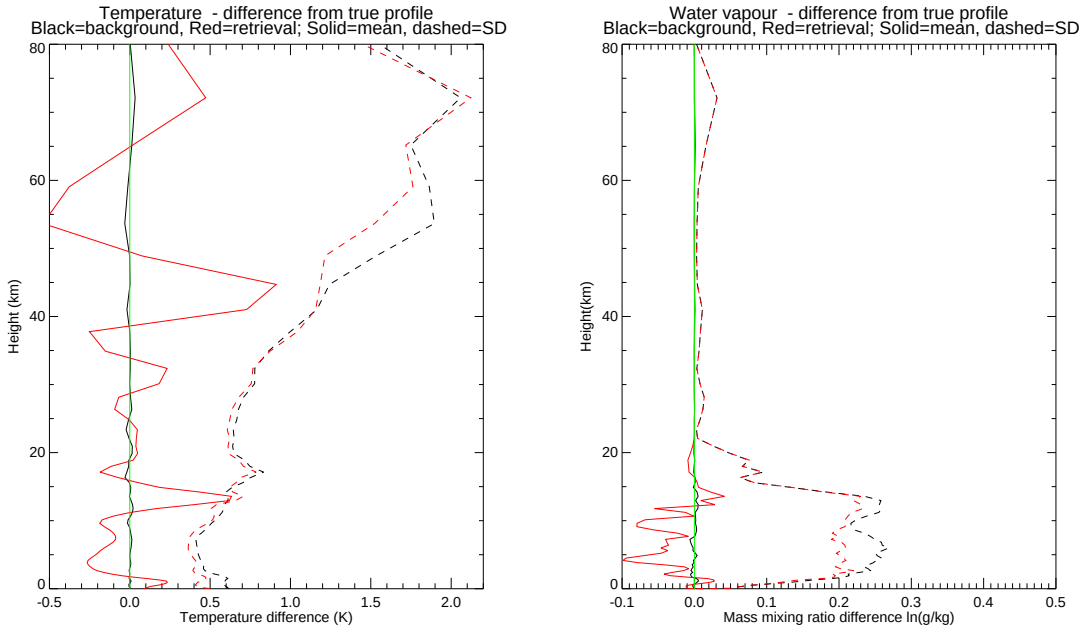


Figure 9.5: IASI Instrument noise for bands 1 and 2, as assumed by the EUMETSAT PC compression (black) and PC-RTTOV (red).

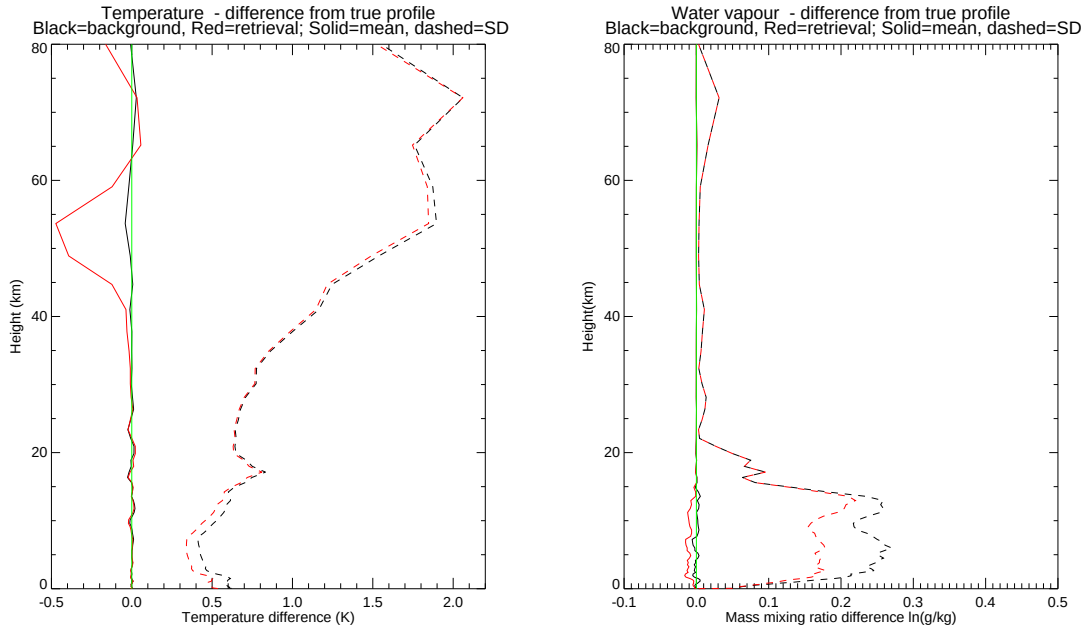


(a) Temperature **(b)** Water Vapour

Figure 9.6: Mean and standard deviation of (Background–Truth) and (Analysis–Truth) for 4180 of the total 4348 observations. PC-RTTOV PC Jacobians and $\mathbf{R}$ equal to the identity matrix.

channel radiances, so the simulated observations for these 500 channels are expected to contain the same information as the 200 PC scores. The results for the same 4180 observations can be seen in figure 9.7. The results in terms of improvement in standard deviation are reasonably consistent with those of the PC scores. The PC scores are slightly better in some areas (e.g. for temperature between 40 and 60 km, but are also less stable, for example the degradation of analysis relative to the background at 72 km. There is no oscillating bias in the channel-based system as seen with the PC scores, although there is a bias of -0.4 K between 45–55 km that is unexplained.

The oscillating bias pattern in figure 9.6 is very similar to that seen for the suboptimal-in-$\mathbf{H}$-and-$\mathbf{R}$ assimilation of reconstructed radiances (figure 9.1). In that case, the likely cause was overweighting of the observations due to neglect of the additional error, $\Delta\tilde{\mathbf{R}}$, that would counteract the use of the raw radiance Jacobian term. In this case, there should be no additional error of that kind that has been neglected. However, this explanation for the problems with the reconstructed radiance assimilation gives a clue for the cause of the oscillating bias in the PC assimilation: careful checking of the experiment showed that the set up was indeed suboptimal-in-$\mathbf{R}$. There was a neglected error term: the observations were simulated with RTTOV-10



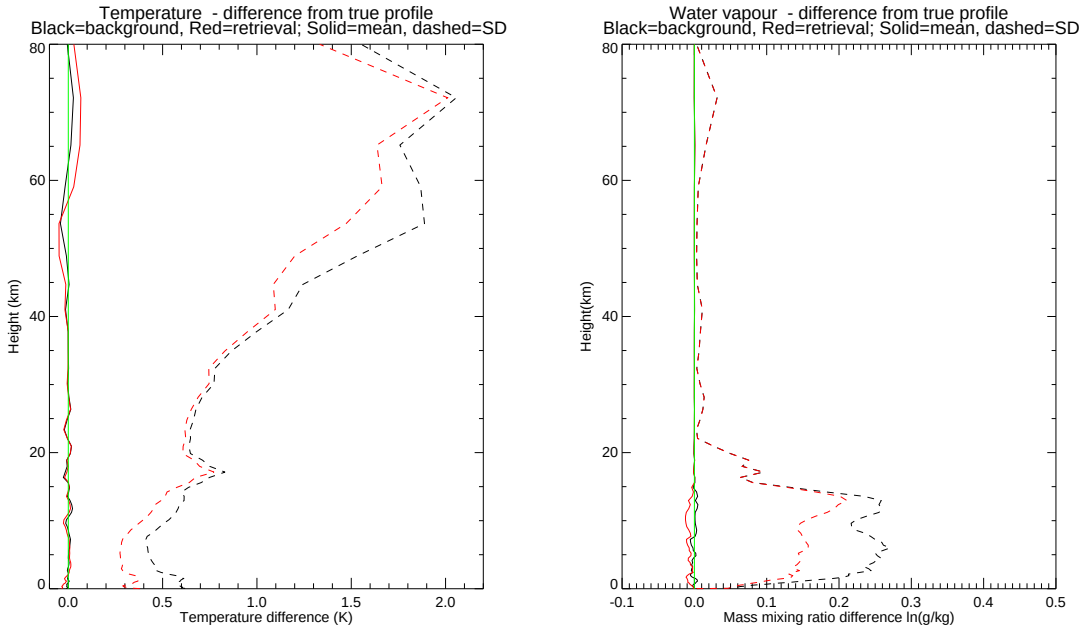
(a) Temperature **(b)** Water Vapour

Figure 9.7: Mean and standard deviation of (Background–Truth) and (Analysis–Truth) for 4180 of the total 4348 observations. PC-RTTOV Channel Jacobians for the 500 PC-RTTOV predictor channel radiances, and $\mathbf{R}$ equal to $\mathbf{L}_p\mathbf{L}_p^{\mathsf{T}}\mathbf{L}_p\mathbf{L}_p^{\mathsf{T}}$ for the PC-RTTOV noise.

rather than PC-RTTOV. This small discrepancy resulted in enough extra forward model error that the oscillating bias resulted. When PC-RTTOV is used to simulate the raw radiances before addition of error and conversion to PC scores, there is no bias in the analysis (figure 9.8) and the temperature errors are greatly reduced. This result is important, as it is a good test of what might be expected due to differences between any forward model and the real world.

The similarity in behaviour here demonstrates near equivalence of the two systems. In both cases, small discrepancies in the $\mathbf{R}$ matrix result in oscillatory biases in the analysis. In the case of PCs, the non-localised Jacobians mean that a small increment at one level is compensated for by increments higher and lower in the atmosphere. In the case of reconstructed radiances, the same result occurs through the significantly non-zero correlation structure of the $\mathbf{R}$ matrix.

The suboptimal-in-$\mathbf{R}$ results are included here because they illustrate the point made in the previous section: in this non-localised system, it is very important not to underestimate the observation errors. If the observation errors are too low, the background will be underweighted, and $\mathbf{B}$ will be unable to filter the oscillating increments correctly. It is important to reiterate that this result applies to PC-



(a) Temperature **(b)** Water Vapour

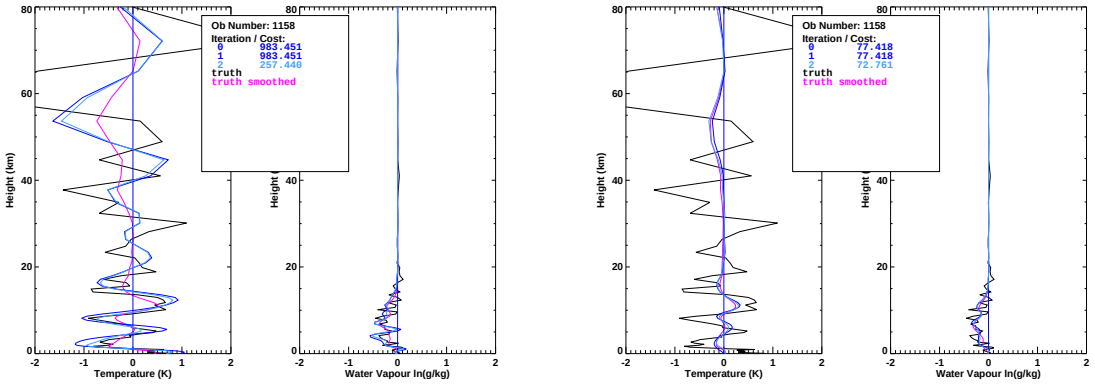
Figure 9.8: Mean and standard deviation of (Background–Truth) and (Analysis–Truth) for 4180 of the total 4348 observations. PC-RTTOV PC Jacobians and $\mathbf{R}$ equal to the identity matrix. Observations simulated with PC-RTTOV rather than RTTOV.

compressed radiances, whether in PC score form or reconstructed radiance form, but not to raw radiances. The effects of overweighting the observations in a raw radiance assimilation system are much less dramatic. This is likely to be because the lower correlations and more localised jacobians mean that the system is unlikely to generate increments with an oscillatory structure.

Empirical strengthening of the diagonal to counteract missing forward model error

It was found that, for PCs also, the underweighting of the background can to some extent be counteracted by strengthening the diagonal of the observation error covariance. An experiment to assimilate PC scores with instrument plus added forward model error (but still missing the additional error from the use of RTTOV for simulation and PC-RTTOV for assimilation), failed when the full error covariance matrix was used. In this system, the $\mathbf{R}$ matrix is non-diagonal because the forward model error does not normalise to $\mathbf{I}$.

Figure 9.9 shows that when a diagonal $\mathbf{R}$ was used instead, the oscillations were damped, but this improvement in bias came at the expense of the ability of the



(a) Instrument noise only

(b) Diagonal of instrument plus RT error matrix

Figure 9.9: Minimisation process for observation 1158 for PC score assimilation with and without RT error. The plotted lines show the difference between a given profile and the background profile. The black line is the background departure from truth, and the pink line is the background departure multiplied by the averaging kernel. The other coloured lines are the departure of the profile on a given numbered iteration.

observation to impart information to the analysis. In particular the temperature analysis is now very similar to the background. Note that the lack of information in the analysis is not a result of the inclusion of radiative transfer errors: the results for the *PC-RTTOV* predictor channels were very similar with and without radiative transfer errors included.

This result seems to contradict the findings of Collard, where using the diagonal of the error covariance matrix made the analysis worse than the background in parts of the profile. However, as will be shown in chapter 10, where a suboptimal-in-$\mathbf{B}$ system is presented, suboptimality does not always result in a degradation of the analysis relative to the background, but it does always result in a degradation relative to an optimal analysis.

9.1.3 Reconstructed radiances or PC scores?

The results of the instrument noise only experiment show that the highly non-diagonal nature of the reconstructed radiance $\tilde{\mathbf{R}}$ matrix can cause problems in a 1D-Var where raw radiance Jacobians are used and this additional forward model error is not accounted for. Where there is only instrument noise, principal component assimilation with an exact forward model does not suffer from this problem. In the real system, there is forward model error, and it is not exactly known, and this apparent advantage of PC score assimilation is shown not to hold when the

observation error term is not properly specified. In the case of incorrectly specified $\mathbf{R}$ matrices, both systems result in oscillatory patterns of bias in the analysis.

It is possible that in a standalone 1D-Var retrieval system, starting from a climatological background source, the precise specification of $\tilde{\mathbf{R}}$ is less critical. In most cases, the background errors would be larger, and there is less danger of underweighting real information that is contained in the background state. It is also likely that problems with the assimilation have been exacerbated here by the very tight definitions of error: an instrument noise only system is very unforgiving. Whilst it is certainly true that care is needed when using these highly correlated data, standalone retrievals that use PC compressed data, such as the EUMETSAT v6 Level 2 temperature and humidity retrievals, can be shown to be very competitive with raw radiance alternatives from sparse channel selections (Thomas August, personal communication). This implies that PC compression and radiance reconstruction are useful techniques to add information in an efficient manner over and above what can be extracted from a sparse channel selection of raw radiances.

The experiments here are for clear-sky cases only. Assimilation of PC scores in cloudy areas poses larger problems than assimilation of reconstructed radiances, because currently used methods of restricting the assimilation to channels that peak above the cloud cannot be used with such non-localised response to the atmosphere. Therefore reconstructed radiances, notwithstanding the problems inherent in specifying their observation error terms, represent a viable alternative.

9.2 Applicability of reconstructed radiance results to the real assimilation system

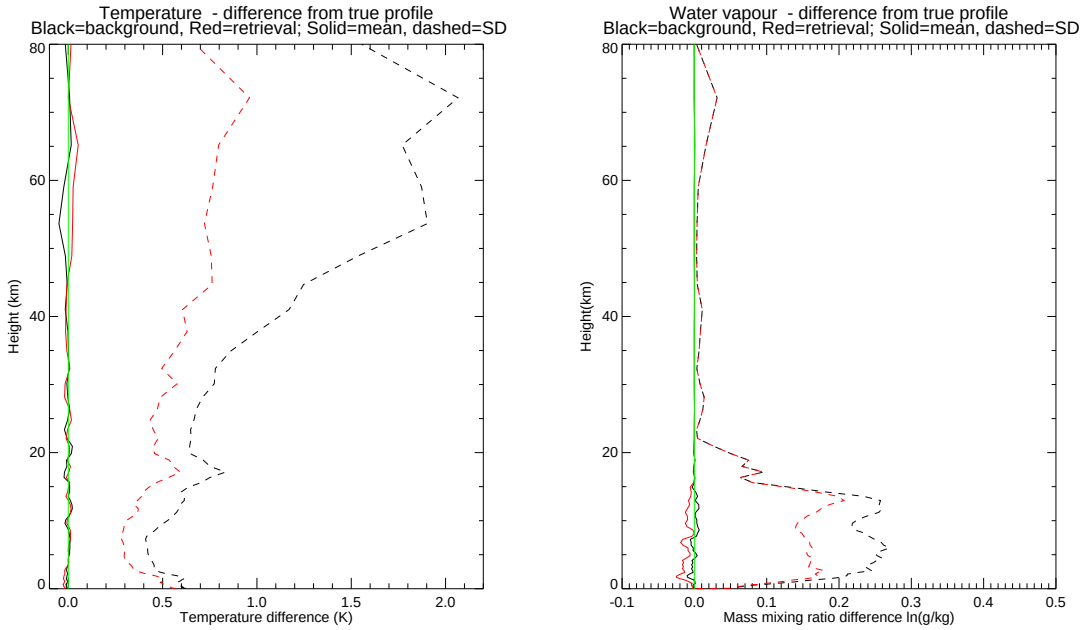
The results of the previous section are somewhat disappointing in terms of the likely benefit to be gained from assimilating reconstructed radiances or PC scores into an NWP system. Whilst a fully optimal system shows significant benefit from the noise reduction inherent in the compression scheme, it is clear that the $\mathbf{R}$ matrix must be very well specified in order to achieve this result in practice. If $\mathbf{R}$ does not properly take account of all sources of error in the system, the results can be very bad indeed. In a real assimilation system, one cannot hope to model all sources of error accurately, and estimating them well enough to prevent degradation to the analysis is a rather daunting task. This is likely to be more critical in an NWP system with very low background errors than in a retrieval with a climatological first guess.

Nevertheless, the $\mathbf{R}_{HL}$ matrix estimated in chapter 7 by the observational method is available, along with the channel selections made. It is, therefore, possible to look at the results of a 1D-Var assuming that $\mathbf{R}_{HL}$ is a good estimate of the true errors. In this case, simulated true brightness temperatures were converted to simulated observations by the addition of noise consistent with $\mathbf{R}_{HL}$ stabilised by incremental multiplication of the diagonal by a small factor until the condition number was reduced to 10^6. This set-up is a little different from those already examined, where noise is added to a radiance profile rather than a brightness temperature profile. Additionally, the observations were not explicitly converted to reconstructed radiance space; rather it was assumed that the diagnosed $\mathbf{R}_{HL}$ represents the departures of the reconstructed brightness temperatures from the true brightness temperatures.

The results in this case are very promising: figure 9.10 shows the overall performance of the 1D-Var with the *RHL-M2-8J-1.3* channels, while figure 9.11 shows the profile increments from this system for the same three observations as shown in figure 9.2. The background errors are reduced considerably by this assimilation system. In fact, the results of the 1D-Var are considerably better, for temperature at least, than predicted by the optimal linear equations (compare figure 9.10 with figure 8.6). The water vapour results agree better with the prediction. The reason for this discrepancy has not been investigated yet, but it could indicate the pseudo-information is being added to the analysis. However, the particularly encouraging result is that no oscillating bias is seen in the solution. If $\mathbf{R}_{HL}$ does represent the observation errors to a reasonable level, these results are a good starting point for further investigations in an operational context.

In terms of speed of minimisation, the number of iterations taken is greater on average than for the instrument noise only systems; performance with the Levenberg-Marquardt minimiser is better, in that only 22 of the observations failed to converge in comparison with 732 for the Newtonian solver (figure 9.12), but the overall fit of analysis to truth for converged observations were almost identical.

As mentioned at the start of this chapter, it is difficult to perform a meaningful intercomparison with raw radiances for the *RHL-M2-8J-1.3* channel selection, because there is currently no raw radiance equivalent to $\tilde{\mathbf{R}}'$. The only comparison possible for the $\mathbf{R}_{HL}$ system is to use the operational 4D-Var *VAR* channel selection and Desroziers error covariance matrix. In a simulated 1D-Var context, this system does not perform very well in terms of incorporating observational information to the analysis (figure 9.13). The adjusted Desroziers matrix is very cautious, and necessarily so as it is very dangerous to destabilise the operational 4D-Var. This



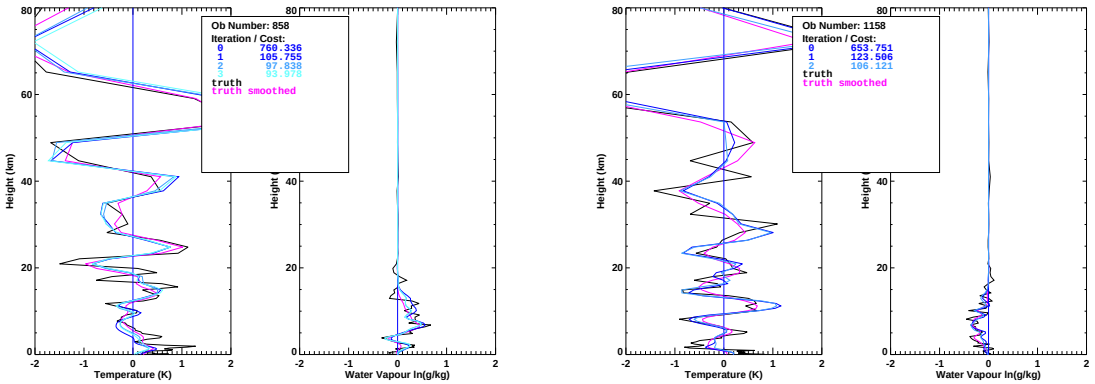
(a) Temperature **(b)** Water Vapour

Figure 9.10: Mean and standard deviation of (Background–Truth) and (Analysis–Truth) for 4348 observations. Raw brightness temperature Jacobians and the derived $\mathbf{R}$ matrix from the observational method, for the *RHL-M2-8J-1.3* channels chosen to work with this matrix.

serves as a reminder that the 1D-Var results for the new channel selection presented in figure 9.10 are extremely optimistic: in the real assimilation system, the error variances will almost certainly need to be increased significantly in order to reduce the number of iterations required in 4D-Var to an acceptable level. Nevertheless, it is a very good result that this empirical observation error covariance matrix and new channel selection does not introduce an oscillating bias and seems to allow the extraction of significant extra information from the observations.

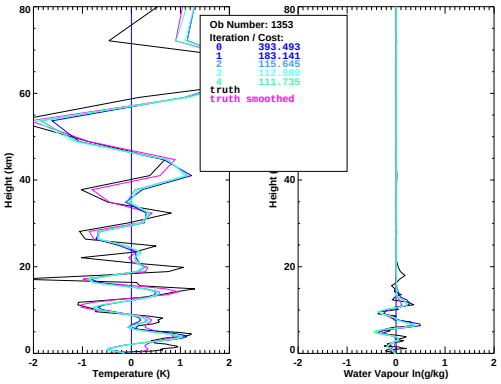
9.3 Summary

The performance of an instrument noise only 1D-Var was assessed for a suboptimal-in-$\mathbf{H}$-and-$\mathbf{R}$ assimilation of reconstructed radiances. Without significant manipulation of the $\mathbf{R}$ to counteract the missing error term that comes from using an incorrect forward model, the 1D-Var performance is very poor with most observations failing to reach covergence. Shrinking the $\mathbf{R}$ matrix towards a diagonally-strengthened matrix does allow an analysis to be calculated in most cases, but oscillating biases are produced. The performance of a similar system is assessed for the assimilation



(a) Ob 858

(b) Ob 1158



(c) Ob 1353

Figure 9.11: Minimisation process for observations 858, 1158 and 1353 for *RHL-M2-8J-1.3* channels assuming $\mathbf{R} = \mathbf{R}_{HL}$. The plotted lines show the difference between a given profile and the background profile. The black line is the background departure from truth, and the pink line is the background departure multiplied by the averaging kernel. The other coloured lines are the departure of the profile on a given numbered iteration.

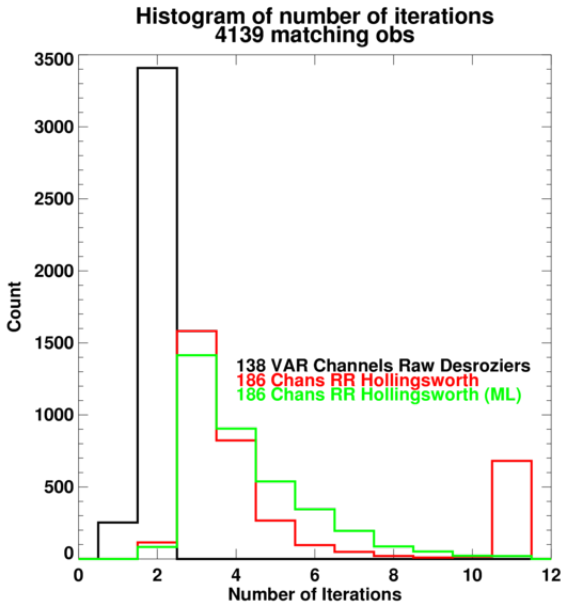
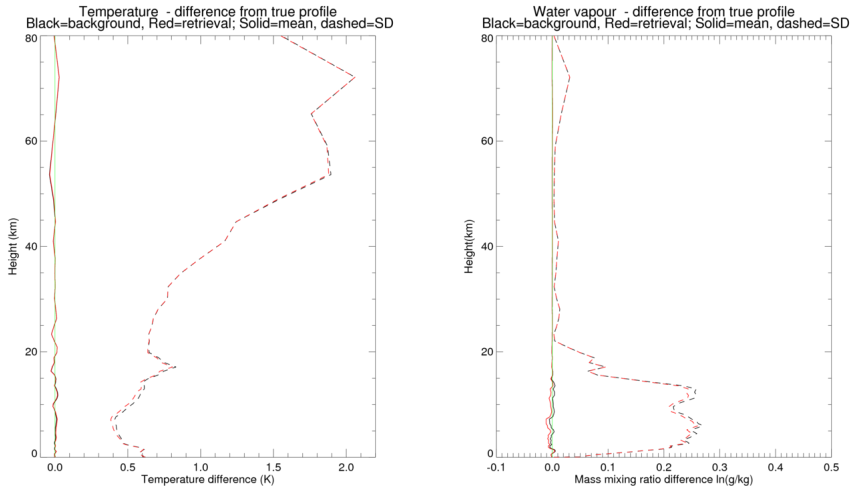


Figure 9.12: Number of iterations taken to convergence for the minimisation comparing the *RHL-M2-8J-1.3* channel selection with Newtonian and Levenberg-Marquardt minimisation methods, and the *VAR* channel selection with Newtonian minimisation. The block of observations at 11 iterations for the Newtonian minimisation scheme in red have not converged (the maximum number of iterations allowed is 10).



(a) Temperature **(b)** Water Vapour

Figure 9.13: Mean and standard deviation of (Background–Truth) and (Analysis–Truth) for 4348 observations. Raw brightness temperature Jacobians and the Desroziers $\mathbf{R}$ matrix used operationally for the 138 4D-Var channels.

of PC scores. This system is quite well behaved as long as all sources of errors are taken into account, but if there is a missing error term, similar oscillating biases result.

The mechanism of production of the oscillating biases is slightly different in each case: for reconstructed radiances it is spurious correlations that generate them, whereas for PC score assimilation it is the non-localised nature of the Jacobians that gives rise to such features. These results suggest that much care needs to be taken when specifying error covariances for use with compressed spectra.

If the error terms are well-enough understood the PC 1D-Var performs well. The preferred channel selection designed for operational implementation at the Met Office, *RHL-M2-8J-1.3*, was also tested in a simulated 1D-Var environment, and gave good results without any oscillating bias. Given that there are difficulties with assimilating PC scores in cloudy areas, these results provide encouragement for further testing of reconstructed radiances within the operational assimilation system.

Chapter 10

Suboptimal-in-B Assimilation

The previous chapters have used idealised 1D-Var systems to investigate the impact of using more of the IASI spectrum. In the experiments presented, it was assumed that the background errors were well described. In reality, the process of estimation of the background error, and our inability to properly model its synoptic evolution (as described in chapter 3) mean that the **B** term will inevitably be misspecified. Furthermore, although parts of the **U** variable transform (see section 3.5.3) vary with latitude, the structure functions are defined globally, meaning that the error covariance structure may not be correct regionally. Hybrid DA methods are now used to include flow dependence into the formulation, but the analysis still relies heavily on the static climatological covariance formulation. Given that the errors are only estimated, they are inevitably misspecified, and the analysis will no longer be optimal. This scenario is referred to as suboptimal-in-**B**. It is the properties of the assumed **B** matrix that define the subspace in which analysis increments can be made, even if the assumed **B** is not the same as the true **B**.

This chapter extends the linear analysis equations presented in section 3.2 to include this suboptimality. Results are shown for a scalar case to identify regimes in which the suboptimal-in-**B** analysis in fact has larger errors than the true background state. The suboptimal-in-**B** analysis is then discussed for IASI, firstly with the *VAR* channel selection and then with the full spectrum. It is important to understand how robust our assimilation system is to errors in the specification of **B**.

10.1 Suboptimal linear analysis: the scalar case

3.8 to 3.12, the optimal analysis equations can be extended to take account of misspecification of **B**. The subscript T now identifies the true background error covariance, $\mathbf{B}_T$, and A identifies the assumed covariance matrix, $\mathbf{B}_A$. The extended theory was presented in Eyre and Hilton (2013).

The optimal analysis error $\mathbf{A}_{opt}(\mathbf{B}_T)$, where $\mathbf{B}_T$ is known is given by:

$$\mathbf{A}_{opt}(\mathbf{B}_T) = (\mathbf{I} - \mathbf{KH})\mathbf{B}_T(\mathbf{I} - \mathbf{KH})^{\mathrm{T}} + \mathbf{KRK}^{\mathrm{T}} \tag{10.1}$$

This is equivalent to equation 3.12. In the real DA system, the analysis weights of equation 3.10, now denoted $\mathbf{K}_{\mathbf{B}_A}$, are calculated using $\mathbf{B}_A$, so the true analysis error, $\mathbf{A}$, is:

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}_{\mathbf{B}_A}\mathbf{H})\mathbf{B}_T(\mathbf{I} - \mathbf{K}_{\mathbf{B}_A}\mathbf{H})^{\mathrm{T}} + \mathbf{K}_{\mathbf{B}_A}\mathbf{R}\mathbf{K}_{\mathbf{B}_A}^{\mathrm{T}} \tag{10.2}$$

This equation is linear in $\mathbf{B}_T$, though non-linear in $\mathbf{B}_A$, which appears as part of $\mathbf{K}_{\mathbf{B}_A}$.

Equation 10.2 can be split into a component representing the expected optimal analysis if $\mathbf{B}_A$ were correct (i.e. $\mathbf{A}_{opt}(\mathbf{B}_A)$), and a misspecification sensitivity component:

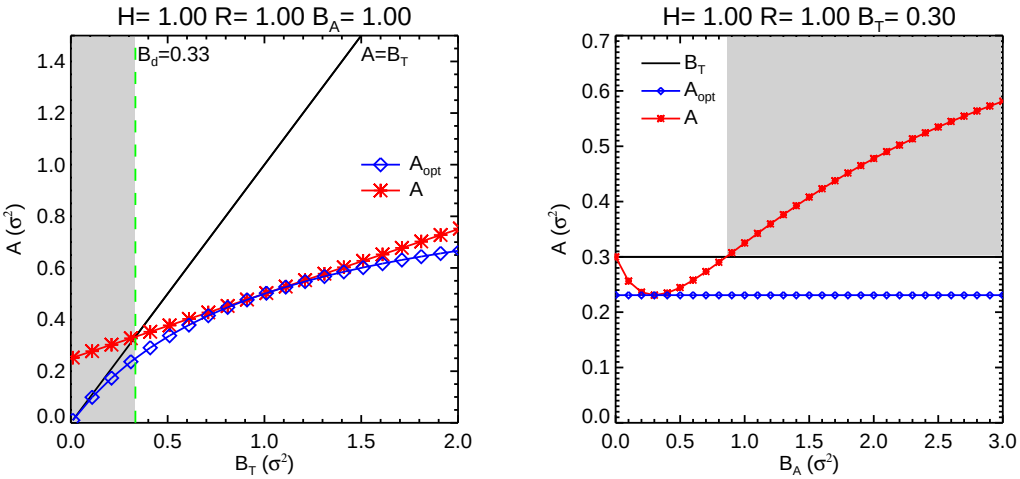
$$\begin{aligned}\mathbf{A} = &\underbrace{(\mathbf{I} - \mathbf{K}_{\mathbf{B}_A}\mathbf{H})\mathbf{B}_A(\mathbf{I} - \mathbf{K}_{\mathbf{B}_A}\mathbf{H})^{\mathrm{T}} + \mathbf{K}_{\mathbf{B}_A}\mathbf{R}\mathbf{K}_{\mathbf{B}_A}^{\mathrm{T}}}_{\mathbf{A}_{opt}(\mathbf{B}_A)} + \\ &\underbrace{(\mathbf{I} - \mathbf{K}_{\mathbf{B}_A}\mathbf{H})(\mathbf{B}_T - \mathbf{B}_A)(\mathbf{I} - \mathbf{K}_{\mathbf{B}_A}\mathbf{H})^{\mathrm{T}}}_{\text{misspecification sensitivity}}\end{aligned} \tag{10.3}$$

Note that $\mathbf{A}_{opt}(\mathbf{B}_A)$ is the quantity that would normally be calculated to predict the performance of a retrieval or assimilation.

As a first step to understanding the sensitivity of the analysis system to misspecification of the background error in general terms, a scalar system was examined, with only one state variable, and one observation with $\mathbf{H} = \mathbf{I}$. For a fixed $\mathbf{B}_A$, $\mathbf{B}$ and $\mathbf{R}$ were varied to test the sensitivity of $\mathbf{A}$.

Figure 10.1(a) shows how the true, i.e. suboptimal, error $\mathbf{A}$ (in red), and $\mathbf{A}_{opt}(\mathbf{B}_T)$ (in blue) vary over a range of $\mathbf{B}_T$. $\mathbf{A}_{opt}$ represents the analysis error that would have been achievable if $\mathbf{B}$ were correctly specified. $\mathbf{A}$ is always bigger than $\mathbf{A}_{opt}$, except where $\mathbf{B}_A = \mathbf{B}_T$. When $\mathbf{B}_A < \mathbf{B}_T$, the observations are underweighted in the analysis, and when $\mathbf{B}_A > \mathbf{B}_T$, useful background information is discarded by the analysis. The black line marks $\mathbf{A} = \mathbf{B}_T$, and a danger zone (Eyre and Hilton, 2013) is defined where $\mathbf{A} > \mathbf{B}_T$, in other words where the analysis is degraded through the misspecification of $\mathbf{B}_A$ which results in the underweighting of the background. The danger zone is marked by the shaded area to the left of the vertical red dashed line. The danger zone occurs for extreme values of $\mathbf{B}_A > \mathbf{B}_T$: the value of $\mathbf{B}_T$ at which the analysis enters the danger zone, $\mathbf{B}_d$, is given by setting $\mathbf{A} = \mathbf{B}_T$ in equation 10.3.

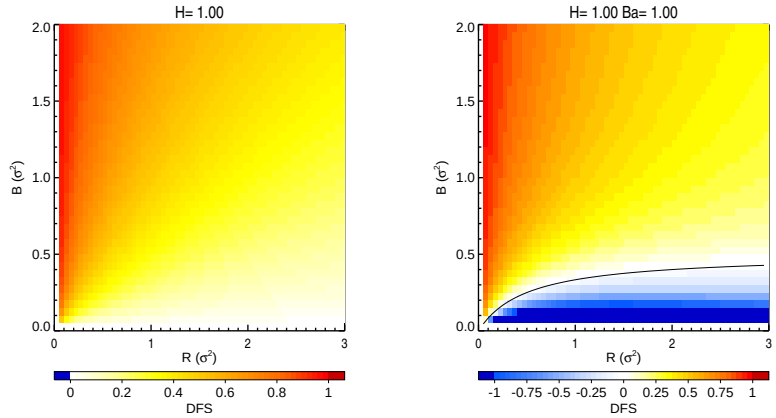
$$\mathbf{B}_d = \mathbf{R}\mathbf{B}_A/(\mathbf{H}^2\mathbf{B}_A + 2\mathbf{R}) \tag{10.4}$$



(a) Analysis error variance, $\mathbf{A}$, as a function of the true background error variance, $\mathbf{B}_T$, with observation error variance, $\mathbf{R} = 1.0$ and $\mathbf{B}_A = 1.0$

(b) Analysis error variance, $\mathbf{A}$, as a function of the assumed background error variance, $\mathbf{B}_A$, with observation error variance, $\mathbf{R} = 1.0$ and $\mathbf{B}_T = 0.3$

Figure 10.1: Properties of the suboptimal scalar analysis system. (a) shows how the suboptimal analysis error ($\mathbf{A}$ in red) and the optimal analysis error ($\mathbf{A}_{opt}(\mathbf{B}_T)$ in blue) vary with $\mathbf{B}_T$ for $\mathbf{B}_A = 1.0$ and $\mathbf{R} = 1.0$. Where $\mathbf{A}$ is larger than $\mathbf{B}_T$ (i.e. where it crosses the line $\mathbf{A} = \mathbf{B}_T$) the analysis enters the danger zone (marked in grey). The value of $\mathbf{B}_T$ at which $\mathbf{A}$ enters the danger zone is $\mathbf{B}_d$, marked with a red dashed line. (b) shows how $\mathbf{A}$ (in red) varies with $\mathbf{B}_A$ for $\mathbf{B}_T = 0.3$ and $\mathbf{R} = 1.0$. Again, the danger zone, where $\mathbf{A} > \mathbf{B}_T$, is marked in grey, which in this case occurs when $\mathbf{B}_A > 0.87$.



(a) DFS Optimal **(b)** DFS Suboptimal

Figure 10.2: Optimal and suboptimal maps of Degrees of Freedom for Signal (DFS) with varying $\mathbf{B}_T$ and $\mathbf{R}$, for $\mathbf{B}_A = 1$. The black line on plot (b) represents the point at which the analysis enters the danger zone: below this line the DFS becomes negative.

Figure 10.1(b) shows the same calculations but this time for a fixed $\mathbf{B}_T = 0.3$ with varying $\mathbf{B}_A$. The suboptimal analysis is in the danger zone when $\mathbf{B}_A$ is >0.87, in this case about 3 times $\mathbf{B}_T$.

Figure 10.2 illustrates the effect of the suboptimal analysis on the information content measures described in chapter 4, for the scalar system of figure 10.1(a) with $\mathbf{B}_A$ =1.0 and $\mathbf{H}$ =1.0. In the suboptimal case, it is necessary to decide whether to calculate DFS relative to $\mathbf{B}_A$ or $\mathbf{B}_T$ (i.e. Tr $(\mathbf{I} - \mathbf{A}\mathbf{B}_A^{-1})$ or Tr $(\mathbf{I} - \mathbf{A}\mathbf{B}_T^{-1})$). It is more relevant to calculate relative to $\mathbf{B}_T$: although $\mathbf{B}_A$ filters the observation information, it is $\mathbf{B}_T$ that describes the true information content of the background forecast. This suboptimal version of DFS becomes negative when the analysis enters the danger zone. This is shown in figure 10.2(b), where the black line marks the variation of $\mathbf{B}_d$ with $\mathbf{B}_T$ and $\mathbf{R}$.

This simple case illustrates that it is best for $\mathbf{B}_A$ not to overestimate the variance of $\mathbf{B}_T$ for any given scalar quantity. This is perhaps counterintuitive, and Daley (1991) states that "if a background is to be used in the objective analysis, it is wiser to overestimate the expected background error variance than to underestimate it" in order to avoid the case that, where $\mathbf{B}_T >> \mathbf{R} >> \mathbf{B}_A$, "the analysis would have been more accurate if the background had been completely ignored". However, completely ignoring the background is impossible in an under-determined problem such as the analysis of a global NWP system as, even with perfect observations the background would be required to fill in the gaps between them. In any case, figure 10.2(b) shows that, regardless of the relative values of $\mathbf{R}$ and $\mathbf{B}_A$, it is always possible for the analysis to be worse than the forecast background if $\mathbf{B}_A >> \mathbf{B}_T$

unless $\mathbf{R}$ is vanishingly small. The black curve marking the danger zone flattens off towards a value of $\mathbf{B}_T = 0.5$ quite quickly: for a reasonable range of $\mathbf{R}$, if $\mathbf{B}_A$ is between 2 and 4 times larger than $\mathbf{B}_T$, the analysis will enter the danger zone. (It seems unlikely that for a realistic application $\mathbf{B}_A$ could overestimate $\mathbf{B}_T$ by more than 4 times but the analysis would still be in the danger zone if it were.)

10.2 Suboptimal linear analysis for Met Office assimilation of IASI

Daley's statement is from an era before NWP centres ran 4D-Var and models were less accurate. Most DA experts would no longer aim to intentionally 'err on the side of caution' in this manner (Andrew Lorenc, pers. comm.). It is certainly of interest, however, to know whether the way in which $\mathbf{B}_A$ is specified (e.g. **MoistCov** or **NMC**) is likely to lead to a situation where the use of accurate observations can degrade the analysis relative to the background. It is also important to understand whether increasing the weight given to IASI observations by using more of the spectrum – as is the aim of this thesis – or decreasing observation errors would make this more or less likely to occur.

This section uses the suboptimal analysis equation 10.3 to examine the analysis of a 1D temperature and humidity profile, for the assimilation of the 138 *VAR* channel subset of IASI (appendix E), assuming the background error properties of **MoistCov** and **NMC**.

10.2.1 Suboptimal analysis in eigenvector space

The prediction of whether the analysis will enter the danger zone for a scalar quantity can be applied approximately to the full system, despite the significant vertical correlations now present, if the eigenvector space of $\mathbf{B}_A$ is used. This projection was shown for the optimal analysis in figures 5.2 and 5.4. Eigenvector decomposition renders modes in which the background errors are uncorrelated, and so the results of the scalar analysis, in terms of the relationship between $\mathbf{B}_A$ and $\mathbf{B}_T$ and the likelihood of entering the danger zone, apply more directly to each mode. $\mathbf{V}$ defines the orthogonal modes of $\mathbf{B}_A$, such that $\mathbf{B}_A = \mathbf{V}\Lambda\mathbf{V}^{\mathrm{T}}$. The optimal and suboptimal analysis error can be projected onto these eigenvectors via pre- and post-multiplication by $\mathbf{V}$, and examination of the diagonal terms of the projected matrix shows in which modes the analysis variance is improved over the background, and in which it

is degraded. It is interesting to establish the number of eigenvectors that enter the danger zone, as well as the extent to which the analysis is degraded, as any method of increasing the overall information content should not decrease robustness toward the misspecification of $\mathbf{B}_A$.

The eigenvectors for **MoistCov** and **NMC** were discussed in section 3.8.1 and are shown in appendices B and C. The total number of eigenvectors is equal to the number of levels, 70 each for temperature and water vapour, but not all of those are significant. Where the associated eigenvalue of each mode is very small, it is not possible for an observation to have much impact on the analysis in that mode because the variational analysis assumes that the background error is very low. Furthermore, IASI may not be sensitive to some of the spatial scales present in the $\mathbf{B}_A$ matrix. The normalised Jacobian method of Rodgers (2000) described in section 4.2 can be adapted to look at each eigenvector of $\mathbf{B}_A$ in turn. Single value decomposition of the normalised Jacobian (equation 4.3), where $\mathbf{B}_A$ is replaced by each of its eigenvectors in turn, indicates that only approximately the lowest 30–35 ranked eigenvectors represent signals that are measurable by the full IASI spectrum to within observation error. In fact, it would not be desirable for observations to exert significant impact on the lower order eigenvectors, as they typically describe small-scale features that are not so well captured by the statistical estimation of $\mathbf{B}$ (Mike Cullen, pers. comm.). Only the first 35 eigenvectors are shown in the plots that follow, and higher-order vectors entering the danger zone are considered insignificant.

10.2.2 True and assumed B pairings

In order to investigate the suboptimal-in-**B** analysis, experimental pairings of $\mathbf{B}_T$ and $\mathbf{B}_A$ are required. Given that $\mathbf{B}_T$ is unknown, it is necessary to explore the possible ways in which $\mathbf{B}_A$ may be incorrectly defined. The two realisations of the Met Office 4D-Var background errors, **NMC** and **MoistCov** are used to construct these experimental pairings: the differences between these matrices represent the actual changes of the covariance structure that were implemented over the course of 18 months of development of the Met Office. We may therefore expect similar changes over the coming years, and these changes may affect the way that IASI observations provide information to the forecast system. We would hope that the impact of IASI would not be unduly changed by altering the covariance statistics, so it is important that the analysis is robust to these differences. However, it should be noted that although the method of gathering the underlying data used to construct

these matrices is different (i.e. NMC method vs ensemble of forecasts) and the moisture control variable is different, these matrices are constructed with the same covariance model and so the difference between them does not capture all aspects of the differences between a modelled and true covariance pair.

MoistCov and **NMC** differ in terms of their correlation and variance components (and consequently in their eigenvectors and eigenvalues). Various intermediate scenarios are possible, for example where the variances are wrong but the correlations correct and vice versa. Considered here is another intermediate, $\mathbf{B}'$, which is relevant to the question of whether the vertical scales of the background error are correctly represented:

$$\mathbf{B}' = \mathbf{V}\mathbf{M}\mathbf{V}^{\mathrm{T}} \tag{10.5}$$

where $\mathbf{M}$ is a diagonal matrix equal to the diagonal elements of the matrix $\mathbf{V}^{\mathrm{T}}\mathbf{B}_T\mathbf{V}$, recalling that $\mathbf{V}$ defines the eigenvector space of $\mathbf{B}_A$. In other words, $\mathbf{B}_T$ is projected onto the eigenvectors of $\mathbf{B}_A$ and $\mathbf{M}$ represents the variances of $\mathbf{B}_T$ associated with each vector. The correlations between the vectors are ignored. If **NMC** is used to define $\mathbf{V}$, then $\mathbf{B}'$ differs from **NMC** in that the variances associated with each vector are different, and $\mathbf{B}'$ differs from **MoistCov** in that the eigenvector structure is different but the eigenvalues are the same. The correlation structure of **MoistCov** compared with $\mathbf{B}'$ (i.e. the part of $\mathbf{B}_T$ that is ignored in constructing $\mathbf{B}'$) is shown in figure 10.3, where $\mathbf{V}^{\mathrm{T}}\mathbf{NMC}_{corr}\mathbf{V}$, is plotted. Although there are non-zero off-diagonal terms, the largest correlations are fairly local, suggesting that **MoistCov** and **NMC** differ more in terms of their eigenvalues than their eigenvectors.

Four pairs of $\mathbf{B}$ and $\mathbf{B}_A$ have been used to explore the range of possible suboptimality scenarios using equation 10.3 and subsequent projection onto the eigenvectors of $\mathbf{B}_A$:

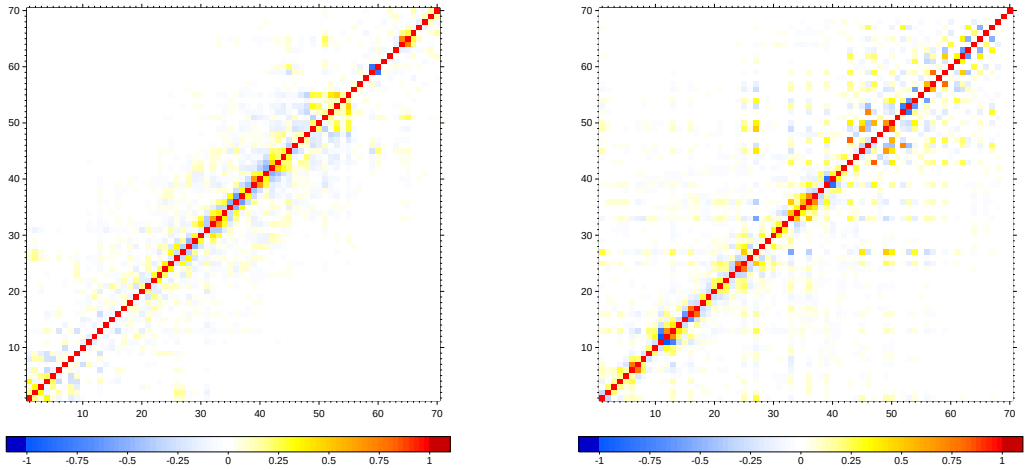
Case 1: Incorrect variance; $\mathbf{B}_T$ = **MoistCov**, $\mathbf{B}_A$ = **MoistCov** with variance $\times 2$

Case 2: Incorrect eigenvalues; $\mathbf{B}_T = \mathbf{B}'$, $\mathbf{B}_A$ = **NMC**

Case 3: Incorrect matrix; $\mathbf{B}_T$ = **MoistCov**, $\mathbf{B}_A$ = **NMC**

Case 4: Incorrect matrix; $\mathbf{B}_T$ = **NMC**, $\mathbf{B}_A$ = **MoistCov**

The suboptimal linear analysis equation 10.3 is used with the *VAR* channel selection for each of these cases, using the set-up described in section 3.8.3 and used in chapter 5, differing only in the inclusion of $\mathbf{B}_A$. As in chapter 5, the results are presented in profile and in eigenvector space. In profile space, the analysis error minus the true background error is plotted. Negative values represent an improvement over



(a) Temperature **(b)** Water Vapour

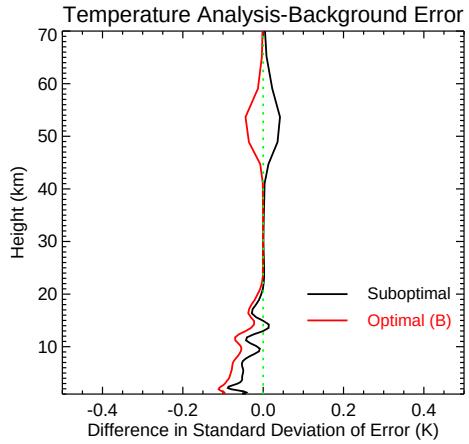
Figure 10.3: Correlations of the **NMC B** matrix projected onto the eigenvectors of the **MoistCov B** matrix

the background and positive values a degradation (i.e. the danger zone is breached). The optimal analysis, $\mathbf{A}_{opt}(\mathbf{B}_T)$ always gives a negative difference.

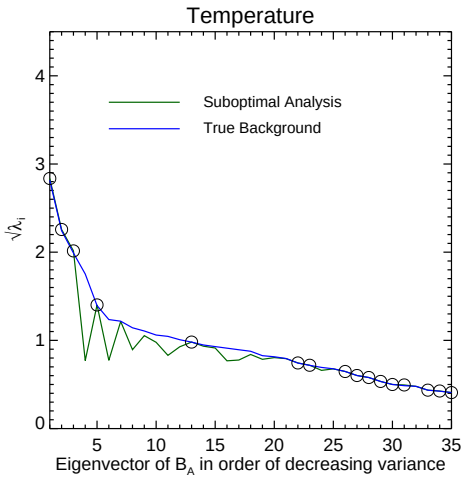
In eigenvector space, the true background error and suboptimal analysis error are projected onto the eigenvectors of $\mathbf{B}_A$, as it is this that defines the analysis modes rather than $\mathbf{B}_T$. For the plots in eigenvector space (for example figure 10.4(b)), where the suboptimal analysis line plots at a higher value of $\sqrt{\lambda_i}$ than $\mathbf{B}_T$ does, the danger zone is breached: such vectors are marked with circles. It should be noted that where the suboptimal analysis line plots exactly on top of $\mathbf{B}_T$, this represents the case where the analysis equals the background; where the suboptimal analysis is very close to the $\mathbf{B}_T$ line, this does not indicate that the danger zone is close to being breached, merely that there is not much information from the observations for that vector.

Case 1: Incorrect variance; B_T = MoistCov, B_A = MoistCov with variance $\times 2$

In this case, the variance of $\mathbf{B}_A$ is too high. Figure 10.4(a) shows the standard deviation of the optimal analysis error, $\mathbf{A}_{opt}(\mathbf{B}_T)$, minus the standard deviation of $\mathbf{B}_T$ in red, and the standard deviation of the suboptimal analysis, $\mathbf{A}$, minus the standard deviation of $\mathbf{B}_T$ in black. The scalar analysis of figure 10.1 shows that the danger zone may be breached where $\mathbf{B}_A$ is at least 2–4 times larger than $\mathbf{B}_T$ (the value depends on the relative size of $\mathbf{B}_A$ and $\mathbf{R}$, and on $\mathbf{H}$ as well). Here, the



(a) Analysis error minus background error. $\mathbf{A}_{opt}(\mathbf{B}_T)$ in red, suboptimal $\mathbf{A}$ in black.



(b) Analysis error and true background error projected onto eigenvectors of $\mathbf{B}_A$. $\mathbf{B}_T$ in blue, $\mathbf{A}$ in green. Eigenvectors marked with circles have entered the danger zone.

Figure 10.4: Suboptimal analysis for temperature for Case 1 where $\mathbf{B}_T$ = **MoistCov** and $\mathbf{B}_A$ has twice the variance of $\mathbf{B}_T$. *VAR* channels. $\mathbf{R}$ contains instrument noise only.

variances of $\mathbf{B}_A$ are set to twice those of $\mathbf{B}_T$, and the danger zone is only breached in the upper stratospheric part of the temperature profile, and at about 15 km. In the lower atmosphere, the retrieval still improves upon the background, although by a smaller amount than in the optimal case. The water vapour analysis error does not enter the danger zone and is not shown.

When shown in the eigenvector space of $\mathbf{B}_A$ (figure 10.4(b)), many vectors have entered the danger zone: despite only a small portion of the retrieval in profile space being degraded, the analysis is actually worse than the background across many spatial scales in the vertical. Where the vectors contain a substantial component from the stratosphere, they are likely to be in the danger zone, and where they are primarily responsive to the lower atmosphere they are improved, in keeping with what is seen in figure 10.4(a). The fact that all the variances of $\mathbf{B}_A$ are greater than those of $\mathbf{B}$, but the retrieval is not degraded across the full profile (and in fact is still improved near the surface) indicates the complex response of the retrieval to the observation.

Case 2: Incorrect eigenvalues; $\mathbf{B}_T = \mathbf{B}'$, $\mathbf{B}_A$ = NMC

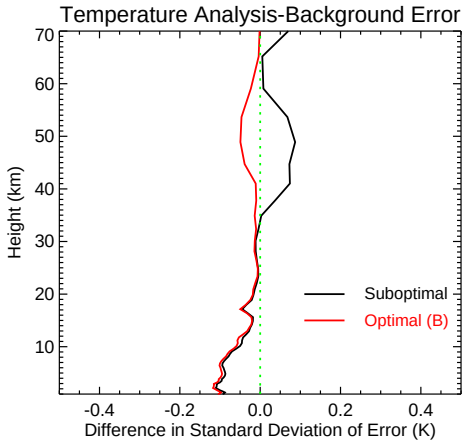
This matrix pairing is more interesting in terms of understanding the relative importance of errors in structure and variance in $\mathbf{B}_A$. $\mathbf{B}_T$ generally has smaller variances

than $\mathbf{B}_A$, and its eigenvalues are also generally smaller, but the eigenvectors of the two matrices are the same, so some aspects of the retrieval might be expected to enter the danger zone. The results, shown in figure 10.5, demonstrate that only a few temperature vectors are degraded, corresponding to strong stratospheric structures. In the lower atmosphere the analysis is virtually indistinguishable from optimal. For water vapour, the analysis enters the danger zone around the tropopause and in the high stratosphere. Many vectors are slightly degraded (although by such a small amount that the lines plot on top of each other for much of figure 10.5(d)). Those that show significant degradation are vectors with large stratospheric components in the **NMC** matrix. These components are not present in the **MoistCov** matrix, by design, as it was features such as these that the new control variable was designed to suppress (hence the marked dips in the eigenvalue spectrum of $\mathbf{B}_T$ projected onto $\mathbf{B}_A$). This issue is explored further in section 10.2.3.

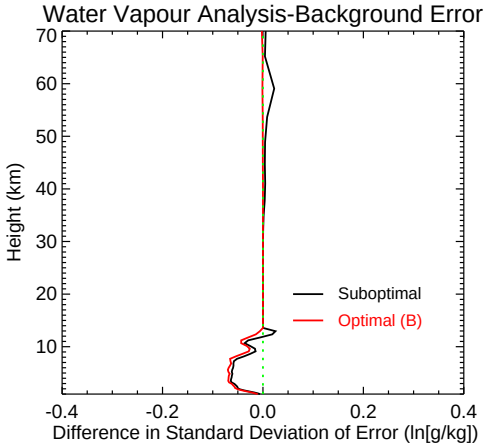
Case 3: Incorrect matrix; B_T = MoistCov, B_A = NMC

In this case, the variances of $\mathbf{B}_A$ are generally larger and some parts of the analysis might reasonably be expected to enter the danger zone. Figure 10.6 shows this to be the case. The results in profile space are very similar to those of Case 2, which suggest that for this matrix pairing, the variance component is more influential than mismatches in the vertical structure functions. Although the stratospheric water vapour analysis is now even worse, in eigenvector space the water vapour retrieval actually seems to fare slightly better. Despite the poor performance in parts of the profile, approximately six eigenvectors each of temperature and water vapour still benefit from the information in the IASI observations.

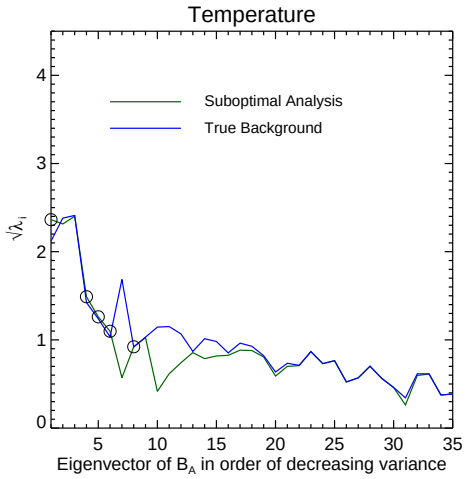
It is worth reiterating that $\mathbf{B}_T$ is unknown. The expected performance of the analysis system can only be measured using $\mathbf{B}_A$, i.e. by calculating $\mathbf{A}_{opt}(\mathbf{B}_A)$. It is quite possible that the suboptimal assimilation actually performs better than expected where $\mathbf{B}_T$ is smaller than $\mathbf{B}_A$ and $\mathbf{A}_{opt}(\mathbf{B}_T)$ is therefore more accurate than $\mathbf{A}_{opt}(\mathbf{B}_A)$. Suboptimal $\mathbf{A}$ falls somewhere between $\mathbf{A}_{opt}(\mathbf{B}_T)$ and $\mathbf{A}_{opt}(\mathbf{B}_A)$, and can therefore be more accurate than the projected performace of the system, i.e. $\mathbf{A}_{opt}(\mathbf{B}_A)$. The problem is that it is the expected performance that is in error, and the true forecast (with error $\mathbf{B}_T$) may still be degraded by the data assimilation process. This is illustrated in figure 10.7, which presents the same results as figure 10.6 but with additional lines showing $\mathbf{A}_{opt}(\mathbf{B}_A)$ and $\mathbf{A}_{opt}(\mathbf{B}_T)$. For many of the leading eigenvectors and for large parts of the profile, the suboptimal $\mathbf{A}$ is closer to $\mathbf{A}_{opt}(\mathbf{B}_T)$, and gives a better result than would have been predicted for this system based on



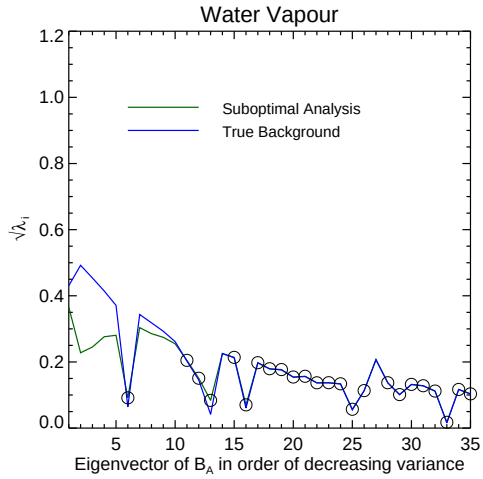
(a) Temperature - analysis error minus background error



(b) Water Vapour - analysis error minus background error

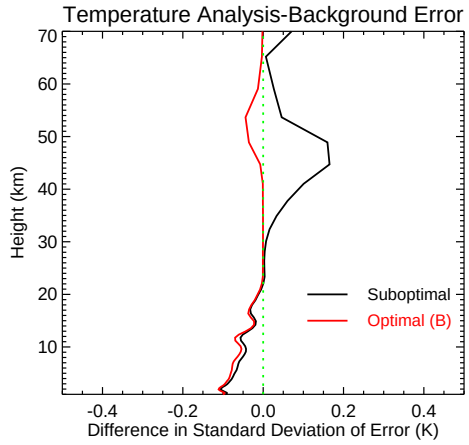


(c) Temperature - Analysis error and true background error projected onto eigenvectors of $\mathbf{B}_A$

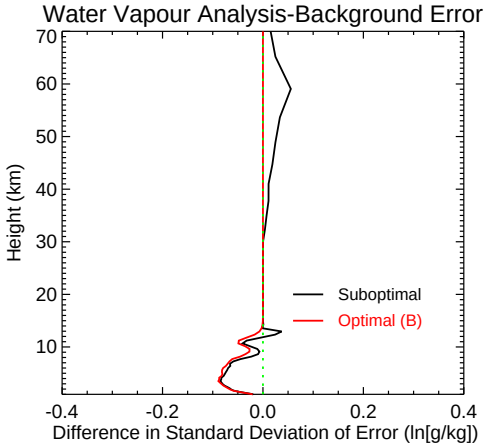


(d) Water Vapour - Analysis error and true background error projected onto eigenvectors of $\mathbf{B}_A$

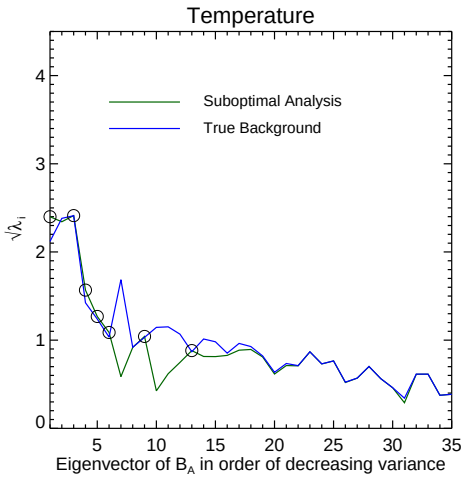
Figure 10.5: Suboptimal analysis for Case 2 where $\mathbf{B}_A$ =**NMC** and $\mathbf{B}_T$ =$\mathbf{B}'$ (see text for details). *VAR* channels. $\mathbf{R}$ contains instrument noise only. In (a), and (b), $\mathbf{A}_{opt}(\mathbf{B}_T)$ is in red and suboptimal $\mathbf{A}$ in black. In (c), and (d) $\mathbf{B}_T$ is in blue, $\mathbf{A}$ in green, and eigenvectors marked with circles have entered the danger zone.



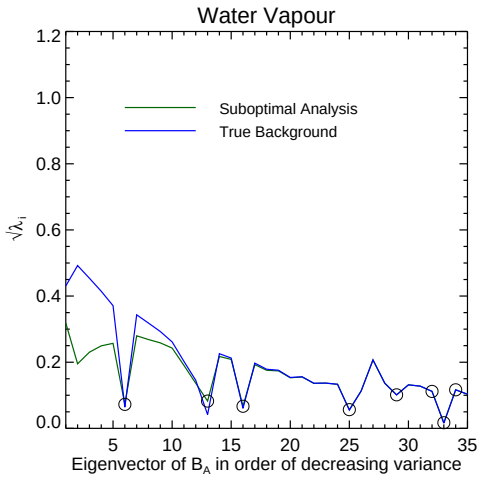
(a) Temperature - analysis error minus background error



(b) Water Vapour - analysis error minus background error

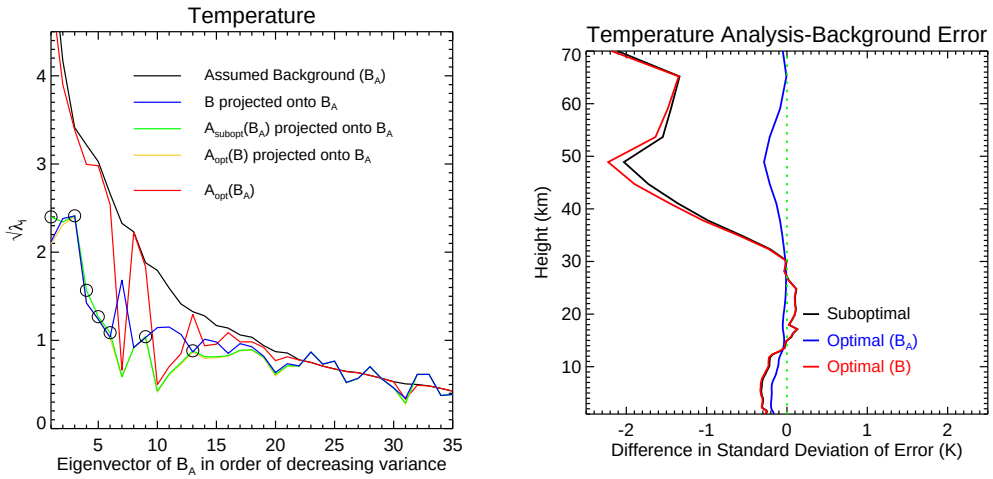


(c) Temperature - Analysis error and true background error projected onto eigenvectors of $\mathbf{B}_A$



(d) Water Vapour - Analysis error and true background error projected onto eigenvectors of $\mathbf{B}_A$

Figure 10.6: Suboptimal analysis for Case 3 where $\mathbf{B}_A$ is **NMC** and $\mathbf{B}$ is **MoistCov**. *VAR* channels. $\mathbf{R}$ contains instrument noise only. In (a), and (b), $\mathbf{A}_{opt}(\mathbf{B}_T)$ is in red and suboptimal $\mathbf{A}$ in black. In (c), and (d) $\mathbf{B}_T$ is in blue, $\mathbf{A}$ in green, and eigenvectors marked with circles have entered the danger zone.



(a) Optimal and suboptimal temperature analysis errors compared to the background errors $\mathbf{B}_A$ and $\mathbf{B}_T$ in the eigenvector space of $\mathbf{B}_A$

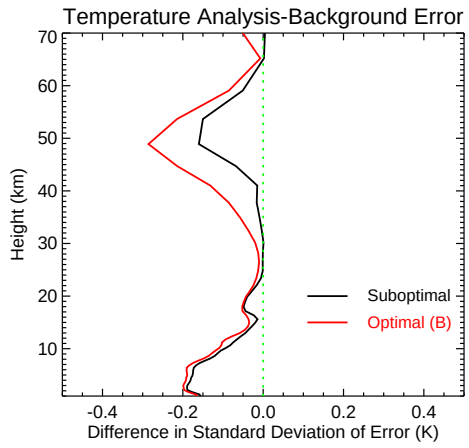
(b) Optimal and suboptimal temperature analysis errors compared to the background error $\mathbf{B}_A$ in profile space

Figure 10.7: Comparison of the suboptimal analysis error $\mathbf{A}$, and the optimal analysis errors $\mathbf{A}_{opt}$ $(\mathbf{B}_T)$ and $\mathbf{A}_{opt}$ $(\mathbf{B}_A)$ for temperature for the *VAR* channel selection with $\mathbf{R}$ containing instrument noise only, for Case 3 where $\mathbf{B}_A$ is **NMC** and $\mathbf{B}_T$ is **MoistCov**. (a) is plotted in the eigenvector space of $\mathbf{B}_A$ and adds the background error terms $\mathbf{B}_A$ and $\mathbf{B}_T$. (b) is in profile space and shows the respective analysis errors minus $\mathbf{B}_A$, so that a negative value is a lower error than $\mathbf{B}_A$ and a positive value is a higher error than $\mathbf{B}_A$.

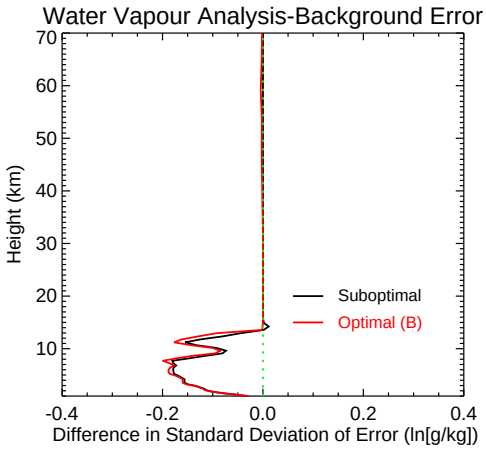
$\mathbf{B}_A$. For a few of the lower order eigenvectors, and in the lower stratosphere (12–28 km), $\mathbf{B}_A$ is in fact smaller than $\mathbf{B}_T$ and the performance of suboptimal $\mathbf{A}$, whilst in this case not degrading the analysis relative to the true forecast error, gives a performance that is worse than expected (i.e. the black line plots to the right of the blue line in figure 10.7(b) for this height range).

Case 4: Incorrect matrix; B_T =NMC, B_A =MoistCov

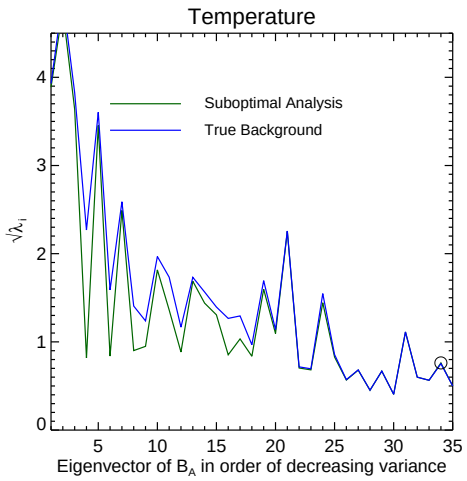
When $\mathbf{B}_T$ has variances generally larger than assumed, as is the case here, the behaviour of the scalar system does not predict that the retrieval would enter the danger zone. Figure 10.8 demonstrates that for this matrix pairing, the expectation holds for most of the profile: although retrieval errors are clearly larger than optimal, only the very top of the temperature profile and the water vapour profile around the tropopause show any degradation, and to a smaller degree than for the reverse matrix pairing shown in Case 3. There is no marked degradation to any significant eigenvectors (vector 34 for temperature is the first to be affected). Eight temperature vectors and six water vapour vectors show substantial improvement relative to the



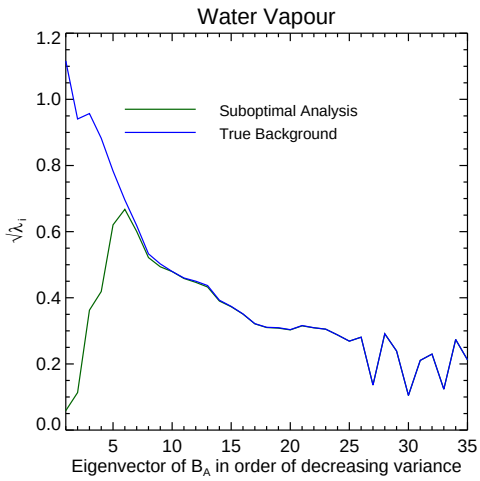
(a) Temperature - analysis error minus background error



(b) Water Vapour - analysis error minus background error



(c) Temperature - Analysis error and true background error projected onto eigenvectors of $\mathbf{B}_A$



(d) Water Vapour - Analysis error and true background error projected onto eigenvectors of $\mathbf{B}_A$

Figure 10.8: Suboptimal analysis for Case 4 where $\mathbf{B}_A$ is **MoistCov** and $\mathbf{B}_T$ is **NMC**. *VAR* channels. $\mathbf{R}$ contains instrument noise only. In (a), and (b), $\mathbf{A}_{opt}(\mathbf{B}_T)$ is in red and suboptimal $\mathbf{A}$ in black. In (c), and (d) $\mathbf{B}_T$ is in blue, $\mathbf{A}$ in green, and eigenvectors marked with circles have entered the danger zone.

background.

Once one understands the suboptimal theory, one can perhaps predict where there might be a risk of degrading the analysis, if the system's true performance does not match the expected performance. If the true system performs worse than expected, there is also a risk that parts of the analysis will be degraded and the covariance statistics should be retuned to make them smaller. If the analysis performs better than expected, it is likely that the background errors may be too restrictive

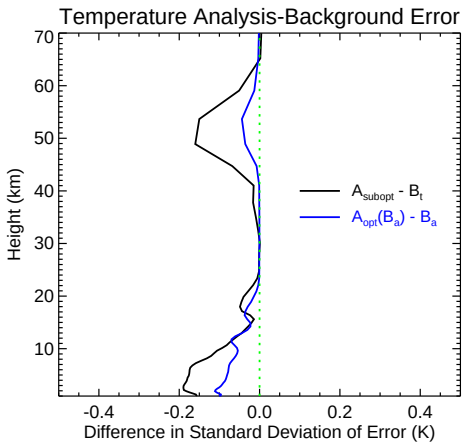


Figure 10.9: Comparison between predicted performance ($\mathbf{A}_{opt}(\mathbf{B}_A)$ minus $\mathbf{B}_A$ in blue) and true performance (Suboptimal $\mathbf{A}$ minus $\mathbf{B}_T$ in black) for the suboptimal analysis for Case 4 where $\mathbf{B}_A$ is **MoistCov** and $\mathbf{B}_T$ is **NMC**. *VAR* channels. $\mathbf{R}$ contains instrument noise only.

and that the observations are underweighted. Centres including ECMWF also use forecast sensitivity to observations (FSO) type measures to determine whether the relative balance between background and observation errors is correct (Daescu and Langland, 2013, for example). Case 4 is an example of where, for substantial portions of the profile, the suboptimal analysis is better than predicted by $\mathbf{B}_A$: figure 10.9 shows in blue the predicted improvement of the analysis, $\mathbf{A}_{opt}(\mathbf{B}_A)$, over the assumed background error and in black the true improvement of the suboptimal analysis, $\mathbf{A}$ over the true background error. In the real system it would not be possible to calculate the 'true' performance in this manner, as we do not know $\mathbf{B}_T$; furthermore we do not know the true state of the atmosphere so cannot construct error terms by comparing analyses against truth. Nevertheless, insight may be gained if the analysis fit to observations is much better than expected. The improvements over background would have been even greater if $\mathbf{B}_A$ did not underestimate the background error, and thus downweight the information from the observations. Of course, this would also assume that the observation errors are well-known, and in reality they are subject to nearly as much uncertainty as the background error statistics.

10.2.3 Stratospheric water vapour analysis

The fact that the problems with the suboptimal retrievals seem worst in the upper reaches of the water vapour profile (e.g. Case 3) should perhaps not be a surprise. It is well-known among satellite data assimilation researchers that attempting to assimilate increased water vapour information, whether by reducing assumed observation

errors, increasing the number of water vapour channels, or adding higher-peaking water vapour channels can cause difficulties.

Although 4D-Var calculates analysis increments for all levels of the model for q_t, when the increments are applied, they are set to zero above a diagonosed tropopause. The danger zone behaviour for vectors that have a stratospheric water vapour component in **NMC** is therefore not signficant in the context of the operational model. This behaviour can be replicated in the linear analysis calculation by restricting the analysis to include only the lowest 38 levels (approximately 10 km)of the water vapour profile. This technique is employed in calculation of the DFS for the suboptimal system, in section 10.2.4.

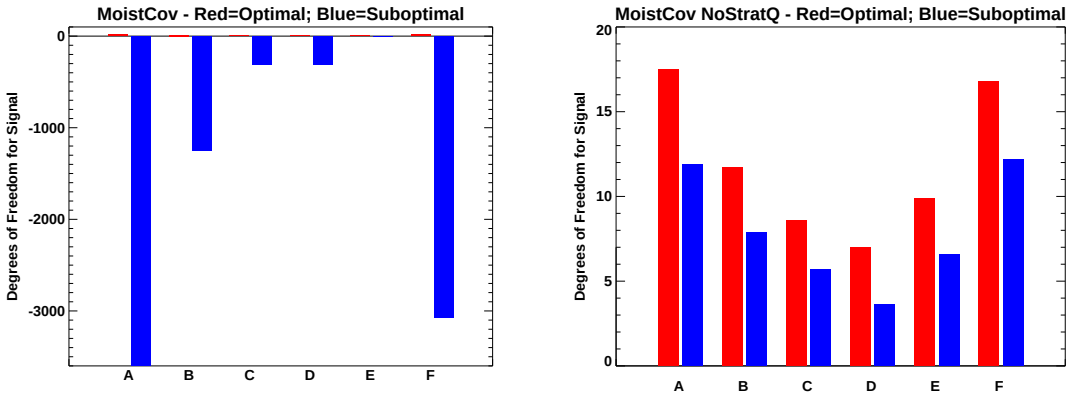
The reduction of increments to zero above the tropopause restricts the influence of these spurious stratospheric water vapour features, but this can lead to undesired features (e.g. the drying of the upper troposphere as described in Hilton et al., 2012b). The spurious stratospheric increments are not present with the new non-linear moisture control variable now in place, but it would still seem sensible to handle the assimilation of water vapour information fairly cautiously, particularly in the UTLS. The reason for the conservative use of water vapour channels in the operational implementation of IASI was because of minimisation problems in 4D-Var when either high-peaking channels or reduced observation errors were used, and it is possible that this was a consequence of increasing the exposure of the assimilation system to the danger zone.

10.2.4 Increasing the spectral coverage

The main aim of this thesis is to establish a method of maximising the overall information content of the analysis using IASI data, improving the extraction of high-vertical resolution structure information from IASI. It is important that the new analysis is no less robust to misspecification of $\mathbf{B}_A$.

Figure 10.10(a) shows the DFS for Case 3 ($\mathbf{B}_T = \mathbf{MoistCov}$, $\mathbf{B}_A = \mathbf{NMC}$), for the same channel selections as shown in figure 4.5 (Full spectrum, *Collard*, *OPS*, *VAR*, *Band1*, PC scores). The suboptimal DFS is strongly negative, and in fact the information content seems to be totally overwhelmed by the parts of the analysis that have entered the danger zone, and in particular by the stratospheric water vapour. Although the DFS measure has alerted us to the danger zone, it has ceased to be a useful measure of information content.

Given that the problems with the stratospheric water vapour analysis are unimportant for the real assimilation system, as described in section 10.2.3, the DFS



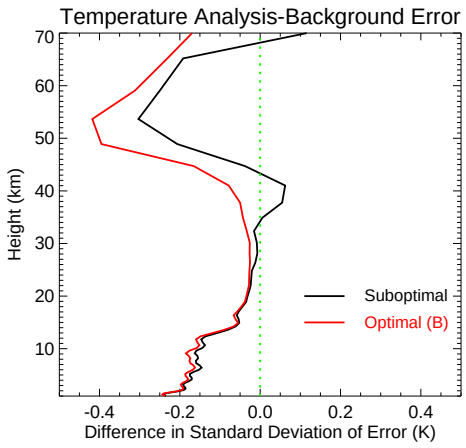
(a) With stratospheric water vapour **(b)** Without stratospheric water vapour

Figure 10.10: DFS for suboptimal analysis with different channel selections for Case 3 where $\mathbf{B}_A$ is **NMC** and $\mathbf{B}_T$ is **MoistCov**. $\mathbf{R}$ is instrument noise only. Red bars represent optimal DFS, and blue bars suboptimal. Note the y-axis scale in (a). A to F represent the same channel selections as in figure 4.5

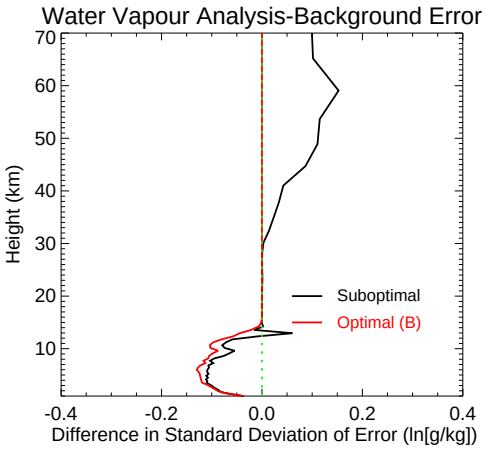
measure can instead be calculated for the lowest 38 levels of the water vapour profile (corresponding to a height of 10.1 km). This measure is shown in figure 10.10(b). The DFS calculation is now what one would expect. It can now be seen that the suboptimal analysis is always worse than the optimal analysis by about 25–30 %, and that increasing the spectral coverage of the assimilated IASI data improves the DFS of both optimal and suboptimal systems by approximately the same amount. When danger zone behaviour is not dominant, the analysis is suboptimal but there is no evidence to contradict the results in chapter 4 that improvements in information content result from the addition of more channels.

On the other hand, examination of the DFS does not show whether increasing the spectral coverage makes retrievals more robust to misspecification of $\mathbf{B}_A$. Figure 10.11 shows the Case 3 retrieval but using the whole spectrum instead of the *VAR* channel selection, to be compared with figure 10.6. Comparison of parts (a) of the figures shows that the use of the full spectrum hugely improves both optimal and suboptimal retrieval errors, and restricts the danger zone to the topmost one or two levels and the region between 35 and 43 km. In this respect one could say that the use of more spectral information somewhat increases robustness to the misspecification in $\mathbf{B}_A$.

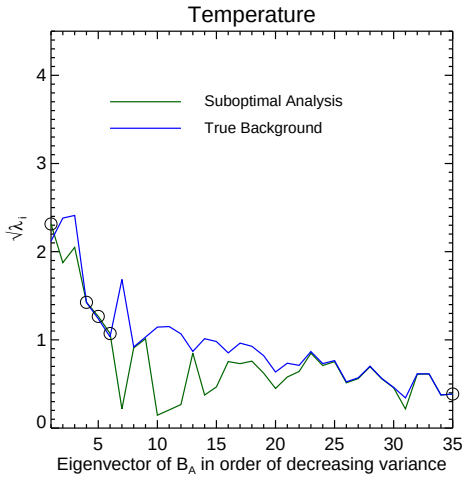
For water vapour, the analysis is improved in the troposphere, but is more significantly degraded in the stratosphere. Although this is not significant in terms of the operational analysis, for reasons stated in section 10.2.3, it illustrates a general point regarding danger zone behaviour that matches the DFS calculations plotted



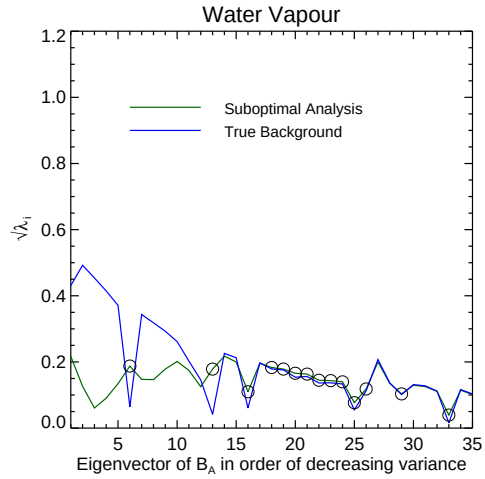
(a) Temperature - analysis error minus background error



(b) Water Vapour - analysis error minus background error



(c) Temperature - Analysis error and true background error projected onto eigenvectors of $\mathbf{B}_A$

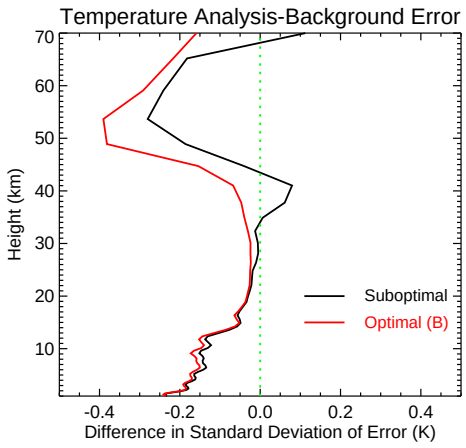


(d) Water Vapour - Analysis error and true background error projected onto eigenvectors of $\mathbf{B}_A$

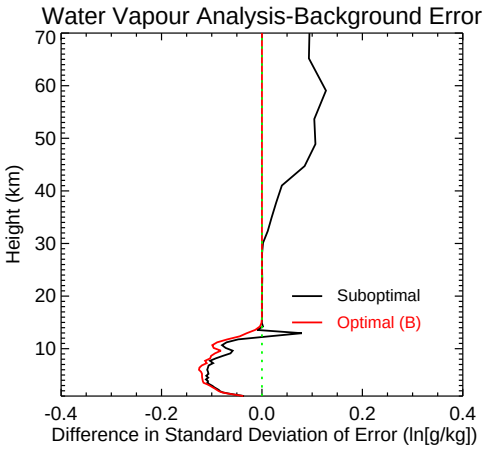
Figure 10.11: Suboptimal analysis for Case 3 where $\mathbf{B}_A$ is **NMC** and $\mathbf{B}$ is **MoistCov**. The full spectrum is used in the analysis. $\mathbf{R}$ contains instrument noise only. In (c), and (d) eigenvectors marked with circles have entered the danger zone.

in figure 10.10(a). If the analysis is in the danger zone and more weight is given to the observation information – in this case by increasing the number of water vapour channels rather than by reducing the observation error – it is possible to make the analysis even worse. This is not behaviour that is predicted by the simple scalar system shown in figure 10.2(b) and reflects the complex response of the atmospheric column to the IASI observation.

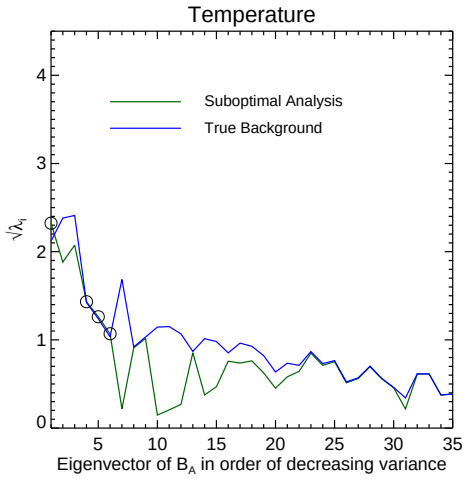
Figure 10.12 shows the Case 3 analysis for the 290 EUMETSAT PC scores



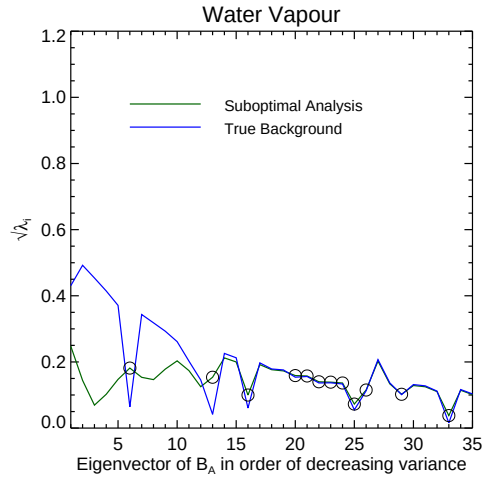
(a) Temperature - analysis error minus background error



(b) Water Vapour - analysis error minus background error



(c) Temperature - Analysis error and true background error projected onto eigenvectors of $\mathbf{B}_A$

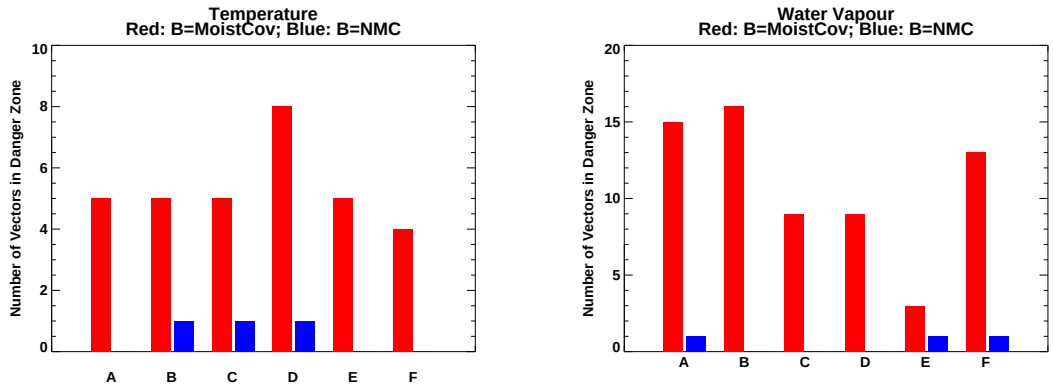


(d) Water Vapour - Analysis error and true background error projected onto eigenvectors of $\mathbf{B}_A$

Figure 10.12: Suboptimal analysis for Case 3 where $\mathbf{B}_A$ is **NMC** and $\mathbf{B}_T$ is **MoistCov**. 290 PC scores are used in the analysis. Instrument noise is assumed to be the only source of observation error. In (c), and (d) eigenvectors marked with circles have entered the danger zone

(Hultberg, 2009). The results are almost identical to those for the full spectrum (figure 10.11). This is encouraging as it means that the promising techniques of data compression explored in previous chapters do not increase susceptibility to the misspecification of $\mathbf{B}$, at least as long as the $\mathbf{R}$ matrix is well specified.

Figure 10.13 shows the number of significant eigenvectors for temperature and water vapour entering the danger zone for the same channel selections as figure 10.10. The red bars correspond to Case 3 and the blue to Case 4.

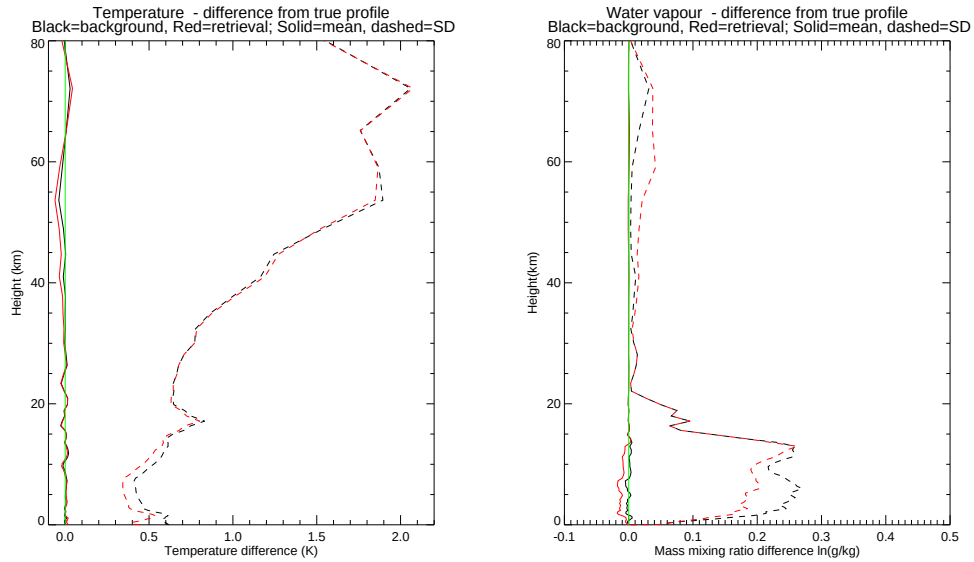


(a) Temperature **(b)** Water Vapour

Figure 10.13: Number of significant eigenvectors entering the danger zone for the suboptimal analysis for different channel selections as in figure 4.5: **A** – full spectrum; **B** – *Collard*; **C** – *OPS*; **D** – *VAR*; **E** – *Band1*; **F** – 290 PC scores. The red columns are for Case 3 where $\mathbf{B}_T$ = **MoistCov** and $\mathbf{B}_A$ = **NMC** and the blue columns are for Case 4 where $\mathbf{B}_T$ = **NMC** and $\mathbf{B}_A$ = **MoistCov**.

For temperature, it seems to be the case that increasing the spectral coverage does improve robustness to misspecification of background error (with this particular pair of matrices), i.e. the number of vectors entering the danger zone decreases. It is interesting to note that there does not seem to be much improvement beyond the 183 *OPS* channel selection: the main difference between the *OPS* and *VAR* channel selections is that the *OPS* selection contains channels peaking high in the stratosphere. This might indicate that using more IASI data aloft would be a priority to improve the analysis.

For water vapour, the situation is considerably more complex for Case 3, where $\mathbf{B}$ = **MoistCov** and $\mathbf{B}_A$ = **NMC**. Avoiding water vapour channels altogether (*Band1*; Column E) means that one is less susceptible to entering the danger zone; however, it also means that one has little chance of improving the model water vapour fields. Beyond that it seems that as more water vapour information is added, one is at greater risk of degrading the analysis. It is worth pointing out that this plot does not take into account the removal of the increments in the stratosphere: because the calculation is not in height space it is not so easy to remove a few levels from the profile. However, the results for Case 4 are not affected, because there are no spurious stratospheric features in $\mathbf{B}_A$ for this scenario. In this case, there seems to be little difference in danger zone behaviour between the different channel selections.



(a) Temperature **(b)** Water Vapour

Figure 10.14: Mean and standard deviation of (Background–Truth) and (Analysis–Truth) for 1D-Var analysis where background errors are calculated using the $\mathbf{B}_T$ matrix **MoistCov** but $\mathbf{B}_A$ is **NMC** (Case 3). *VAR* channels. $\mathbf{R}$ contains instrument noise only.

10.3 Suboptimal 1D-Var

This section investigates the behaviour of the sub-optimal 1D-Var, matching Case 3 of section 10.2, where $\mathbf{B} = \mathbf{MoistCov}$ and $\mathbf{B}_A = \mathbf{NMC}$, for the *VAR* channel selection. Figure 10.14 compares the analysis and background errors for the 4348 profile set. The results are broadly similar to those predicted by the linear suboptimal analysis equations, in that some parts of the profile, particularly in the high stratosphere, are degraded through the misspecification of $\mathbf{B}_A$, and the tropospheric profiles are improved by about 0.1 K and up to 0.07 $\ln(q)$ units. However, the linear analysis implied a slightly better improvement in $\ln(q)$, and the 1D-Var results are subject to radiance non-linearity error (see chapter 5), which manifests as an analysis bias for water vapour.

In terms of danger zone degradations, they are slightly less severe than predicted, other than in the 60 km region of the water vapour profile. No substantial suboptimal behaviour is visible in the water vapour profile at the tropopause, and the temperature profile at 40 km is only degraded by about 0.05 K rather than the 0.2 K predicted. There is, however, a larger bias in retrieved temperature at the same altitude.

The results presented in this section show that the suboptimal behaviour pre-

dicted by theory is a real phenomenon. They confirm that linear analysis is a useful tool to predict suboptimal behaviour in the full 1D-Var case, but care should be taken when assessing the magnitude of the effects.

10.4 A discussion on observation errors

All of the calculations in the linear analysis, and all of the 1D-Var runs presented in this thesis used a diagonal **R** matrix of level 1c instrument noise only, with no attempts to add forward model error or representativeness error. This is clearly unrealistic, and represents an idealised scenario. Although the main purpose of this thesis is not to investigate changes to the observation error term, it is instructive to see how the behaviour of the system changes under different, more realistic, observation error assumptions.

Figure 10.15 shows the DFS for the optimal and suboptimal analyses for Case 4 for the *VAR* channel selection. The results are shown for Case 4 rather than Case 3, because the latter suffer from the effects of the stratospheric water vapour analysis seen in figure 10.10(a). Four settings are shown:

Inst: Diagonal level 1c instrument noise only, identical to the assumptions in previous sections.

AddRT: Diagonal level 1c instrument noise, plus 0.2 K radiative transfer error, plus additional error relating to climatological variability of unmodelled trace gases. This final term is negligible for the channel selection used in this study

RVar: The diagonal error covariance matrix used operationally in 4D-Var for IASI assimilation until January 2013. Most tropospheric CO_2 channels were assimilated with 0.5 K errors, window channels and stratospheric temperature channels with 1 K errors, and water vapour channels with 4 K errors.

Deroz: A full error covariance matrix calculated using the Desroziers method (Desroziers et al., 2005) from the operational 4D-Var system in 2011 (Weston, 2011). This matrix is used operationally for IASI assimilation in 4D-Var since January 2013. As well as the introduction of inter-channel correlations, the Desroziers error standard deviations are generally slightly smaller than the AddRT error standard deviations for temperature sounding channels (about 0.05 K on average), and somewhat larger for the water vapour sounding channels (about 0.4 K).

Figure 10.15 shows that as the observation error increases, the DFS of the optimal retrieval is reduced, as expected. Although the Desroziers matrix has quite

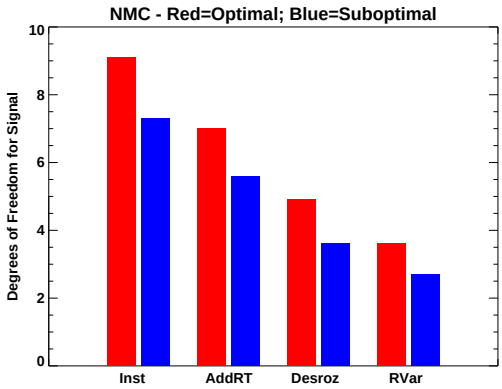
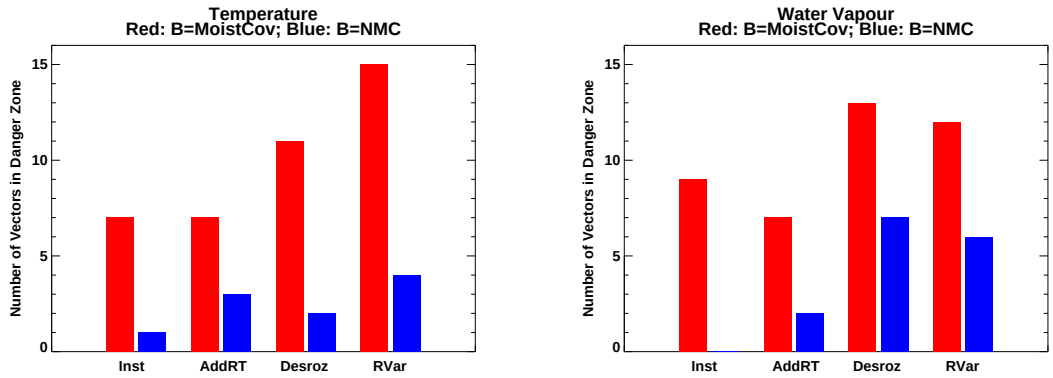


Figure 10.15: Comparison of DFS for optimal and suboptimal analysis, for the *VAR* channel selection with different observation errors, for Case 3, where $\mathbf{B}_T$ = **MoistCov** and $\mathbf{B}_A$ = **NMC**. Column labels are as follows: **Inst**: Instrument noise only; **AddRT**: Instrument noise plus forward model error; **Desroz**: $\mathbf{R}$ matrix calculated using the Desroziers diagnostic, operational at the Met Office since January 2013; **RVar**: $\mathbf{R}$ matrix used at the Met Office in 4D-Var prior to January 2013. See text for further details. The **Inst** column is equivalent to column D in figure 10.10(b)

small observation errors, for temperature at least, the DFS with this matrix is still smaller than that with the AddRT error matrix: taking proper account of the error correlations and the errors of representation in 4D-Var results in a relatively low weight being give to the data overall. The results using the Desroziers matrix are probably the best we can hope to achieve taking reasonable account of errors that genuinely exist in the system. The old 4D-Var error assumptions (RVar) are clearly more conservative, and the improvement in DFS seen here was reflected in modest forecast impact when the Desroziers matrix was adopted (Weston et al., 2014).

Similarly to figure 10.10(a), the suboptimal results are worse than the optimal results, and the degree to which they are worse is not greatly influenced by the observation errors chosen.

Figure 10.16 shows the number of eigenvectors entering the danger zone for the suboptimal analysis with each of these observation error scenarios for Cases 3 and 4. Here we see some evidence that reducing observation errors protects against the number of significant modes entering the danger zone, which fits with the scalar predictions of figure 10.2(b). Again, the results for water vapour for Case 3 (in red) include the degradation to the spurious stratospheric modes, and should be treated with caution. Note that this figure does not demonstrate the effect of reducing the assumed observation errors, but rather that of reducing the true observation errors, which would require improvements in, for example, radiative transfer modelling, or in the vertical and horizontal resolutions of the model, to represent more closely the



(a) Temperature **(b)** Water Vapour

Figure 10.16: Number of eigenvectors entering the danger zone. The red columns are for Case 3 where $\mathbf{B}_T$ = **MoistCov** and $\mathbf{B}_A$ = **NMC** and the blue columns are for Case 4 where $\mathbf{B}_T$ = **NMC** and $\mathbf{B}_A$ = **MoistCov**.

volume of atmosphere to which the IASI data are sensitive.

10.5 Consequences for assimilation and retrieval

The results presented in this chapter suggest that guarding against exaggeration of background errors is important, and that care should be taken in the construction of background error terms for water vapour, and in particular to avoid the introduction of spurious vertical modes in the stratosphere. In an operational NWP setting, there are numerous considerations regarding the construction of background error terms, and it may not be possible to ensure that $\mathbf{B}_A$ does not exaggerate $\mathbf{B}_T$ in any respect. The use of Hybrid DA systems is a definite step forward, and helps to define a $\mathbf{B}_A$ that is regionally rather than globally valid. An alternative approach being investigated at Météo-France is to define different error distributions for different synoptic conditions that have more homogeneous error properties, for example in precipitating and non-precipitating areas (Thibaut Montmerle, personnal communication).

There are more obvious consequences for standalone 1D retrieval schemes where background error terms can be tailored to the application. However, in general, the background profiles used in standalone schemes are from a climatological source, and there is little information in them that is likely to be lost through underweighting the background. However, if background profiles can be defined regionally or come from an NWP model, it is desirable to maintain realistic error distributions. It may also be useful to examine the eigenvectors of any diagnosed $\mathbf{B}_A$ to check for spurious

modes that could be damped down.

10.6 Summary

The behaviour of the retrieval system in the case where $\mathbf{B}_T$ is not known exactly has been investigated by extending the linear analysis equations to take account of differences between $\mathbf{B}_T$ and $\mathbf{B}_A$. The behaviour of a scalar system is investigated, and it is shown that the true analysis error, $\mathbf{A}$, is larger than the error that would be predicted if $\mathbf{B}_T$ were correctly specified ($\mathbf{A}_{opt}(\mathbf{B}_T)$). If $\mathbf{B}_A$ is too large, in the extreme case it is possible for the analysis to be worse than the background, and this degradation is referred to as the danger zone.

The suboptimal analysis equations were then applied to a full column analysis of temperature and water vapour. It was found that where $\mathbf{B}_A$ is too large, the retrieved profile may be degraded relative to the background profile, both in physical space and in vertical eigenmode space. A degradation in a small vertical range can be spread over numerous modes in the vertical.

A suboptimal version of the DFS calculation has been performed. This becomes negative for a scalar system when the danger zone is breached. However, the behaviour of a full column analysis is more complex, and once a substantial breach of the danger zone is made, the DFS becomes overwhelmed and is no longer a reasonable indicator of information content. In terms of information content where the danger zone is not a significant contribution, the suboptimal retrieval always has a lower information content than the optimal, by about 25–30 % in the cases investigated here.

The danger zone behaviour has also been demonstrated using 1D-Var with the *VAR* channel selection, confirming that linear analysis is a useful tool to predict performance of a 1D-Var retrieval system, although the magnitude of the degradation is somewhat smaller in a 1D-Var setting than is predicted by the linear equations.

The results presented in this chapter suggest that guarding against exaggeration of background errors is important, and that care should be taken in the construction of background error terms for water vapour, and in particular to avoid the introduction of spurious vertical modes in the stratosphere. Examination of the eigenvectors of $\mathbf{B}_A$ is a useful tool to investigate possible problems arising from spurious modes of $\mathbf{B}_A$.

Increasing the weight to the observation by decreasing observation errors results in an increase in DFS, both for the optimal and suboptimal analysis, and without

stratospheric water vapour effects, reduced exposure to the danger zone. The effect of incorrectly specified observation error has not been investigated here.

Increasing the spectral coverage reduces exposure to the danger zone for temperature. For water vapour, once stratospheric effects are taken into account, an increase in spectral coverage does not make a significant difference to the danger zone behaviour for the matrix pairing of **MoistCov** and **NMC**. These results are encouraging because it means that there is no greater risk to the analysis if more spectral information is provided: this is demonstrated here for the full spectrum and for the EUMETSAT PC scores.

Chapter 11

Conclusions

IASI and other hyperspectral sounders have been shown to provide significant impact to NWP systems. However, current assimilation schemes use less than 2% of the channels and results presented in this thesis show that it may be possible to double the information content provided by IASI if the full spectrum were exploited. The techniques of PC compression and radiance reconstruction have been explored, and it has been shown that in an optimal linear assimilation, it is possible to extract almost the full information content of the spectrum using data in a compressed form.

11.1 The performance of the current assimilation scheme

Chapter 4 explored the information content of IASI in the current Met Office assimilation system. New Forecast Sensitivity to Observations results that split the impact into different regions in the vertical suggest that IASI is generally having the expected effect within the assimilation system, with water vapour and window channels providing information to the humidity fields, while temperature channels contribute to potential temperature and wind fields. Higher peaking channels have impact high in the atmosphere, while lower peaking channels contribute more information to the lower and mid troposphere. This suggests that there are no obvious deficits that need correcting, and that to improve the amount of information extracted from IASI, other than by reducing observation errors through a more accurate forward operator, the best approach is to increase the spectral coverage presented to the 4D-Var analysis.

The optimal estimation equations predict that much more impact could be gained from IASI if the full spectrum could be assimilated, in particular smaller-scale features in the vertical would be improved for the water vapour analysis, and the stratospheric temperature analysis could also be improved. Inspection of the averaging kernels also suggests that more information could be provided to the lowest levels of the atmosphere on both temperature and humidity.

Chapter 5 compared theoretical linear optimal estimation results in profile space and in the eigenvector space of the background error covariance matrix. The humidity field is improved by around $0.08\ln(q)$ units at wavelengths of up to 7km. Although the improvement is approximately 25% in terms of standard deviation, the vertical resolution is much less than the instrument specification predicted. Approximately 7 temperature vectors are improved, with mostly tropospheric features. The DFS for the whole analysis is 7.3, split almost equally between temperature and humidity. These improvements fall well short of the impact predicted by pre-launch studies (Collard, 1998; Prunet et al., 1998), and this assimilation scenario is already somewhat unrealistic because the only source of error assumed is instrument noise. The results of the linear optimal estimation were verified using a simulated 1D-Var experiment, giving consistent results in terms of standard deviation, but demonstrating the introduction of a bias in the water vapour analysis because of errors resulting from non-linearity in the radiance response to changes in water vapour amount.

11.2 Principal components and reconstructed radiances

Chapters 6 and 7 introduced principal component (PC) compression and radiance reconstruction. The importance of the rank of the reconstructed radiance observation error covariance matrix was emphasised, and the impact this has on the channel selection for reconstruced radiances. Practical implementation of reconstructed radiances for assimilation is likely to make use of raw radiance forward model Jacobians (i.e. be suboptimal-in-$\mathbf{H}$), and an extra error term, $\Delta\tilde{\mathbf{R}}$ was demonstrated. $\Delta\tilde{\mathbf{R}}$ is orthogonal to the instrument noise covariance matrix for reconstructed radiances, and therefore has implications for the rank of the problem. Neglect of this term renders the assimilation suboptimal-in-$\mathbf{H}$-and-$\mathbf{R}$. In practice, however, the necessary use of a diagnosed error covariance matrix that reflects all sources of error including $\Delta\tilde{\mathbf{R}}$ reduces the requirement to have an accurate estimate of this quantity. One example of such a diagnosed matrix, $\mathbf{R}_{HL}$, was presented, and it was found that the rank of the matrix was close to the number of retained PCs.

Two new channel selection methods for reconstructed radiances were presented. The first is a mathematical approach using the eigenvectors themselves to select the most information-rich channels. The second is based on information content, taking full account of observation error covariances, and penalising the selection of a channel

that raises the condition number of the resultant **R** matrix. These methods have been used to explore the selection of channel sets for reconstructed radiances, which have a high degree of error correlation between channels, by varying the inputs for the channel selection process. The new channel selections were compared with the *Collard* selection for information content and matrix conditioning. It is not possible to determine a definitive 'best' channel selection with such a simple test. However, tailoring the channel selection to NWP by choosing on the basis of DFS appears to provide a more information-rich channel selection in fewer channels than when choosing on the basis of condition number of the **R** matrix alone or when using the eigenvectors themselves.

To test whether the reconstructed radiance channel selection method is capable of finding channels that represent the full information content of the observation, further channel selections were made assuming instrument noise only. DFS calculations showed that the reconstructed radiance assimilation is capable of providing the full information content of the 210 PC scores, close to that of the full spectrum, but the **R** matrix is not well-conditioned, risking the addition of pseudo-information to the retrieval and an unstable minimisation. Shrinkage has been explored as a pragmatic way to modify the **R** matrix towards a diagonal to correct for this. This technique was shown to reduce the pseudo-information and to stabilise the solution, but as a consequence reduces the DFS considerably.

The performance of an instrument noise only 1D-Var, for both PC scores and reconstructed radiances, was assessed for a Suboptimal-in-**H**-and-**R** assimilation of reconstructed radiances. For reconstructed radiances, without significant manipulation of the **R** to counteract the missing error term that comes from using an incorrect forward model, the 1D-Var performance is very poor with most observations failing to reach covergence. Shrinking the **R** matrix towards a diagonally-strengthened matrix does allow an analysis to be calculated in most cases, but oscillating biases are produced. The performance of a similar system for PC scores is quite well behaved as long as all sources of error are taken into account, but if there is a missing error term, similar oscillating biases result.

These results suggest that much care needs to be taken when specifying error covariances for use with compressed spectra. It is likely that the tests performed in this thesis exaggerated the effects of the misspecification of the **R** matrix because the errors applied were so small.

The preferred channel selection designed for operational implementation at the Met Office, *RHL-M2-8J-1.3*, was also tested with the $\mathbf{R}_{HL}$ error covariance matrix

in a simulated 1D-Var environment, and ecouragingly gave good results without any oscillating bias. If $\mathbf{R}_{HL}$ does a reasonable job of representing the structure of the inter-channel errors, this suggests that reconstructed radiances could be used in an operational context and that further tests with real data are warranted.

11.3 Assimilation with a misspecified B matrix

The behaviour of the retrieval system in the case where $\mathbf{B}$ is not known exactly was investigated by extending the linear analysis equations to take account of differences between $\mathbf{B}$ and $\mathbf{B}_A$. The behaviour of a scalar system was investigated, and it was shown that the true analysis error, $\mathbf{A}$, is larger than the error that would be predicted if $\mathbf{B}_T$ were correctly specified. If $\mathbf{B}_A$ is too large, in the extreme case it is possible for the analysis to be worse than the background, and this degradation is referred to as the danger zone.

The suboptimal analysis equations were then applied to a full column analysis of temperature and water vapour. It was found that where $\mathbf{B}_A$ is too large, the retrieved profile may be degraded relative to the background profile, both in physical space and in vertical eigenmode space. A degradation in a small vertical range can be spread over numerous modes in the vertical. Where the danger zone is not a significant contribution, the suboptimal retrieval always has a lower information content than the optimal, by about 25–30% in the cases investigated here. The danger zone behaviour was also demonstrated using 1D-Var with the *VAR* channel selection, confirming that linear analysis is a useful tool to predict performance of a 1D-Var retrieval system, although the magnitude of the degradation is somewhat smaller in a 1D-Var setting than is predicted by the linear equations.

The results of chapter 10 suggest that guarding against exaggeration of background errors is important, and that care should be taken in the construction of background error terms for water vapour, in particular to avoid the introduction of spurious vertical modes in the stratosphere. Although this cannot always be achieved, it is useful to know how the observations interact with the vertical structures in $\mathbf{B}$, and confirms that work to improve the representation of background error in NWP models by representing synoptic structure will be beneficial to the assimilation of satellite observations that have broad sensitivities in the vertical.

Increasing the weight to the observation by decreasing observation errors results in an increase in DFS, both for the optimal and suboptimal analysis, and without stratospheric water vapour effects, reduced exposure to the danger zone. Increasing

the spectral coverage reduces exposure to the danger zone for temperature. For water vapour, once stratospheric effects are taken into account, an increase in spectral coverage does not make a significant difference to the danger zone behaviour. This result is encouraging because it means that there is no greater risk to the analysis if more spectral information is provided by the observations.

11.4 Further work

The specific aims of this thesis were to investigate the relative performance of IASI PC scores and reconstructed radiances in a clear sky assimilation setting, in terms of information content and vertical structure, and to find a set of reconstructed radiance channels with the maximimum information content possible that could be used in an operational setting.

There are several ways in which this work should be extended:

Clear sky PC/reconstructed radiance comparison: The intercomparison 1D-Vars in this thesis assumed similar – but not identical –instrument noise profiles were the only source of error. It would be useful to rerun this intercomparison with identical instrument noise and with addition of reasonable representation and radiative transfer errors, as this would make the experiment more relevant to assimilation in the real world.

Clear sky raw/reconstructed radiance comparison: A meaningful intercomparison between raw and reconstructed radiances was hampered by the lack of a dataset of observation minus background statistics for the full spectrum of raw radiances. It would be useful to gather this dataset to allow better diagnostics for the reconstructed radiance system.

Cloudy assimilation: There is much work to be done on the assimilation of cloud-affected PC score profiles, which is well beyond the scope of this thesis. However, the assimilation of cloud-affected reconstructed radiances should be tested, because it is possible that subtle changes in the error correlation structure could be introduced by the spreading of cloud information between channels that would affect cloud detection algorithms resulting in detrimental performance.

Operational implementation: One of the outputs of this thesis is a channel selection for reconstructed radiances tuned to the Met Office assimilation scheme. This channel selection should be tested in an operational context as soon as possible.

Surface parameter retrieval: The results shown here were for the analysis

of temperature and water vapour profiles only in order to mimic 4D-Var as closely as possible. In the operational Met Office assimilation scheme a 1D-Var pre-processor is used to perform a retrieval of skin temperature and surface emissivity for use in 4D-Var, along with diagnosed cloud properties. The channel selection for reconstructed radiances does not contain many window channels, which could hamper the retrieval of emissivity. An alternative channel selection should be devised including the retrieval of skin and surface parameters in the DFS calculation.

The impact of PC score dissemination on PC score assimilation: One of the main motivations for the use of reconstructed radiances is that data providers such as EUMETSAT already disseminate PC-compressed radiances, and this method of dissemination is the baseline for geostationary hyperspectral sounders such as MTG-IRS. PC-based forward models typically do not cater for the eigenvector basis used for dissemination (partly because these may be changed with time, for example following unusual atmospheric events). Little work has been done to date on the construction of standard PC scores for assimilation from PC scores for dissemination (or their reconstructed radiances). To be properly prepared for the arrival of these new instruments, it needs to be verified that this additional projection would not cause any difficulties.

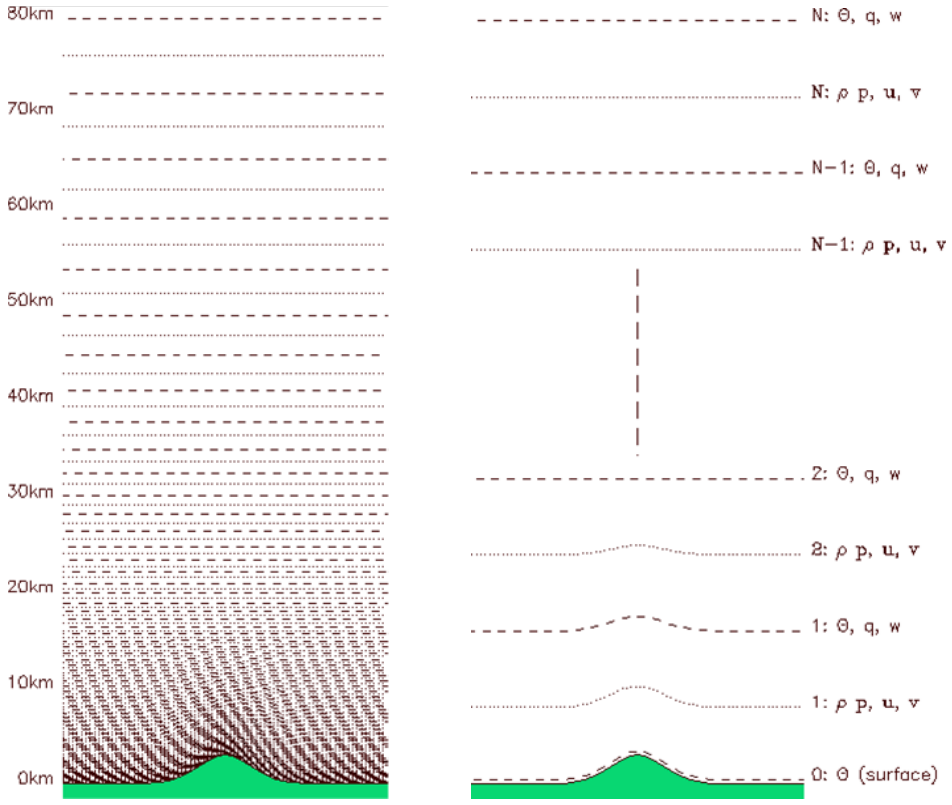
Dissemination of PC scores for future sounders: The baseline for hyperspectral sounders in geostationary orbit is that they will be disseminated in PC score format. The use of PC-compressed data, whether as PC scores or reconstructed radiances is therefore relevant to the wider remote sensing community. Further research into the use of compressed data for other applications, such as trace gas retrievals, is very important. Outside of an NWP context, it may not be so easy to use diagnostic techniques to determine observation errors and the implications of a suboptimal analysis should be investigated.

Appendices

Appendix A

The levels of the Met Office Unified Model

The Met Office Unified Model has 70 vertical levels in a staggered Charney-Phillips grid with potential temperature, moisture and the vertical component of the wind defined on one set of levels, and the horizontal wind components, pressure and density defined on the half-levels in between. The levels are expressed as height, and are terrain-following in the lower atmosphere. Figure A.1(a) shows the distribution of the levels in height, and figure A.1(b) shows the staggering of levels in the Charney-Phillips grid.



(a) The vertical levels of the UM

(b) The staggering of levels on the Charney-Phillips grid

Figure A.1: Vertical levels of the Charney-Phillips grid in the version of the Met Office Unified Model used in this thesis. Theta, moisture variables and the vertical wind component are defined on the levels marked with dashes, and the horizontal wind components, pressure and density are defined on the levels marked with dots.

Appendix B

Eigenvectors of the MoistCov B matrix

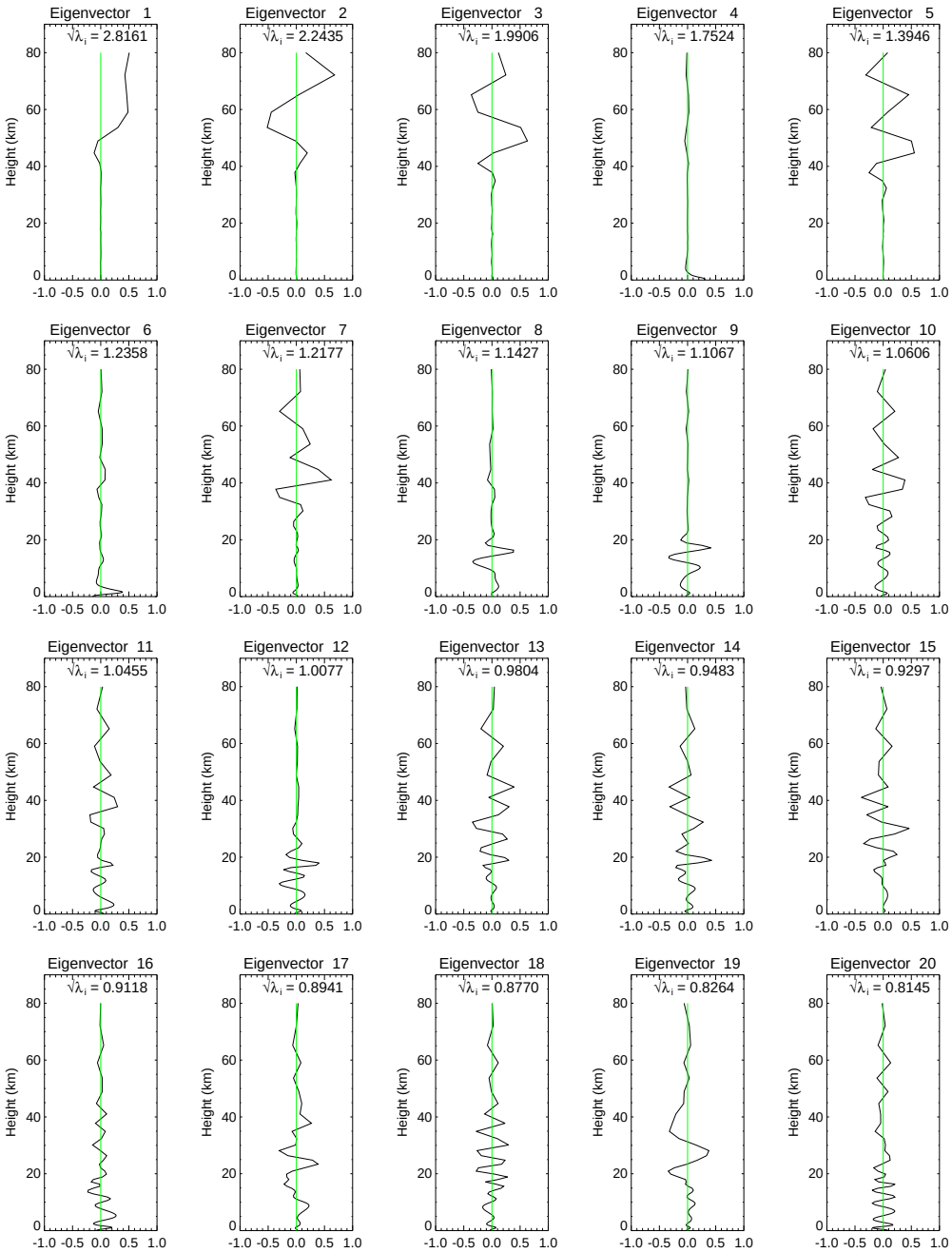


Figure B.1: Temperature eigenvectors 1–20 of the **MoistCov** $\mathbf{B}$ Matrix

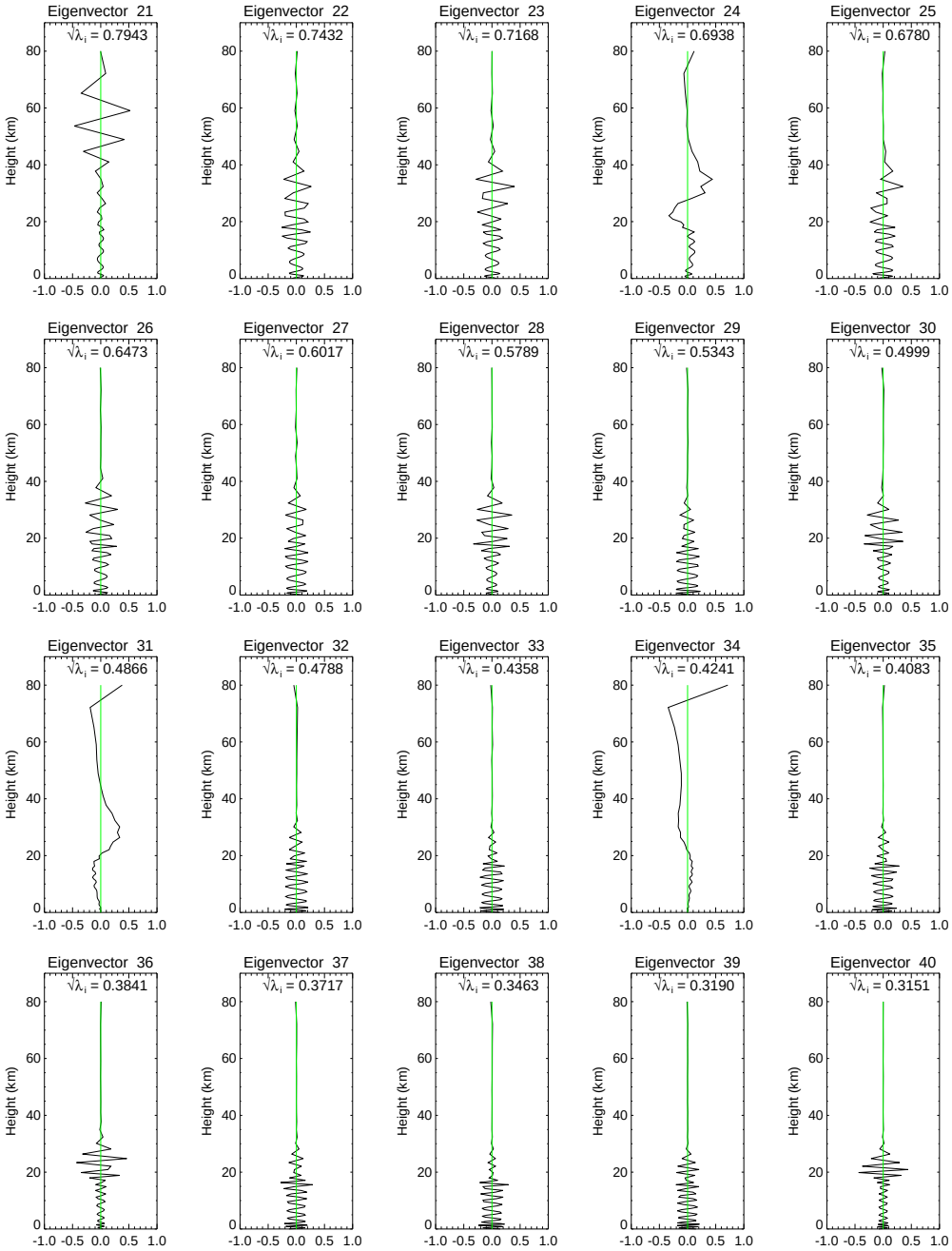


Figure B.2: Temperature eigenvectors 21–40 of the **MoistCov** $\mathbf{B}$ Matrix

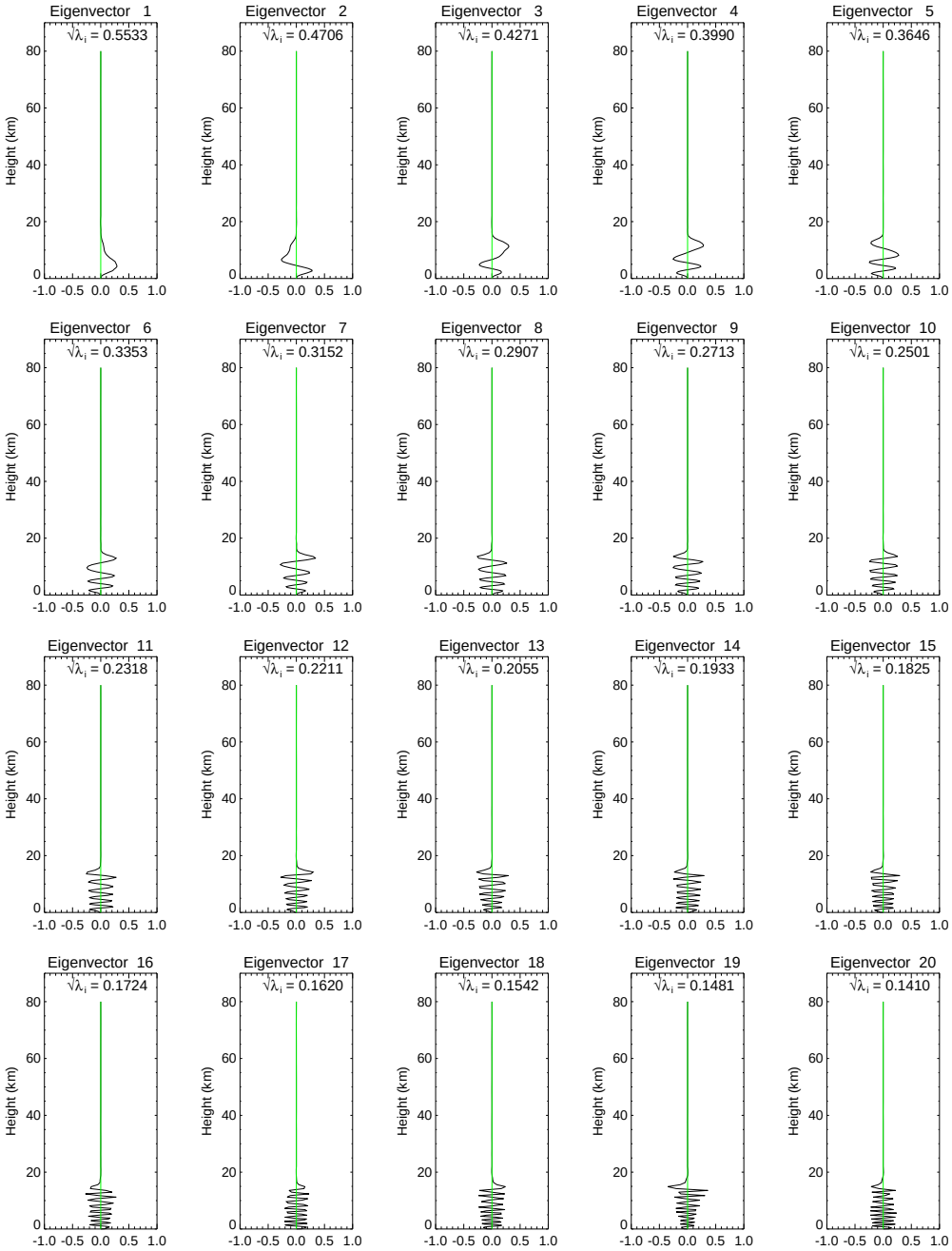


Figure B.3: Water vapour eigenvectors 1–20 of the **MoistCov** $\mathbf{B}$ Matrix

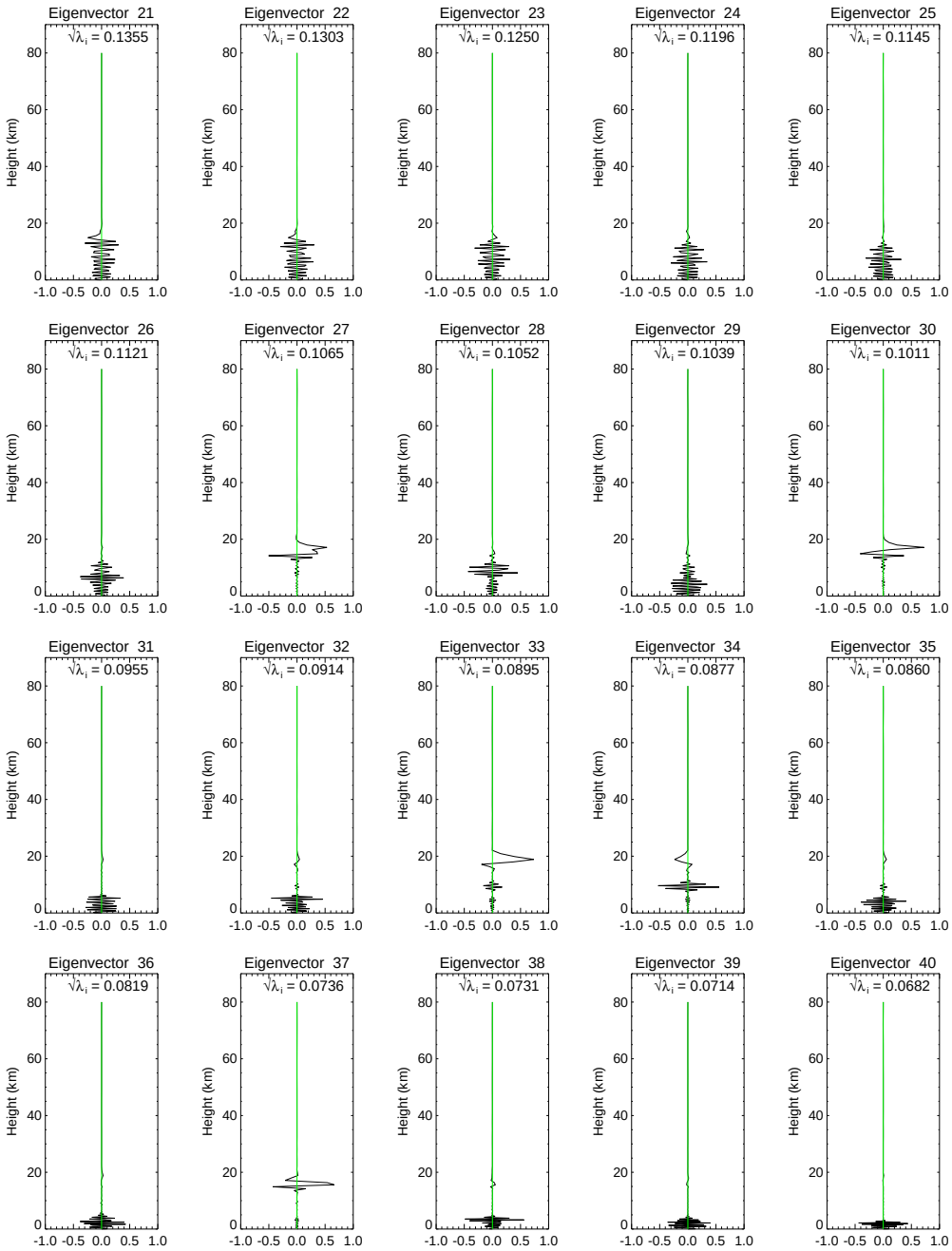


Figure B.4: Water vapour eigenvectors 21–40 of the **MoistCov** $\mathbf{B}$ Matrix

Appendix C

Eigenvectors of the NMC $\mathbf{B}$ matrix

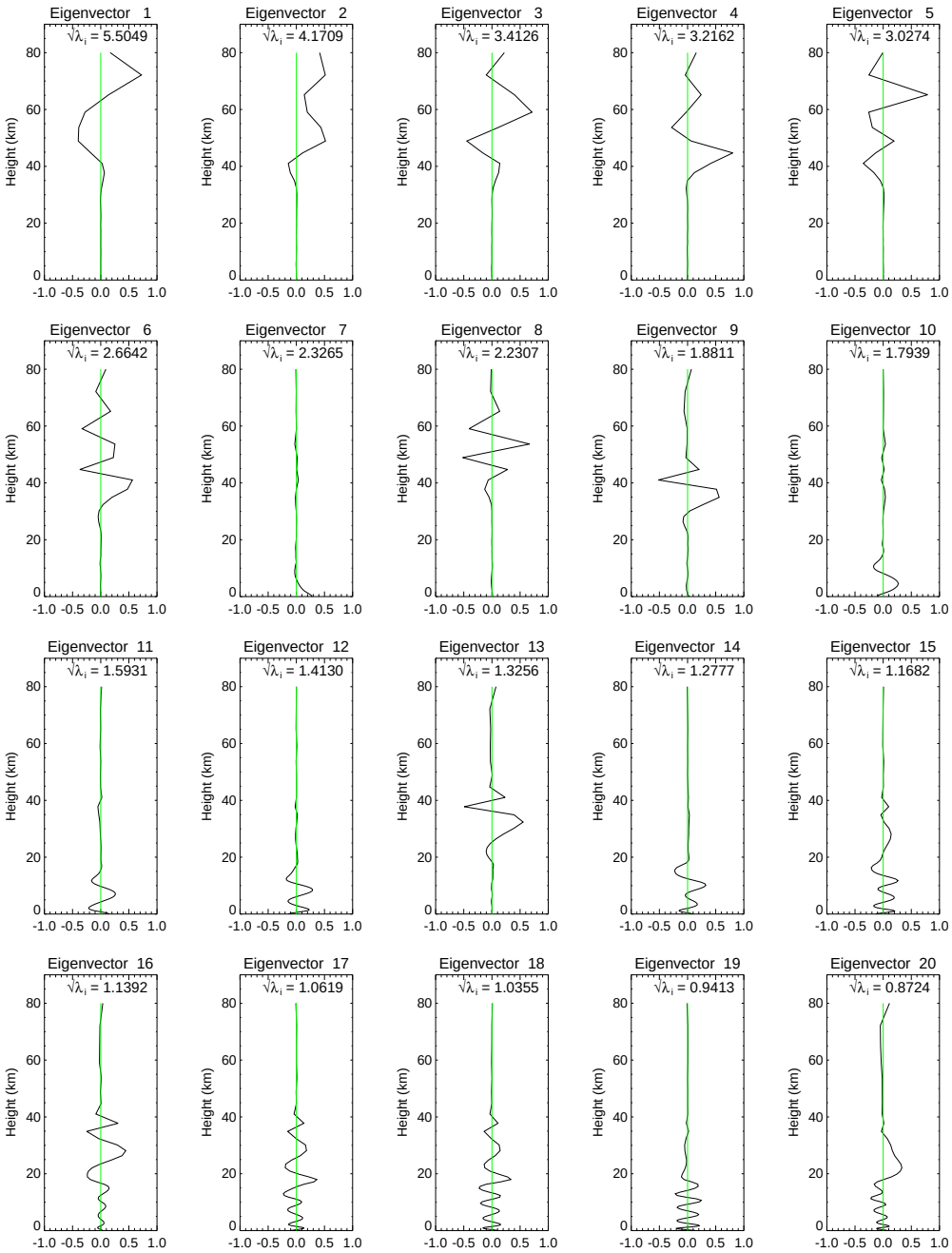


Figure C.1: Temperature eigenvectors 1–20 of the **NMC** $\mathbf{B}$ Matrix

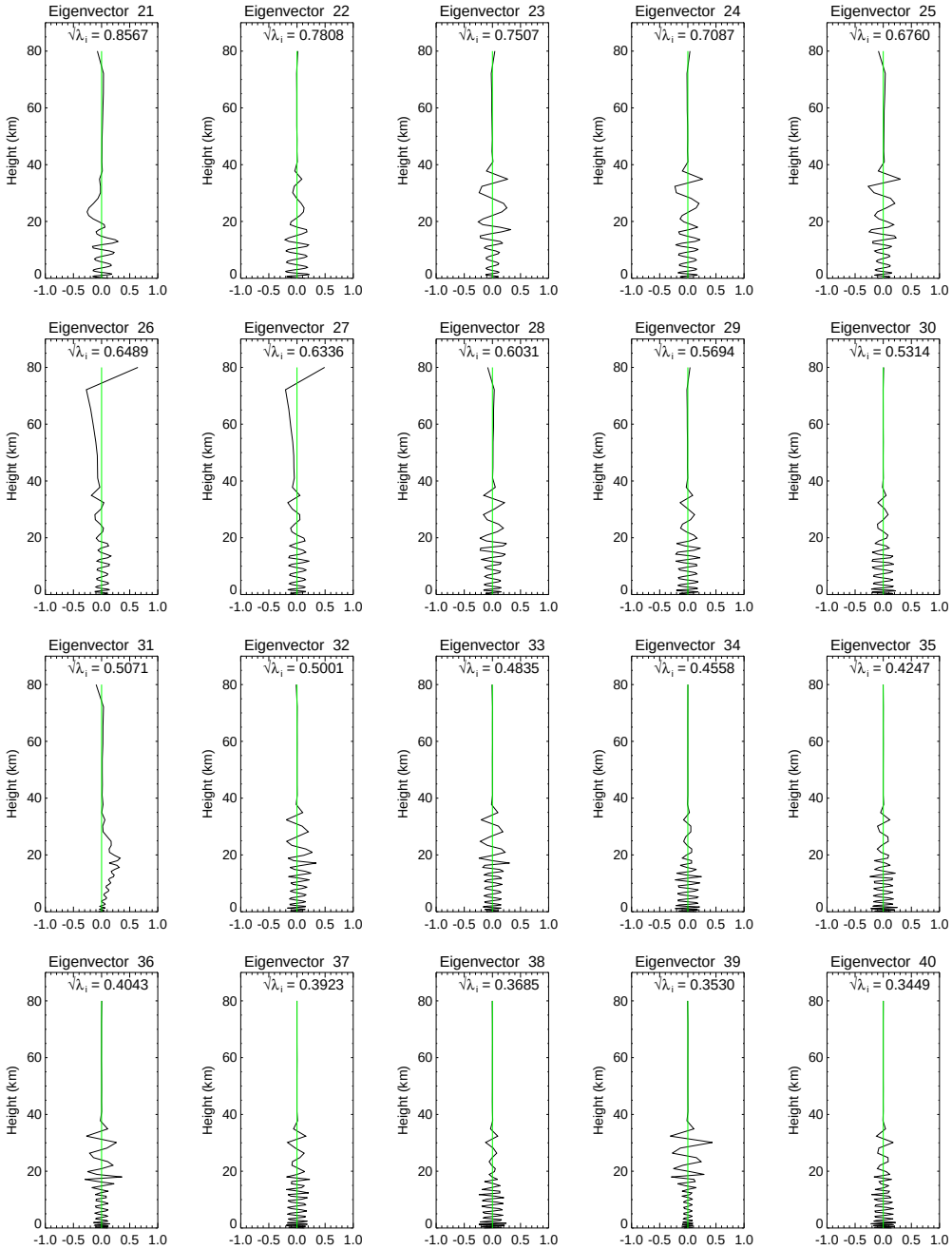


Figure C.2: Temperature eigenvectors 21–40 of the **NMC** $\mathbf{B}$ Matrix

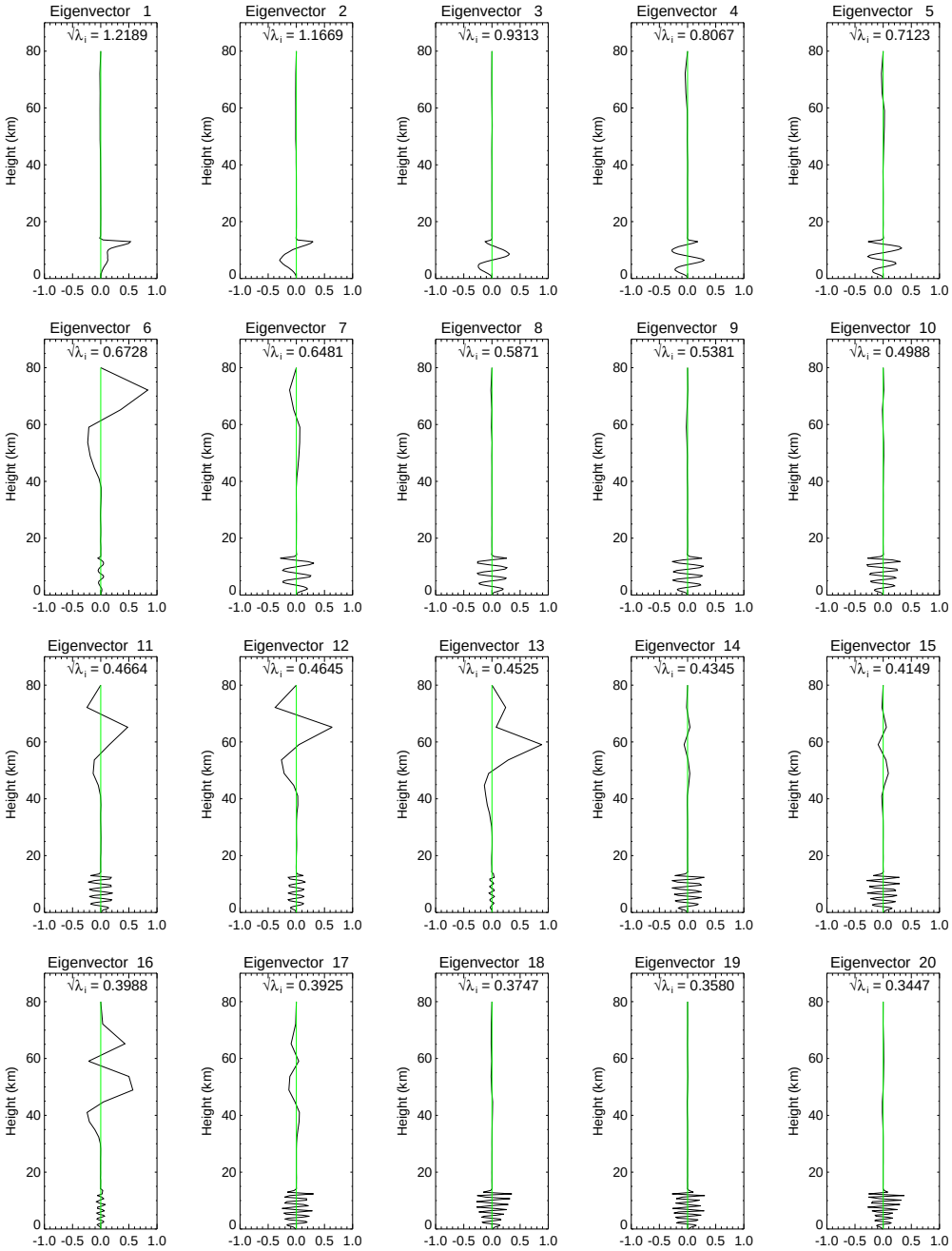


Figure C.3: Water vapour eigenvectors 1–20 of the **NMC** $\mathbf{B}$ Matrix

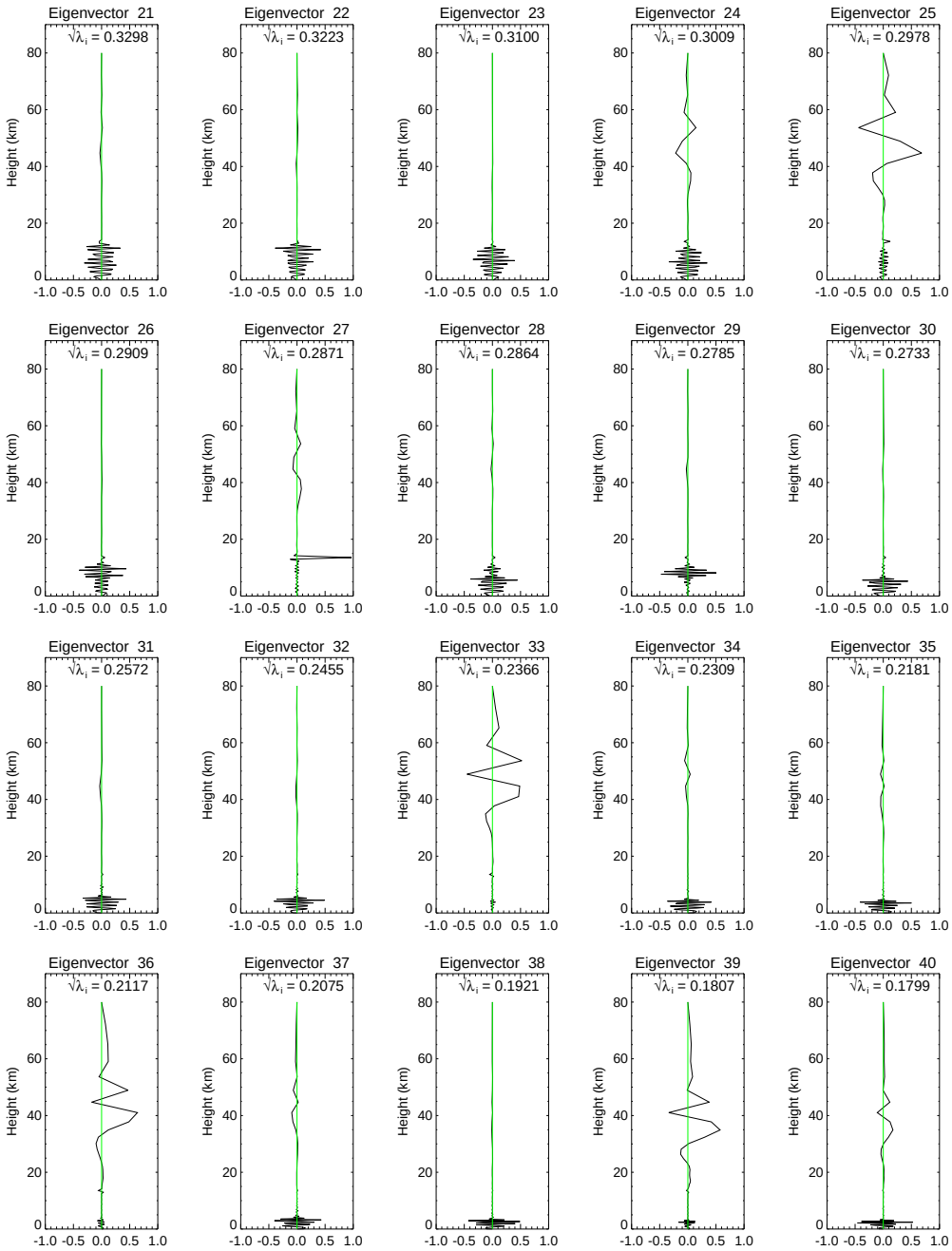


Figure C.4: Water vapour eigenvectors 21–40 of the **NMC** $\mathbf{B}$ Matrix

Appendix D

Derivation of the reconstruction error Φ_{rec}

This section repeats the derivation of the additional error, $\mathbf{\Phi}$, that results from the use of a forward model that calculates raw radiances rather than reconstructed radiances in section 6.3, but assuming that the true state is given by $\mathbf{y}_t$ rather than by $\tilde{\mathbf{y}}_t$. In this case, one would probably describe $\mathbf{\Phi}$ as a reconstruction error, so we will give it a subscript $\mathbf{\Phi}_{rec}$:

$$\begin{aligned} \delta\tilde{\mathbf{y}}' = & \quad (\tilde{\mathbf{y}} - \mathbf{y}_t) & - & \quad (H(\mathbf{x}) - \mathbf{y}_t) \\ = & \quad (\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{y} - \mathbf{y}_t) & - & \quad (H(\mathbf{x}) - \mathbf{y}_t)) \end{aligned} \tag{D.1}$$

and

$$\tilde{\mathbf{R}}' = \quad \mathbf{E}' \qquad + \quad \mathbf{F} \tag{D.2}$$

Let ϵ_{obs} be the observation error (instrument noise), then:

$$\mathbf{y} = \mathbf{y}_t + \epsilon_{obs} \tag{D.3}$$

The left-hand term in equation D.1 gives the reconstructed radiance departure from truth:

$$\begin{aligned} \delta\tilde{\mathbf{y}}_{obs} = & \quad (\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{y} - \mathbf{y}_t) \\ = & \quad \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}(\mathbf{y}_t + \epsilon_{obs}) - \mathbf{y}_t \\ = & \quad (\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}} - \mathbf{I})\mathbf{y}_t + \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\epsilon_{obs} \end{aligned} \tag{D.4}$$

Using $< \cdots >$ to represent expectation over climatological conditions

$$\mathbf{F}' = \quad < \delta\tilde{\mathbf{y}}_{obs}\, \delta\tilde{\mathbf{y}}_{obs}^{\mathrm{T}} > \tag{D.5}$$

$$\begin{aligned} = & \quad (\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}} - \mathbf{I}) < \mathbf{y}_t\mathbf{y}_t^{\mathrm{T}} > (\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}} - \mathbf{I})^{\mathrm{T}} & + & \quad \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{E}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}} \\ = & \quad \mathbf{\Phi}_{rec} & + & \quad \mathbf{L}_p\mathbf{L}_p^{\mathrm{T}}\mathbf{E}\mathbf{L}_p\mathbf{L}_p^{\mathrm{T}} \end{aligned} \tag{D.6}$$

Note that $\mathbf{\Phi}_{rec}$ in equation D.5 is the same as $\mathbf{\Phi}$ of equation 6.28 (when the

quadratic is multiplied out), and the total error budget is given by:

$$\tilde{\mathbf{R}}' = \mathbf{L}_p \mathbf{L}_p^{\mathrm{T}} \mathbf{E} \mathbf{L}_p \mathbf{L}_p^{\mathrm{T}} + \mathbf{F} + \mathbf{\Phi} \tag{D.7}$$

This equivalence assumes that the forward modelled radiance and the observed radiance are each given by an error fully characterised by the covariance matrix about the same true radiance. Biases often exist between observed and forward modelled radiances, but it is assumed that these biases are adequately corrected for before assimilation.

Appendix E

Channel Selections

E.1 The *Collard* channel selection

16	38	49	51	55	57	59	61	63	66	70	72
74	79	81	83	85	87	89	92	95	97	99	101
104	106	109	111	113	116	119	122	125	128	131	133
135	138	141	144	146	148	151	154	157	159	161	163
167	170	173	176	179	180	185	187	193	199	205	207
210	212	214	217	219	222	224	226	230	232	236	239
242	243	246	249	252	254	260	262	265	267	269	275
280	282	294	296	299	303	306	323	327	329	335	345
347	350	354	356	360	366	371	373	375	377	379	381
383	386	389	398	401	404	407	410	414	416	426	428
432	434	439	445	457	515	546	552	559	566	571	573
646	662	668	756	867	906	921	1027	1046	1121	1133	1191
1194	1271	1479	1509	1513	1521	1536	1574	1579	1585	1587	1626
1639	1643	1652	1658	1671	1786	1805	1884	1991	2019	2094	2119
2213	2239	2245	2271	2321	2398	2701	2741	2819	2889	2907	2910
2919	2939	2944	2948	2951	2958	2977	2985	2988	2991	2993	3002
3008	3014	3027	3029	3036	3047	3049	3053	3058	3064	3069	3087
3093	3098	3105	3107	3110	3127	3136	3151	3160	3165	3168	3175
3178	3207	3228	3244	3248	3252	3256	3263	3281	3303	3309	3312
3322	3339	3375	3378	3411	3438	3440	3442	3444	3446	3448	3450
3452	3454	3458	3467	3476	3484	3491	3497	3499	3504	3506	3509
3518	3522	3527	3540	3555	3575	3577	3580	3582	3586	3589	3599
3653	3658	3661	3943	4032	5130	5368	5371	5379	5381	5383	5397
5399	5401	5403	5405	5455	5480	5483	5485	5492	5502	5507	5509
5517	5558	5988	5992	5994	6003	6350	6463	6601	6962	6980	6982
6985	6987	6989	6991	6993	6995	6997	7267	7269	7424	7426	7428
7885	8007										

E.2 The *OPS* channel selection

16	38	49	51	55	57	59	61	63	66	70	72
74	79	81	83	85	87	89	95	97	99	101	104
106	109	111	113	116	119	122	125	128	131	133	135
138	141	144	146	148	151	154	157	159	161	163	167
170	173	176	179	180	185	187	193	199	205	207	210
212	214	217	219	222	224	226	230	232	236	239	242
243	246	249	252	254	260	262	265	267	269	275	280
282	294	296	299	303	306	323	327	329	335	345	347
350	354	356	360	366	371	373	375	377	379	381	383
386	389	398	401	404	407	410	414	416	426	428	432
434	439	445	457	515	546	552	566	571	573	646	662
668	756	867	906	921	1027	1046	1121	1133	1191	1194	1271
1786	1805	1884	1991	2019	2094	2119	2239	2245	2741	2889	2907
2944	2948	2951	2958	2988	3027	3029	3053	3058	3064	3168	3248
3506	3577	3582	3589	5130	5368	5371	5379	5381	5383	5397	5399
5401	5403	5405									

E.3 The *VAR* channel selection

38	51	57	63	109	116	122	128	135	141	148	154
161	167	173	179	180	185	187	193	199	205	207	210
212	214	217	219	222	224	226	230	232	236	239	242
243	246	249	252	254	260	262	265	267	269	275	280
282	294	296	299	306	323	327	329	335	345	347	350
354	356	360	366	371	373	375	377	379	381	383	386
389	398	401	404	407	410	414	416	426	428	432	434
439	445	457	515	546	552	566	571	573	662	668	756
867	921	1027	1133	1194	1271	1805	1884	1991	2094	2239	2245
2741	2889	2907	2944	2948	2951	2958	2988	3027	3029	3053	3058
3064	3168	3248	3506	3577	3582	3589	5130	5368	5371	5379	5381
5383	5397	5399	5401	5403	5405						

E.4 The *Method 1* channel selections for R_{HL}

E.4.1 Original method (*Lp-M1*)

2	10	15	21	26	32	71	86	92	94	97	102
125	154	192	212	243	262	292	299	302	306	314	326
338	340	345	361	387	393	399	437	449	497	527	546
580	588	591	604	616	629	660	679	697	722	742	763
821	832	847	892	912	930	936	953	973	1005	1056	1073
1136	1219	1276	1290	1315	1374	1443	1479	1521	1539	1555	1560
1586	1594	1646	1668	1686	1696	1711	1730	1791	1828	1842	1860
1905	1951	1960	1964	1966	1991	2001	2007	2015	2020	2026	2034
2041	2053	2064	2070	2075	2082	2094	2132	2151	2157	2166	2168
2170	2187	2199	2217	2227	2238	2245	2261	2266	2268	2270	2272
2278	2291	2302	2306	2319	2323	2328	2332	2342	2359	2379	2399
2402	2411	2422	2426	2430	2439	2461	2468	2475	2486	2503	2509
2516	2536	2540	2550	2554	2557	2562	2585	2589	2617	2632	2640
2645	2647	2661	2675	2684	2692	2699	2708	2712	2716	2721	2724
2733	2746	2757	2763	2771	2777	2782	2802	2807	2814	2819	2821
2837	2861	2867	2870	2875	2879	2885	2888	2900	2907	2913	2917
2940	2947	2964	2971	2985	3002	3074	3111	3248	3252	3390	3412
3580	3817	4209	4533	4873	5021						

E.4.2 Modified method (*RHL-M1*)

698	616	2	2001	16	10	92	832	21	72	588	94
32	339	97	387	936	399	1842	497	302	243	1219	1521
26	1966	299	313	2042	3014	1964	678	2034	2075	1596	292
87	1686	102	2027	578	358	2150	973	345	2010	2065	1290
558	953	2018	306	912	449	1586	2053	439	1951	547	1242
1661	326	3028	1991	1695	192	632	847	812	591	1904	1554
1712	125	930	2168	1276	1859	1005	1479	262	3244	526	2070
1960	1056	1730	2171	892	154	379	721	1560	1630	648	1931
2094	2211	394	1089	2746	2888	1539	763	2081	1373	2347	1802
232	1434	2717	2329	2404	2320	2270	2868	2378	1315	2134	2426
2724	3251	2685	2267	2341	2535	2227	2430	2589	2777	2246	2323
2814	881	2644	2675	2461	2278	2182	2800	2238	2486	2856	2411

2273 2561 2806 2558 2157 2439 3098 2409 2699 2940 2750 2302
2571 2467 2649 2866 2733 2475 2399 5021 2764 2551 2832 2174
2617 2640 2878 2264 2721 2782 2632 3074 2306 2920 2860 3268
2907 2287 2766 2564 3002 2365 3580 2198 2692 3773 2913 2947
2445 2731 2757 2504 2219 2508 2971 2556 2984 4699 3435 2516
2900 2816 4565 2489 3201 5104

E.5 The *Method 2* channel selections for R_{HL}

E.5.1 Jacobians – 8 Clear; Threshold – 1.0. (*RHL-M2-8J-1.0*)

2 9 15 21 26 32 37 50 63 71 86 91
94 98 103 125 180 199 242 259 298 302 306 314
325 338 345 361 381 387 395 399 440 451 526 544
584 589 607 616 635 641 688 783 819 832 836 893
906 927 935 957 970 976 1007 1086 1120 1194 1270 1289
1351 1392 1424 1479 1502 1534 1539 1554 1560 1587 1589 1601
1621 1646 1662 1684 1686 1694 1712 1720 1761 1828 1844 1854
1906 1908 1925 1964 1965 1988 1998 2008 2018 2020 2031 2038
2041 2052 2067 2076 2082 2110 2119 2145 2186 2199 2212 2227
2266 2278 2324 2329 2342 2346 2370 2379 2398 2411 2454 2475
2497 2508 2515 2554 2562 2578 2589 2616 2640 2645 2646 2670
2683 2689 2714 2724 2729 2732 2745 2751 2771 2779 2801 2808
2861 2866 2873 2886 2893 2902 2921 2925 2948 2955 2977 3015
3058 3117 3247 3352 3449 3546 3715 3809 3826 3888 4032 4034
4037 4042 4121 4158 4203 4207 4209 4213 4245 4291 4326 4341
4359 4381 4426 4508 4588 4592 4605 4607 4620 4627 4691 4708
4740 4742 4770 4798 4803 4832 4890 4898 4938 4949 4957 4979
5003 5026 5035 5038 5093 5108

E.5.2 Jacobians – 8 Clear; Threshold – 1.3 (*RHL-M2-8J-1.3*)

1 2 13 14 16 17 18 19 22 23 24 25
34 37 40 41 42 44 50 56 67 68 69 71
72 77 87 88 89 90 91 92 93 94 95 96
97 100 101 102 109 111 120 121 124 125 126 127
129 130 132 137 138 139 140 141 143 144 145 155

157 160 162 165 166 167 169 170 172 173 174 175
179 180 189 191 192 193 194 195 196 197 198 199
203 205 206 207 208 209 210 221 222 223 226 231
235 242 243 248 249 252 253 261 266 269 274 275
276 278 295 296 297 298 299 300 304 336 378 382
385 386 387 390 395 397 410 414 415 431 449 587
732 840 1587 2018 2270 2488 2507 2508 2509 2510 2612 2623
2636 2648 2654 2670 2672 2676 2689 2711 2733 2744 2760 2770
2819 2842 2919 2926 2927 2933 2959 3117 3194 3195 3196 3249
3327 3335 3410 3442 3452 3580 3601 3787 3829 3966 3972 3974
4008 4032 4635 5025 5048 5094

E.5.3 Jacobians – 8 Clear; Threshold – 100.0 (*RHL-M2-8J-100.0*)

5 18 19 21 92 94 100 135 136 144 173 190
217 224 418 1602 1616 2270 2323 2378 2455 2464 2465 2489
2499 2502 2507 2508 2509 2510 2511 2512 2606 2607 2630 2633
2652 2670 2671 2672 2676 2689 2706 2760 2818 2842 2843 2925
2928 2932 2933 2984 3008 3009 3021 3069 3070 3071 3072 3073
3074 3075 3117 3118 3134 3135 3143 3155 3161 3224 3264 3278
3301 3327 3423 3448 3488 3509 3523 3549 3553 3562 3584 3587
3588 3631 3662 3670 3765 3776 3779 3807 3815 3846 4635 4930
5067 5072 5073

E.5.4 Jacobians – 8 Clear 2 Zen; Threshold – 1.3 (*RHL-M2-8J2Z-1.3*)

1 2 5 16 29 31 32 34 35 36 37 38
46 47 48 60 62 64 70 71 72 73 83 84
86 87 89 90 91 92 93 94 95 96 98 99
100 101 103 105 107 108 109 118 119 120 124 130
137 139 142 143 147 149 150 152 155 156 157 159
162 166 173 176 181 183 184 189 193 199 213 214
215 216 225 231 235 247 248 249 252 253 263 274
275 284 296 300 303 317 320 324 330 334 351 355
379 384 385 389 390 397 402 454 505 507 520 548

587 604 622 653 688 694 819 842 850 1097 1331 1403
1404 1531 1563 1576 1686 1941 1973 2437 2488 2496 2507 2508
2509 2510 2555 2569 2570 2572 2584 2634 2648 2654 2655 2670
2671 2672 2674 2675 2678 2733 2778 2933 2992 3004 3008 3117
3222 3247 3250 3252 3318 3462 3591 3780 3782 3972 4032 4037
4088 4162 4333 4352 4359 4560 4636 4798 4801 5047

E.5.5 Jacobians – 11 Cloudy A; Threshold – 1.3 (*RHL-M2-11J-1.3*)

1 6 10 16 26 37 57 59 70 71 73 87
88 89 92 95 96 97 99 100 101 102 132 135
136 138 168 172 174 178 191 194 199 203 228 278
296 440 1579 1650 2325 2462 2488 2494 2495 2498 2499 2500
2501 2502 2503 2506 2507 2508 2509 2510 2571 2584 2602 2606
2607 2608 2623 2636 2646 2649 2652 2654 2655 2656 2670 2671
2672 2675 2676 2677 2678 2679 2680 2689 2690 2691 2695 2697
2699 2700 2701 2702 2703 2710 2731 2760 2788 2816 2842 2843
2864 2873 2874 2926 2928 2933 2942 2947 2951 2954 2995 2996
2997 3009 3011 3012 3022 3023 3040 3047 3056 3082 3083 3090
3095 3123 3152 3197 3198 3208 3228 3229 3230 3231 3246 3247
3248 3249 3250 3251 3261 3262 3315 3318 3322 3324 3326 3327
3451 3452 3453 3460 3461 3469 3506 3508 3509 3524 3528 3537
3555 3562 3563 3580 3585 3588 3589 3590 3612 3621 3635 3671
3681 3726 3749 3806 3966 3996 4032 4043 4094 4150 4380 4514
4561 4674 4883 5101

E.5.6 Jacobians – 9 Cloudy B; Threshold – 1.3 (*RHL-M2-9J-1.3*)

8 9 10 11 12 16 26 30 37 54 55 57
71 83 84 88 89 90 91 92 93 94 95 96
97 101 103 105 113 134 138 141 171 178 180 189
192 199 202 220 278 296 304 431 612 1579 1609 1632
1637 2460 2461 2463 2464 2482 2486 2487 2492 2493 2495 2496
2500 2501 2502 2503 2505 2506 2507 2508 2509 2510 2511 2512
2568 2569 2570 2571 2572 2573 2574 2579 2583 2598 2636 2670

2671 2672 2675 2676 2689 2690 2710 2712 2761 2768 2769 2771
2772 2787 2789 2818 2839 2840 2841 2842 2843 2864 2868 2872
2889 2896 2900 2913 2923 2924 2925 2928 2933 3002 3009 3013
3015 3044 3045 3097 3116 3120 3150 3157 3164 3168 3169 3229
3239 3244 3245 3246 3247 3248 3249 3250 3251 3255 3259 3291
3325 3326 3327 3328 3329 3334 3346 3349 3359 3362 3402 3407
3430 3459 3460 3461 3462 3463 3469 3470 3482 3488 3492 3494
3495 3506 3526 3528 3534 3537 3540 3570 3580 3583 3588 3597
3615 3621 3623 3681 3729 3742 3767 3768 3787 3808 3857 3966
3972 3979 4030 4032 4034 4099 4107 4342 4355 4465 4576 4640
4676 4784 5066 5101

E.5.7 Jacobians – 14 Cloudy C; Threshold – 1.3 (*RHL-M2-14J-1.3*)

1 11 17 38 63 70 71 73 78 88 90 91
92 93 96 101 103 116 117 123 124 125 126 170
180 192 193 205 243 293 381 430 1587 2459 2460 2461
2462 2468 2483 2505 2506 2507 2508 2509 2522 2570 2571 2576
2580 2583 2585 2606 2607 2628 2640 2670 2671 2672 2676 2679
2689 2690 2691 2692 2693 2694 2767 2787 2793 2855 2889 2925
2928 2933 3002 3008 3009 3011 3013 3021 3042 3053 3090 3096
3116 3121 3136 3144 3149 3157 3164 3168 3169 3170 3198 3229
3246 3247 3250 3251 3255 3283 3284 3287 3294 3300 3302 3308
3318 3319 3326 3327 3328 3329 3334 3365 3373 3374 3411 3419
3451 3461 3470 3502 3503 3504 3506 3509 3523 3550 3555 3563
3587 3588 3606 3617 3621 3645 3651 3658 3668 3670 3685 3723
3806 3850 3859 3923 3924 3945 3962 3991 3999 4028 4029 4030
4033 4160 4247 4248 4277 4289 4565 4571 4860 4865 4878 5054
5095

E.5.8 Jacobians – 8 Clear; Threshold – 1.3; Added Water Vapour Error for 100 channels (*RHL-M2-8J-1.3-WVE*)

1 2 4 11 16 23 37 41 43 55 60 65
69 71 72 81 82 85 87 88 89 90 91 92
93 94 95 97 98 99 100 101 102 104 105 106

107 108 112 118 119 121 125 126 128 129 130 131
132 133 137 138 139 141 143 145 157 158 159 161
163 164 166 169 170 171 172 176 177 180 181 182
183 184 186 187 188 189 190 191 192 194 195 196
198 199 200 202 205 216 217 218 219 220 221 228
238 240 241 243 244 245 246 247 248 249 250 251
252 265 266 268 274 276 277 278 280 281 294 295
296 297 298 301 306 309 328 335 336 337 338 339
340 341 342 343 387 388 389 392 401 403 415 440
654 839 1609 1619 2520 2637 2676 2830 2841 2905 2911 2928
2933 2945 2949 2952 2991 3008 3027 3169 3191 3207 3209 3396
3397 3398 3399 3400 3469 3653 3658 3757 4103 4223 4465 4509
4560 4742 4751 5073

E.6 The *Method 2* channel selections for $\tilde{\mathrm{N}}$

E.6.1 Jacobians – 8 Clear; Threshold – 1.0 (*Ñ-M2-8J-1.0*)

93 107 133 159 195 210 215 248 273 282 317 323
370 408 428 479 503 519 543 560 568 638 666 703
715 728 768 770 786 823 858 883 885 921 948 985
1001 1032 1044 1075 1096 1097 1128 1134 1151 1199 1203 1228
1233 1238 1284 1294 1305 1321 1328 1348 1359 1371 1401 1415
1419 1440 1441 1449 1482 1490 1496 1519 1523 1541 1568 1610
1615 1627 1638 1655 1673 1698 1729 1749 1766 1774 1816 1839
1855 1894 1900 1919 1948 1976 1998 2012 2021 2027 2044 2054
2063 2070 2142 2159 2165 2177 2190 2204 2216 2243 2252 2257
2262 2265 2269 2272 2280 2293 2295 2298 2301 2313 2329 2334
2340 2347 2361 2367 2375 2380 2393 2407 2413 2418 2432 2442
2453 2456 2461 2467 2476 2484 2491 2504 2510 2518 2523 2526
2528 2536 2547 2555 2567 2579 2588 2601 2608 2618 2627 2643
2654 2661 2668 2695 2701 2716 2721 2726 2729 2733 2739 2744
2759 2764 2769 2779 2783 2796 2804 2821 2826 2835 2846 2849
2862 2869 2880 2902 2904 2915 2924 2938 2947 2977 2999 3001
3029 3056 3080 3087 3252 3353 3449 3570 3575 3611 3816 4190
4208 4528 4691 4700 4934 5099

E.6.2 Jacobians – 8 Clear; Threshold – 1.3 (*Ñ-M2-8J-1.3*)

6	17	46	75	81	89	90	92	93	95	96	99
105	106	107	135	136	142	143	144	163	167	169	175
186	199	201	215	216	234	235	244	263	267	281	298
300	317	339	381	385	408	418	456	488	521	551	555
559	572	597	599	628	654	656	715	739	769	823	870
876	882	925	931	945	953	988	1056	1118	1304	1331	1367
1388	1441	1467	1496	1537	1580	1615	1638	1645	1648	1688	1719
1766	1824	1833	1894	1945	1971	2073	2146	2172	2200	2204	2214
2232	2273	2276	2282	2295	2298	2305	2314	2323	2349	2352	2353
2354	2355	2368	2369	2378	2381	2393	2397	2410	2412	2428	2442
2449	2451	2459	2494	2496	2505	2515	2528	2541	2557	2563	2571
2574	2584	2612	2616	2636	2638	2643	2656	2661	2667	2671	2675
2694	2695	2701	2712	2727	2733	2736	2738	2739	2740	2755	2760
2768	2771	2775	2806	2827	2830	2835	2856	2865	2872	2875	2881
2885	2896	2912	2935	2959	2979	2980	2982	2983	2994	3015	3017
3061	3064	3078	3094	3125	3126	3152	3201	3203	3257	3300	3309
3311	3335	3338	3350	3426	3445	3460	3504	3611	3634	3662	3673
3794	3943	3958	4005	4140	4699						

E.7 The *PC-RTTOV* 500 predictor channels

8	69	98	114	118	124	135	138	142	152	161	162
165	174	176	179	181	182	184	186	187	188	197	200
206	209	214	220	225	226	237	240	241	246	251	252
254	263	265	270	272	274	277	281	289	301	302	303
307	309	311	323	324	325	339	340	345	346	365	369
392	394	398	405	415	416	440	441	442	480	496	497
505	561	578	587	600	616	617	654	721	732	1393	1416
1420	1453	1465	1474	1493	1503	1536	1566	1613	1646	1661	1719
2216	2220	2239	2258	2265	2266	2273	2283	2296	2309	2317	2319
2320	2322	2324	2342	2343	2391	2398	2399	2400	2401	2405	2409
2426	2427	2432	2440	2467	2479	2482	2483	2485	2493	2494	2497
2499	2503	2510	2511	2514	2516	2518	2520	2530	2533	2535	2544
2550	2554	2555	2557	2560	2563	2564	2565	2573	2579	2593	2595
2597	2608	2610	2613	2633	2636	2637	2642	2645	2649	2651	2653

2661 2668 2673 2676 2677 2684 2686 2693 2694 2695 2697 2698
2699 2701 2702 2707 2710 2715 2720 2723 2726 2732 2746 2750
2751 2755 2760 2767 2781 2783 2786 2794 2801 2808 2811 2815
2820 2821 2822 2835 2836 2842 2865 2875 2886 2888 2890 2913
2914 2915 2919 2920 2923 2924 2926 2932 2955 2984 2998 2999
3000 3002 3017 3021 3025 3030 3052 3068 3078 3080 3097 3098
3101 3102 3105 3107 3110 3142 3149 3167 3169 3171 3245 3246
3247 3248 3249 3252 3253 3254 3256 3279 3284 3320 3325 3370
3371 3406 3416 3444 3445 3446 3448 3451 3452 3453 3459 3460
3487 3488 3502 3510 3542 3554 3566 3571 3576 3577 3579 3581
3621 3647 3648 3651 3661 3664 3679 3700 3740 3744 3767 3787
3795 3848 3861 3949 3972 3982 4020 4106 4123 4131 4140 4190
4204 4239 4340 4440 4478 4495 4606 4631 4641 4663 4677 4778
4797 4838 5035 5036 5102 5169 5370 5375 5382 5384 5386 5388
5389 5391 5393 5394 5402 5415 5416 5417 5418 5419 5420 5421
5422 5423 5435 5440 5485 5489 5491 5497 5528 5529 5530 5536
5547 5571 5574 5580 5581 5594 5620 5645 5648 5657 5664 5665
5666 5672 5681 5682 5683 5689 5696 5710 5711 5714 5716 5720
5722 5726 5732 5737 5738 5774 5782 5794 5810 5823 5824 5826
5829 5835 5845 5849 5882 5886 5887 5893 5899 5910 5932 5938
5955 5974 6030 6031 6056 6069 6106 6124 6179 6197 6222 6230
6238 6243 6257 6276 6338 6348 6369 6371 6378 6387 6396 6412
6420 6428 6446 6455 6461 6464 6472 6474 6476 6477 6488 6496
6500 6509 6512 6523 6531 6542 6551 6554 6564 6578 6581 6593
6596 6598 6602 6603 6604 6606 6624 6632 6655 6666 6667 6669
6674 6686 6687 6696 6710 6714 6717 6719 6723 6736 6745 6762
6765 6775 6804 6874 6875 6890 6913 6929 6972 7036 7102 7247
7416 7795 7808 7851 7909 8019 8301 8302

Bibliography

Aires, F., W. B. Rossow, Scott N. A., and Chédin (2002) "Remote sensing from the infrared atmospheric sounding interferometer instrument. 1. Compression, denoising, and first-guess retrieval algorithms," *J. Geophys. Res*, Vol. 107, DOI: http://dx.doi.org/10.1029/2001JD000955.

Andersson, E., M. Fisher, R. Munro, and A. McNally (2000) "Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation scheme, and the explanation of a case of poor convergence," *Q. J. R. Meteorol Soc.*, Vol. 126, pp. 1455–1472.

Antonelli, P., H. E. Revercomb, L. A. Sromovsky, W. L. Smith, R. O. Knuteson, D. C. Tobin, R. K. Garcia, H. B. Howell, H.-L. Huang, and F. A. Best (2004) "A principal component noise filter for high spectral resolution infrared measurements," *J. Geophys. Res.*, Vol. 109, DOI: http://dx.doi.org/10.1029/2004JD004862.

Baker, N. L. and R. Daley (2000) "Observation and background adjoint sensitivity in the adaptive observation-targetting problem," *Q. J. R. Meteorol Soc.*, Vol. 126, pp. 1431–1454.

Bormann, N. and P. Bauer (2010) "Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data," *Q. J. R. Meteorol. Soc.*, Vol. 136, pp. 1036–1050, DOI: http://dx.doi.org/10.1002/qj.616.

Bormann, N., A. Collard, and P. Bauer (2010) "Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. II: Methods and application to AIRS and IASI data," *Q. J. R. Meteorol. Soc.*, Vol. 136, pp. 1051–1063, DOI: http://dx.doi.org/10.1002/qj.615.

Bouttier, F. and P. Courtier (1999) "Data Assimilation Concepts and Methods," ECMWF Meteorological Training Course Lecture Notes.

Bowler, N.E., A. Arribas, K.R. Mylne, K.B. Robertson, and S.E. Beare (2008) "The MOGREPS short-range ensemble prediction system," *Q. J. R. Meteorol. Soc.*, Vol. 134, pp. 703–722, DOI: http://dx.doi.org/10.1002/qj.234.

Bowler, N.E., J. Flowerdew, and S.R. Pring (2013) "Tests of different flavours of EnKF on a simple model," *Q. J. R. Meteorol. Soc.*, Vol. 139, pp. 1505–1519, DOI: http://dx.doi.org/10.1002/qj.2055.

Cameron, J. and J. Eyre (2008) "Application of the observational method to IASI data," Unpublished Met Office Technical Report.

Cardinali, C. (2009) "Monitoring the observation impact on the short-range forecast," *Q. J. R. Meteorol. Soc.*, Vol. 135, pp. 239–250, DOI: http://dx.doi.org/10.1002/qj.366.

Chalon, G., F. Cayla, and D. Diebel (2001) "IASI: An Advanced Sounder for Operational Meteorology," in *Proceedings of the 52nd Congress of IAF, Toulouse France, 1-5 Oct. 2001.*

Clayton, A. M., A. C. Lorenc, and D. M. Barker (2012) "Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office," *Q. J. R. Meteorol. Soc.*, Vol. 139, pp. 1445–1461, Accepted 13 September 2012.

Collard, A. D. (1998) "Notes on IASI performance,"Technical Report 256, Met Office.

——— (2007) "Selection of IASI channels for use in numerical weather prediction," *Q. J. R. Meteorol. Soc.*, Vol. 133, pp. 1977–1991, DOI: http://dx.doi.org/10.1002/qj.178.

——— (2012) "IASI Channel Selection for Reconstructed Radiances," NWPSAF Visiting Scientist report NWPSAF-MO-VS-047 http://research.metoffice.gov.uk/research/interproj/nwpsaf/vs.html.

Collard, A. D., F. I. Hilton, M. J. Forsythe, and B. Candy (2011) "From Observations to Forecasts — Part x. Satellite Observations in Numerical Weather Prediction," *Weather*, Vol. 66, pp. 31–36, DOI: http://dx.doi.org/10.1002/wea.736.

Collard, A. D. and A. P. McNally (2009) "The assimilation of Infrared Atmospheric Sounding Interferometer radiances at ECMWF," *Q. J. R. Meteorol. Soc.*, Vol. 135, pp. 1044–1058, DOI: http://dx.doi.org/10.1002/qj.410.

Collard, A. D., A. P. McNally, F. I. Hilton, S. B. Healy, and N. C. Atkinson (2010) "The Use of Principal Component Analysis for the Assimilation of High-Resolution Infrared Sounder Observations for Numerical Weather Prediction,"

Q. J. R. Meteorol. Soc., Vol. 136, pp. 2038–2050, DOI: http://dx.doi.org/10.1002/qj.701.

Courtier, P., Thépaut. J-N., and A. Hollingsworth (1994) "A strategy for operational implementation of 4D-Var, using an incremental approach," *Q. J. R. Meteorol. Soc.*, Vol. 120, pp. 1389–1408.

Daescu, D. N. and R. H. Langland (2013) "Error covariance sensitivity and impact estimation with adjoint 4D-Var: theoretical aspects and first applications to NAVDAS-AR," *Q. J. R. Meteorol. Soc.*, Vol. 139, pp. 226–241.

Daley, R. (1991) *Atmospheric Data Analysis*: Cambridge University Press, Cambridge, UK.

Daniels, M. J. and R. E. Kass (2001) "Shrikage Estimators for Covariance Matrices," *Biometrics*, Vol. 57, pp. 1173–1184, DOI: http://dx.doi.org/10.1111/j.0006-341X.2001.01173.x.

Dee, D.P. (2005) "Bias and data assimilation," *Q. J. R. Meteorol. Soc.*, Vol. 131, pp. 3323–3342, DOI: http://dx.doi.org/10.1256/qj.05.137.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli (2005) "Diagnosis of observation, background and analysis error statistics in observation space," *Q. J. R. Meteorol. Soc.*, Vol. 131, pp. 3385–3396.

Efron, B. and C. Morris (1977) "Stein's paradox in statistics," *Scientific American*, Vol. 236, pp. 119–127, DOI: http://dx.doi.org/10.1038/scientificamerican0577-119.

Evensen, G. (2007) *Data assimilation: the ensemble Kalman filter*: Springer, Berlin, pp.279.

Eyre, J. R. and A. D. Collard (1999) "The effects of nonlinearity on retrieval errors: implications for the interpretation of advanced infra-red sounder data," in *Proceedings of the 4th International TOVS Study Conference, 27 January–2 February 1999, Boulder, Colorado, USA*.

Eyre, J. R. and F. I. Hilton (2013) "Sensitivity of analysis error covariance to the mis-specification of background error covariance," *Q. J. R. Meteorol. Soc.*, Vol. 139, pp. 524–533, DOI: http://dx.doi.org/10.1002/qj.1979.

Fairbairn, D. (2014) "Comparing Variational and Ensemble Data Assimilation Methods for Numerical Weather Prediction," Ph.D. dissertation, University of Surrey, Guildford, UK.

Fairbairn, D., S.R. Pring, A.C. Lorenc, and I. Roulstone (2014) "A comparison of 4DVar with ensemble data assimilation methods," *Q. J. R. Meteorol. Soc.*, Vol. 140, pp. 281–294, DOI: http://dx.doi.org/10.1002/qj.2135.

Fisher, M. (1997) "Efficient Minimization of Quadratic Penalty Functions," Unpublished manuscript.

——— (2003) "Background error covariance modelling," in *ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, 8–12 September 2003*, pp. 45–64, http://old.ecmwf.int/publications/library/do/references/list/17111.

Fisher, M. and H. Auvinen (2011) "Long Window 4D-Var," in *ECMWF Seminar on Data Assimilation for Atmosphere and Ocean, 6–9 September 2011*, pp. 45–64, http://old.ecmwf.int/publications/library/do/references/list/3052012.

Fisher, M., M. Leutbecher, and G. A. Kelly (2006) "On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation," *Q. J. R. Meteorol. Soc.*, Vol. 131, pp. 3235–3246, DOI: http://dx.doi.org/10.1002/qj.04.142.

Gadd, A. J., B. R. Barwell, S. J. Cox, and R. J. Renshaw (1995) "Global processing of satellite sounding radiances in numerical weather prediction," *Q. J. R. Meteorol. Soc.*, Vol. 121, pp. 615–630, DOI: http://dx.doi.org/10.1002/qj.49712152308.

Garand, L., N. Baker, and C. Clerbaux (2009) "Working group reports," in *ECMWF/EUMETSAT NWP-SAF Workshop on Assimilation of IASI in NWP, 6–8 May 2009*, http://old.ecmwf.int/publications/library/ecpublications/_pdf/workshop/2009/IASI/.

Garand, L., S. Heilliette, and M. Buehner (2007) "Interchannel Error Correlation Associated with AIRS Radiance Observations: Inference and Impact in Data Assimilation," *J. Applied Meteorology*, Vol. 46, pp. 714–725.

Gelaro, R. and Y. Zhu (2009) "Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models," *Tellus*, Vol. 61A, pp. 179–193.

Guidard, V., P. Brousseau, N. Fourrié, and F. Rabier (2010) "Impact of IASI assimilation in convective scale model AROME," `http://smsc.cnes.fr/IASI/PDF/conf2/session3/Guidard.pdf`.

Han, W. and A.P. McNally (2010) "The 4D-Var assimilation of ozone-sensitive infrared radiances measured by IASI," *Q. J. R. Meteorol. Soc.*, Vol. 136, pp. 2025–2037, DOI: `http://dx.doi.org/10.1002/qj.708`.

Harris, B. A. and G. Kelly (2001) "A satellite radiace-bias correction scheme for data assimilation," *Q. J. R. Meteorol. Soc.*, Vol. 127, pp. 1453–1468.

Hilton, F., P. Antonelli, X. Calbet, T. Hultberg, L. Lavanant, X. Liu, G. Masiello, S. Newman, J. Taylor, C. Serio, and D. Zhou (2009b) "An investigation into the performance of retrievals of temperature and humidity from IASI," in *Proceedings of the Eumetsat Meteorological Satellite Conference, Bath, UK, 21-25 September 2009*.

Hilton, F., R. Armante, T. August, C. Barnet, A. Bouchard, C. Camy-Peyret, V. Capelle, L. Clarisse, C. Clerbaux, P.-F. Coheur, A. Collard, C. Crevoisier, G. Dufour, D. Edwards, F. Faijan, N. Fourrié, A. Gambacorta, M. Goldberg, V. Guidard, D. Hurtmans, S. Illingworth, N. Jacquinet-Husson, T. Kerzenmacher, D. Klaes, L. Lavanant, G. Masiello, M. Matricardi, A. McNally, S. Newman, E. Pavelin, S. Payan, E. Péquignot, S. Peyridieu, T. Phulpin, J. Remedios, P. Schlüssel, C. Serio, L. Strow, C. Stubenrauch, J. Taylor, D. Tobin, W. Wolf, and D. Zhou (2012a) "Hyperspectral Earth Observation from IASI," *Bull. Am. Meteorol. Soc.*, Vol. 93, pp. 347–370, DOI: `http://dx.doi.org/10.1175/BAMS-D-11-00027.1`.

Hilton, F., N. C. Atkinson, S. J. English, and J. R. Eyre (2009a) "Assimilation of IASI at the Met Office and assessment of its impact through observing system experiments," *Q. J. R. Meteorol. Soc.*, Vol. 135, pp. 495–505, DOI: `http://dx.doi.org/10.1002/qj.379`.

Hilton, F., A. Collard, V. Guidard, R. Randriamampianina, and M. Schwaerz (2009c) "Assimilation of IASI radiances at European NWP Centres," in *ECMWF/EUMETSAT NWP-SAF Workshop on Assimilation of IASI*

in NWP, 6–8 May 2009, http://old.ecmwf.int/publications/library/ecpublications/_pdf/workshop/2009/IASI/.

Hilton, F. I. and A. D. Collard (2009) "Recommendations for the use of principal component-compressed observations from infrared hyperspectral sounders,"Technical Report 536, Met Office.

Hilton, F. I., S. M. Newman, and A. D. Collard (2012b) "Identification of NWP humidity biases using high-peaking water vapour channels from IASI," *Atmos. Sci. Lett.*, Vol. 13, pp. 73–28, DOI: http://dx.doi.org/10.1002/asl.36.

Hollingsworth, A. and P. Lönnberg (1986) "The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field," *Tellus*, Vol. 38A, pp. 111–136.

Hultberg, T. (2009) "IASI Principal Component Compression (IASI PCC) FAQ,"Technical report, EUMETSAT, Available from http://www.eumetsat.int/website/home/Data/Products/Level1Data/index.html.

Hultberg, T. and T. August (2013) "On the equivalence of reconstructed radiances and PC scores," Unpublished EUMETSAT internal report.

Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc (1997) "Unified notation for data assimilation: operational, sequential and variational," *J. Meteorol. Soc. Japan*, Vol. 75, pp. 181–189.

Ingleby, B., A. C. Lorenc, K. Ngan, F. Rawlins, and D. R. Jackson (2013) "Improved variational analyses using a nonlinear humidity control variable," *Q. J. R. Meteorol. Soc.*, Vol. 139, pp. 1875–1887, DOI: http://dx.doi.org/10.1002/qj.2073.

Joo, S., J. Eyre, and R. Marriott (2013) "The impact of Metop and other satellite data within the Met Office global NWP system using an adjoint-based sensitivity method," *Mon. Wea. Rev.*, Vol. 141, pp. 3331–3342, DOI: http://dx.doi.org/:10.1175/MWR-D-12-00232.1.

Kalman, R. E. (1960) "A new approach to linear filtering and prediction problems," *Trans. AMSE - J. Basic Engineering*, Vol. 82 D, pp. 35–45.

Kalnay, E. (2003) *Atmospheric Modelling, Data Assimilation and Predictability*: Cambridge University Press, Cambridge, UK.

Kelly, G. and J-N. Thépaut (2007) "Evaluation of the impact of the space component of the Global Observing System through Observing System Experiments," in *Seminar on recent developments in the use of satellite observations in numerical weather prediction, 3–7 September 2007*, http://old.ecmwf.int/publications/library/do/references/show?id=88345.

Ledoit, O. and M. Wolf (2004) "A well conditioned estimate for large-diemsional covariance matrices," *J. Multivariate Analysis*, Vol. 88, pp. 365–411.

——— (2012) "Nonlinear shrinkage estimation of large-diemsional covariance matrices," *Annals of Statistics*, Vol. 40, pp. 1024–1060.

Levenberg, K. (1944) "A method for the solution of certain non-linear problems in least squares," *Quarterly of Applied Mathematics*, Vol. 2, pp. 164–168.

Liu, Xu, Daniel K. Zhou, Allen Larar, William L. Smith, and Stephen A. Mango (2007) "Case-study of a principal-component-based radiative transfer forward model and retrieval algorithm using EAQUATE data," *Q. J. R. Meteorol. Soc.*, Vol. 133, pp. 243–256, DOI: http://dx.doi.org/10.1002/qj.156.

Lorenc, A. C. and R. T. Marriott (2013) "Forecast sensitivity to observations in the Met Office Global numerical weather prediction system," *Q. J. R. Meteorol. Soc.*, Vol. 140, pp. 209–224, DOI: http://dx.doi.org/10.1002/qj.2122.

Lorenc, A.C. (2003) "Modelling of error covariances by 4D-Var data assimilation," *Q. J. R. Meteorol. Soc.*, Vol. 129, pp. 3167–3182.

Lorenc, A.C., S.P. Ballard, R.S. Bell, N.B. Ingleby, P.L.F. Andrews, D.M. Barker, J.R. Bray, A.M. Clayton, T. Dalby, D. Li, T.J. Payne, and F.W. Saunders (2000) "The Met. Office Global 3-Dimensional Variational Data Analysis Scheme," *Q. J. R. Meteorol. Soc.*, Vol. 125, pp. 2991–3012.

Lorenc, A.C. and T.J. Payne (2007) "4D-Var and the butterfly effect: Statistical four-dimensional data assimilation for a wide range of scales," *Q. J. R. Meteorol. Soc.*, Vol. 133, pp. 607–614, DOI: http://dx.doi.org/:10.1002/qj.36.

Martinet, P., N. Fourrié, V. Guidard, F. Rabier, T. Montmerle, and P. Brunel (2013) "Towards the use of microphysical variables for the assimilation of cloud-affected infrared radiances," *Q. J. R. Meteorol. Soc.*, Vol. 139, pp. 1402–1416, DOI: http://dx.doi.org/:10.1002/qj.2046.

Masiello, G., C. Serio, and P. Antonelli (2012) "Inversion for atmospheric thermodynamical parameters of IASI data in the principal components space," *Q. J. R. Meteorol. Soc.*, Vol. 138, pp. 103–117, DOI: http://dx.doi.org/10.1002/qj.909.

Matricardi, M. (2010) "A principal component based version of the RTTOV fast radiative transfer model," *Q. J. R. Meteorol. Soc.*, Vol. 136, pp. 1823–1835, DOI: http://dx.doi.org/10.1002/qj.680.

Matricardi, M. and A.P. McNally (2014) "The direct assimilation of principal components of IASI spectra in the ECMWF 4D-Var," *Q. J. R. Meteorol. Soc.*, Vol. 140, pp. 573–582, DOI: http://dx.doi.org/10.1002/qj.2156.

McNally, A. P. and P. D. Watts (2003) "A cloud detection algorithm for high-spectral-resolution infrared sounders," *Q. J. R. Meteorol. Soc.*, Vol. 129, pp. 3411–3423.

Migliorini, S., A. Geer, M. Matricardi, and S. English (2014) "All-sky assimilation of selected water vapour IASI channels at ECMWF: strategy and initial trials," in *Proceedings of the 19th International TOVS Study Conference, 26 Mar–14 April 2014, Jeju Island, SOuth Korea*, Poster available from http://cimss.ssec.wisc.edu/itwg/itsc/itsc19/program/posters/9p_07_migliorini.pdf.

Parrish, D. F. and J. C. Derber (1992) "The National Meteorological Center's spectral statistical interpolation analysis system," *Mon. Wea. Rev.*, Vol. 120, pp. 1747–1763.

Pavelin, E. G. and B. Candy (2013) "Assimilation of surface-sensitive infrared radiances over land: Estimation of land surface temperature and emissivity," *Q. J. R. Meteorol. Soc.*, Vol. 140, pp. 1198–1208, DOI: http://dx.doi.org/10.1002/qj.2218.

Pavelin, E. G., S. J. English, and J. R. Eyre (2008) "The assimilation of cloud-affected infrared satellite radiances for numerical weather prediction," *Q. J. R. Meteorol. Soc.*, Vol. 134, pp. 737–749, DOI: http://dx.doi.org/10.1002/qj.243.

Payne, T. (2007) "VAR Scientific Documentation Paper 5: Minimisation and Convergence," Unpublished Met Office Documentation.

Piccolo, C. (2011) "Growth of forecast errors from covariances modeled by 4DVar amd ETKF methods," *Mon. Weather Rev.*, Vol. 139, pp. 1505–1518.

Prunet, P., J.-N. Thépaut, and V. Cassé (1998) "The information content of clear sky IASI radiances and their potential for numerical weather prediction," *Q. J. R. Meteorol. Soc.*, Vol. 124, pp. 211–241, DOI: http://dx.doi.org/10.1002/qj.49712454510.

Rabier, F. (2005) "Overview of global data assimilation developments in numerical weather-prediction centres," *Q. J. R. Meteorol. Soc.*, Vol. 131, pp. 3215–3233, DOI: http://dx.doi.org/10.1256/qj.05.129.

Rodgers, C. D. (2000) *Inverse Methods for Atmospheric Sounding: Theory and Practice*: World Scientific, Singapore.

Saunders, R. W., M. Matricardi, A. Geer, and P. Rayer (2010) *RTTOV-9 Science and Validation Report*, http://research.metoffice.gov.uk/research/interproj/nwpsaf/rtm/rttov9_files/rttov9_svr.pdf.

Siméoni, D., C. Singer, and G. Chalon (1997) "Infrared atmospheric sounding interferometer," *Acta Astronautica*, Vol. 40, pp. 113–118.

Simmons, A. J. and A. Hollingsworth (2002) "Some aspects of the improvement in skill of numerical weather prediction," *Q. J. R. Meteorol. Soc.*, Vol. 128, pp. 647–677, DOI: http://dx.doi.org/10.1256/003590002321042135.

Smith, A., N. Atkinson, W. Bell, and A. Doherty (2015) "An Initial Assessment of Observations from the Suomi-NPP Satellite: Data from the Cross-track Infrared Sounder (CrIS)," *Atmospheric Science Letters*, DOI: http://dx.doi.org/10.1002/asl2.551.

Stengel, M., M. Lindskog, P. Undén, N. Gustafsson, and R. Bennartz (2010) "An extended observation operator in HIRLAM 4D-VAR for the assimilation of cloud-affected satellite radiances," *Q. J. R. Meteorol. Soc.*, Vol. 136, pp. 1064–1074, DOI: http://dx.doi.org/10.1256/qj.621.

Stewart, L. M. (2010) "Correlated observation errors in data assimilation," Ph.D. dissertation, University of Reading, Reading, UK.

Twomey, S. (1996) *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements, 1977.*: Dover Publications, Inc., Mineola, New York.

Ventress, L. and A. Dudhia (2013) "Improving the selection of IASI channels for use in numerical weather prediction," *Q. J. R. Meteorol. Soc.*, Vol. 140, pp. 2111–2118, DOI: http://dx.doi.org/:10.1002/qj.2280.

Wang, C., G. Pan, and L. Cao (2012) "A shrinkage estimation for large dimensional precision matrices using random matrix theory," arXiv:1211.2400.

Watts, P. D. and A. P. McNally (1988) "The sensitivity of a minimum variance retrieval scheme to the value of its principal parameters," in *Proceedings of the 4th International TOVS Study Conference, Igls, Austria*, pp. 399–412.

Weston, P. (2011) "Progress towards the implemetation of correlated observation errors in 4D-Var,"Technical Report 560, Met Office.

Weston, P., E. Pavelin, and A. Collard (2013) *NWPSAF Met Office 1D-Var User Manual*, Available from `http://research.metoffice.gov.uk/research/interproj/nwpsaf/metoffice_1dvar/nwpsaf_mo_ud_006_NWPSAF_1DVar_Manual.html`.

Weston, P.P., W. Bell, and J.R. Eyre (2014) "Accounting for correlated error in the assimilation of high resolution sounders," *Q. J. R. Meteorol. Soc.*, Vol. 140, pp. 2420–2429, DOI: `http://dx.doi.org/:10.1002/qj.2306`.

Xu, Q., L. Wei, and S. Healy (2009) "Measuring information content from observations for data assimilation: connection between different measures and application to radar scan design," *Tellus*, Vol. 61A, pp. 144–153.