EPS Product Validation Report: IASI L1 PCC PPF (Part 2)

Doc.No. : EUM/OPS-EPS/REP/11/0036
Issue : v1A
Date : 16 February 2011

EUMETSAT
Eumetsat-Allee 1, D-64295 Darmstadt, Germany
Tel: +49 6151 807-7
Fax: +49 6151 807 555
http://www.eumetsat.int

Document Change Record

Issue / Revision	Date	DCN No.	Changed Pages / Paragraphs
v1	02 Feb 2011		First Issue
v1A	16 Feb 2011		Minor editorial updates.

Table of Contents

List of Figures

1 INTRODUCTION

1.1 Purpose and Scope

The purpose of this document is to summarise the results of the validation of the Principal Component (PC) compressed IASI Level 1C data. The new eigenvectors, validated in this report, deviate from the previous which are described in [RD14] in two ways:

- 73 outlier spectra collected over fires in Russia, August 2010, have been added to the training spectra.
- The noise-normalised radiance means stored in the field “Mean” of the eigenvector files are no longer based on the mean radiances of the training set of spectra. Instead they are based on the mean L1C radiances from the day of 16 January 2011 processed on GS2, using the “new” spectral database ([RD16]).

Furthermore, the coefficients and threshold for outlier detection have been adjusted and the total number of PC scores in Band 3 has been increased from 80 to 90. These two changes affect the configuration only. As the principal component methodology and the generation of the previous set of eigenvector files is described in [RD14], in this report we will concentrate on describing the changes and the validation of the resulting PCS/PCR products.

1.2 Document Structure

The evolution of the IASI L1C product is discussed in Section 2. The test configuration is specified in Section 3. The changes to the eigenvectors and the configuration are presented in Sections 4 and 5. Section 6 contains the validation results, and finally in Section 7 the conclusion is presented.

1.3 Abbreviations and Acronyms

EARS	EUMETSAT Advanced Retransmission Service
EPS	EUMETSAT Polar System
GEADR	Global External Auxiliary Data Record
GIADR	Global Internal Auxiliary Data Record
GPFS	Generic Product Format Specification
GS	Ground Segment
HDF	Hierarchical Data Format
IASI	Infrared Atmospheric Sounding Interferometer
MDR	Measurement Data Record
Noise	Random measurement error
PCC	Principal Component Compression
PCR	Principal Component Residuals
PCS	Principal Component Scores
PFS	Product Format Specification
PPF	Product Processing Facility

PVRB Product Validation Review Board
RMS Root Mean Square
SAA South Atlantic Anomaly

1.4 Reference Documents

Ref.	Document Title and Reference Number
[RD1]	IASI Level 1 PCC Product Generation Specification, EUM/OPS-EPS/SPE/08/0199
[RD2]	Future Dissemination Approach for IASI level 1 Products, EUM/STG-OPSWG/23/08/DOC/11
[RD3]	Operational Dissemination of IASI Data using PC Compression, EUM/OPS-EPS/TEN/08/0202
[RD4]	Generation of eigenvector files for the IPCC PPF, EUM/OPS-EPS/SPE/08/0200
[RD5]	EPS Programme Auxiliary Data Inventory, EUM/EPS/SYS/LIS/00/002
[RD6]	EARS Operational Service Specification, EUM/OPS/SPE/01/0839
[RD7]	Product Validation Review Board for IASI L1 Day-2 Product, EUM/OPS-EPS/MIN/10/0103
[RD8]	Introduction of IASI L1 Data reduction in EPS, EUM/OPS-EPS/TEN/08/0185
[RD9]	Engineering Change Proposal EPS_AB_ECP_319, in EUMETSAT Dimensions data base.
[RD10]	OPS(L1) Change Request, CNES RF_001 (Hummingbird #300724)
[RD11]	N. C. Atkinson, F. I. Hilton, S. M. Illingworth, J. R. Eyre, and T. Hultberg: “Potential for the use of reconstructed IASI radiances in the detection of atmospheric trace gases”, AMT, 3, 991-1003, 2010.
[RD12]	Technical Note. Noise Covariance Matrix. IA-TN-0000-3271-CNE
[RD13]	Compression des Données IASI par Composantes Principales - Analyse d'Impact sur les Gaz Trace, NOV-3788-NT-9112
[RD14]	EPS Product Validation Report: IASI L1 PCC PPF. EUM/OPS-EPS/REP/10/0148
[RD15]	Product Validation Review Board for IASI Principal Component Compressed products demonstration status. EUM/OPS-EPS/MIN/10/0239
[RD16]	Validation of the new spectral database used by IASI L1 PFF. EUM/OPS-EPS/TBD
[RD17]	N. Bormann, A. Collard and P. Bauer: “Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. : Application to AIRS and IASI data”, Q.J.R. Meteorol. Soc., 136: 1051-1063, April 2010
[RD18]	Tobin et al.: “Principle component analysis of IASI spectra with a focus on non-uniform scene effects on the ILS”

2 ASSUMPTIONS AND OPEN ISSUES

2.1 Evolution of IASI L1C Product

Following the IASI L1C processing configuration change on GS1 on 18 May 2010 (as described in [RD14]) a further configuration change on GS1 was done on 7 February 2011 by activation of the following set of configuration files:

IASI_BRD_xx_M02_20101222080000Z_xxxxxxxxxxxxxxZ_20101222072008Z_IAST_0000000012
IASI_GRD_xx_M02_20101220110000Z_xxxxxxxxxxxxxxZ_20101220101609Z_IAST_0000000022
IASI_ODB_xx_M02_20101222080000Z_xxxxxxxxxxxxxxZ_20101222072116Z_IAST_0000000010

This new configuration aims at improving the pixel differences raised in the NCR-3962 (see [RD16]) and has been active in GS2 since 16 January 2011. The validation in this report has been performed with spectra computed with this configuration.

3 TEST CONFIGURATION

The products under validation have been processed with the IASI L1 PCC Product Processing Facility (PPF) version 1.2.1 configured with the following eigenvector files:

IASI_EV[1|2|3]_xx_M02_20110125000000Z_xxxxxxxxxxxxxxZ_20110119000104Z_xxxx_xxxxxxxxxx

This configuration has been running in GS2 since 1 February 2011 13.08.53 UTC sensing time. The eigenvectors have been changed in two aspects with respect to the previous eigenvectors

IASI_EV[1|2|3]_xx_M02_20100720000000Z_xxxxxxxxxxxxxxZ_20100629000103Z_xxxx_xxxxxxxxxx

which, except for using a different channel number convention (AI-1 in [RD15]), are identical to eigenvectors

IASI_EV[1|2|3]_xx_M02_20100428000000Z_xxxxxxxxxxxxxxZ_20100212000102Z_xxxx_xxxxxxxxxx

which were validated in [RD14] and on which the start of the trial dissemination was based.

4 CHANGES TO THE GENERATION OF EIGENVECTORS

The eigenvectors used for the compression are fully determined by:

1. The noise normalisation matrix.
2. The training set of spectra.
3. The number of eigenvectors to retain.

The choice of the noise normalisation matrix is described in [RD14] and has not been changed for the generation of the set of eigenvectors under validation. We have used the same training set of 101,829 spectra as described in [RD14] plus 73 additional outlier spectra as described in Section 4.1. The number of scores included in the PCS product has been increased from 80 to 90 in Band 3 in recognition of the relatively (compared with Band 1 and 2) high spatial correlation observed for the ‘last’ (i.e. least significant) of the included PC scores. Finally, we have decided not to use the mean radiance of the training set, but rather the mean radiance of the latest configuration L1c spectra, for the computation of the PC scores. This choice is further discussed in Section 4.2.

4.1 Addition of Further Outlier Spectra to the Training Set

Outlier spectra with unusual high residual RMS are systematically being identified and collected in the IASI_IPO auxiliary products. There is evidence that most of the detected outliers correspond to undetected spikes (mainly occurring over the SAA), but some can be attributed to very unusual atmospheric situations not adequately represented in the training set. Such situations are often associated with fires. As Russia was heavily affected by fires in the summer of 2010, we collected the IASI_IPO files from 1 to 10 August 2010 and selected all identified outliers occurring over a geographical area roughly corresponding to Russia; a total of 73 outlier spectra were selected in this way. In the reconstruction residuals computed from these spectra we identify distinct spectral features in the mean and standard deviation, which seem to correlate well with absorption lines of HCN and NH_3.

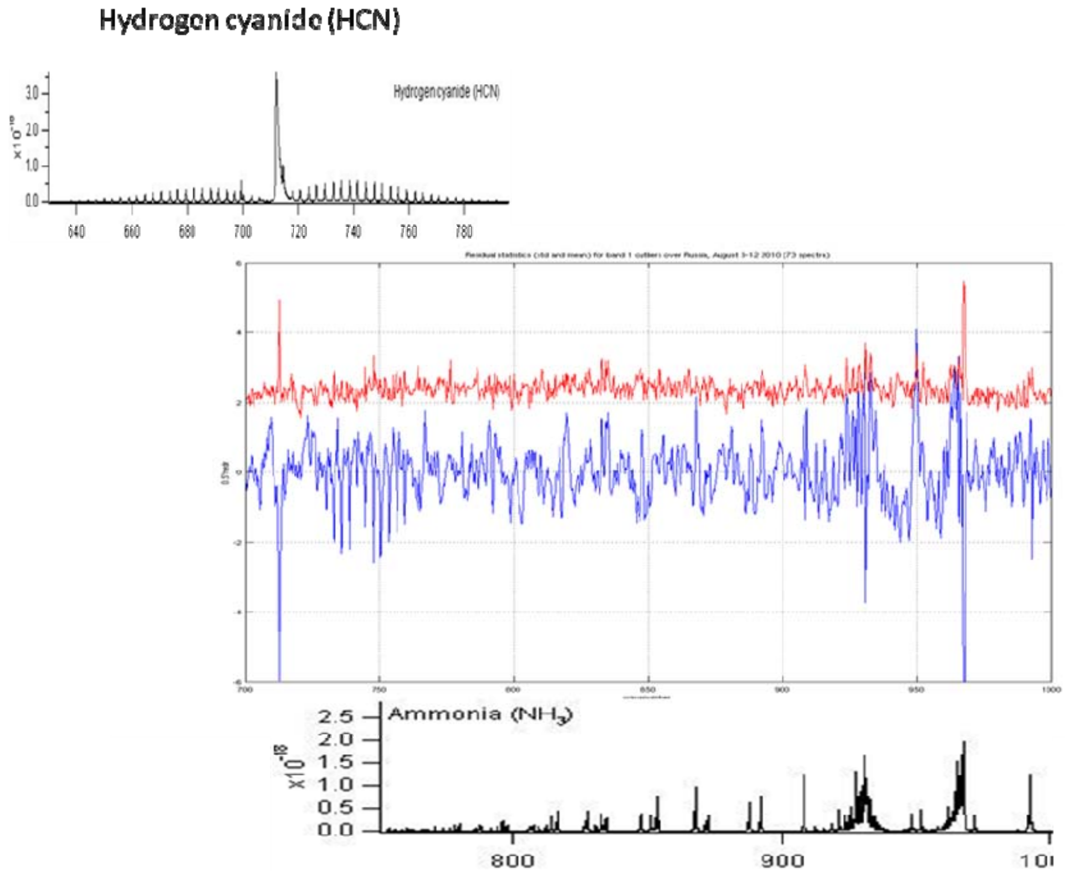


Figure 1: Residual mean and standard deviation for 'Russian fire cases' plotted together with HCN and NH_3 absorption lines, for wave numbers 700-1000 cm^{-1}

Although it has been shown that the NH_3 signal is in general very well represented in the eigenvectors ([RD11]), this indicates that in some extreme situations, the HCN and NH_3 signals could not be fully captured using the previous set of eigenvectors. After adding the selected outliers to the training set, the spectral features observed in the reconstruction residuals were significantly attenuated, but not entirely eliminated. This suggests that the 73 added spectra are not enough to get these type of spectra fully represented by the eigenvectors, and we will continue looking for outlier spectra which can improve the ability of the eigenvectors to represent extreme situations further.

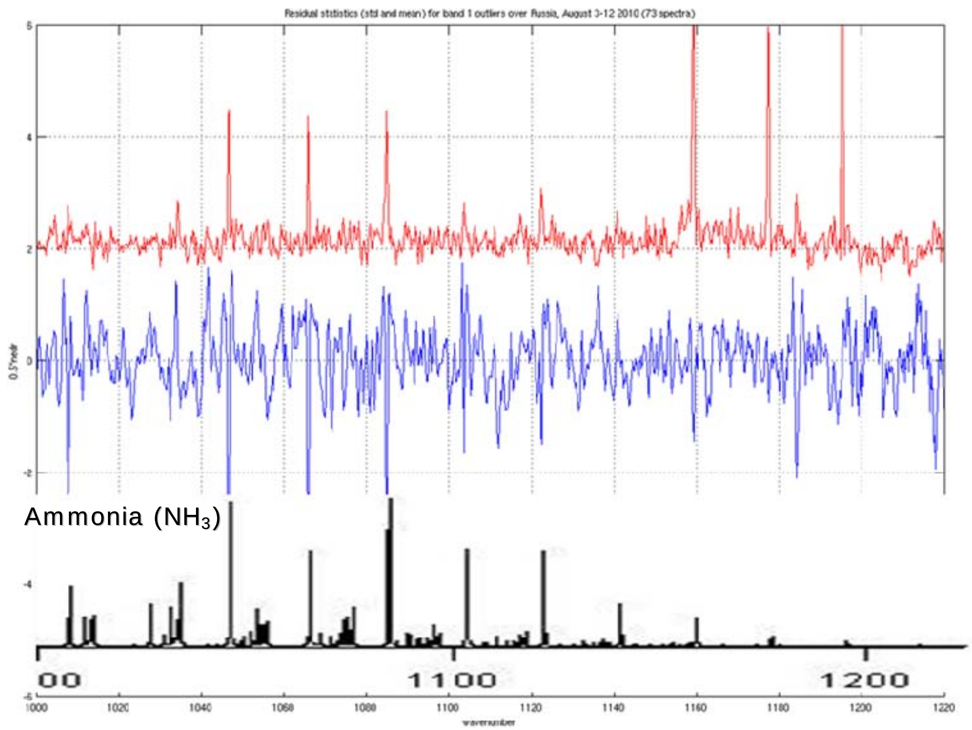


Figure 2: Residual mean and standard deviation for 'Russian fire cases' plotted together with NH_3 absorption lines, for wave numbers 1000-1220 cm^{-1}

4.2 Adaptation of Mean Radiances in Eigenvector Files to the Latest IASI L1C Configuration

As described in the first validation report, the IASI L1C processing configuration has undergone some changes with the result that there was a mismatch between the configuration used for most of the spectra in the training data set and the most recent spectra. This led to a global bias in the residuals, which although very small (around 0.02 K at 280 K for the most affected channels), was considered an obstacle by the first PVRB for declaring the PC score product operational.

As mentioned in Section 2.1, an additional IASI L1C configuration change is in the pipeline and the resulting spectra has been the target of the eigenvectors being validated. We recall that by selecting a training set consisting only of these new spectra (available from GS2 since 16 January 2011) we would lose spectral information about rare events and would encounter non-negligible reconstruction errors in connection with fires, volcanic eruptions, etc. Furthermore it is not possible to capture the effect of the configuration change in a few eigenvectors computed from the residuals since the configuration change 'signal' is lower than the noise level. To get rid of the bias in the residuals we have therefore instead changed the mean radiance within the eigenvector files to correspond to the new configuration without changing the eigenvectors themselves. So in a sense the present set of eigenvector files is inconsistent, in that the eigenvectors have been computed from the training set of spectra, but the mean radiance has been computed from a full day's (16 January 2011) worth of spectra computed with the new configuration. By making sure that the mean radiance applied in the compression corresponds to the mean of the actual spectra being compressed, we trivially ensure the absence of residual bias.

5 CHANGES TO THE CONFIGURATION

5.1 Increase of the Number of Scores in Band 3

As we observed (see detailed plots in Section 6) that the spatial correlation of the PC scores of the PCs with least rank was higher in Band 3 than in the other bands, we decided to increase the total number of scores in Band 3 from 80 to 90. Users can still decide to use only the first 80 scores for reconstruction, if desired.

5.2 Thresholds for Outlier Detection

Following the update to the eigenvector files, the slopes and threshold for outlier detection have been recomputed as described in [RD14]. The resulting values are found in Appendix A. It is worth noting that the thresholds have been set to the mean 'signal-corrected' residual RMS for each detector plus five times its standard deviation. Assuming a perfect Gaussian distribution of the residual RMS and the absence of 'real' outliers, statistically we should expect about one case per day to be classified as an outlier.

6 VALIDATION RESULTS

The validation is based on one full day (2 February 2011) of IASI PCC products processed on GS2 taking 'new' configuration IASI L1C spectra as input. The core validation consists of an examination of the residuals (Section 6.1), but we also take a close look at the PC scores and the outliers in the following sections.

6.1 Examination of the Residuals

The mean and standard deviation of the noise-normalised residuals, computed over a full day (2 February 2011), are shown in Figure 3. The figure looks as expected: the mean is close to zero and the standard deviation is close to one. A few downward spikes can be noticed in the standard deviation; those channels where the standard deviation of the residuals is clearly lower than the noise are the ones where the noise filtering is less efficient and a relatively high proportion of the original noise is still contained in the reconstructed radiances.

When the mean radiance is computed separately for each of the four pixels (Figure 4) the bias remains low but gets slightly higher. This is not surprising considering the differences in mean radiance of the original spectra between the four detectors.

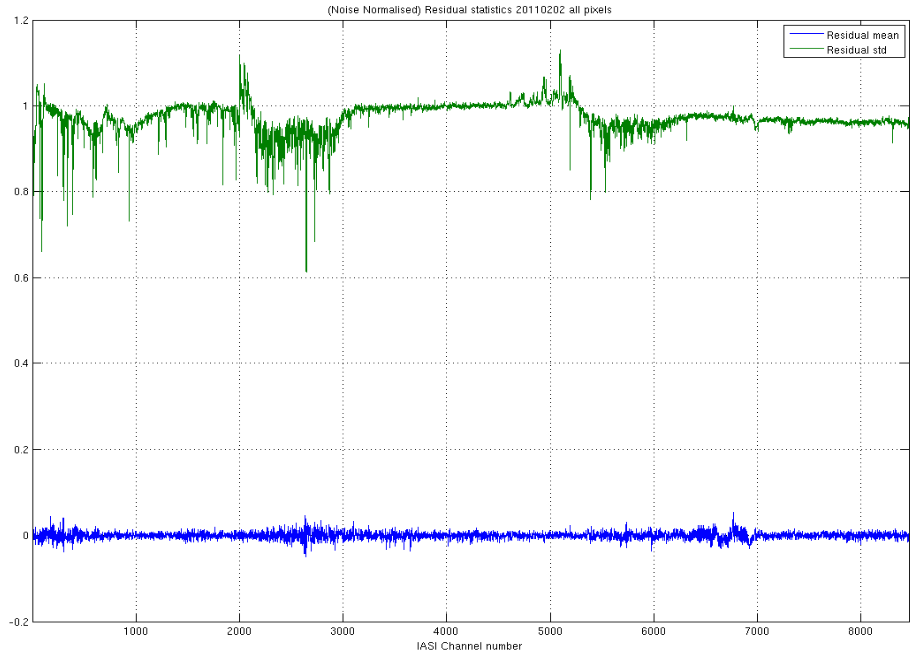


Figure 3: Residual mean and standard deviation

While the residual mean is similar among the four pixels, there are clear differences between the residual standard deviations for each of the four detectors, as can be seen in Figure 5.

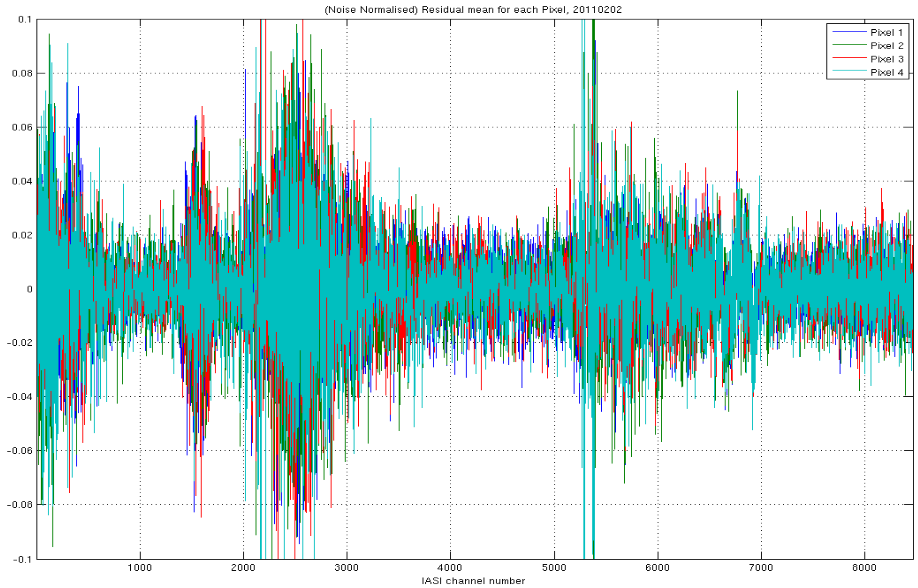


Figure 4: Residual mean for each detector

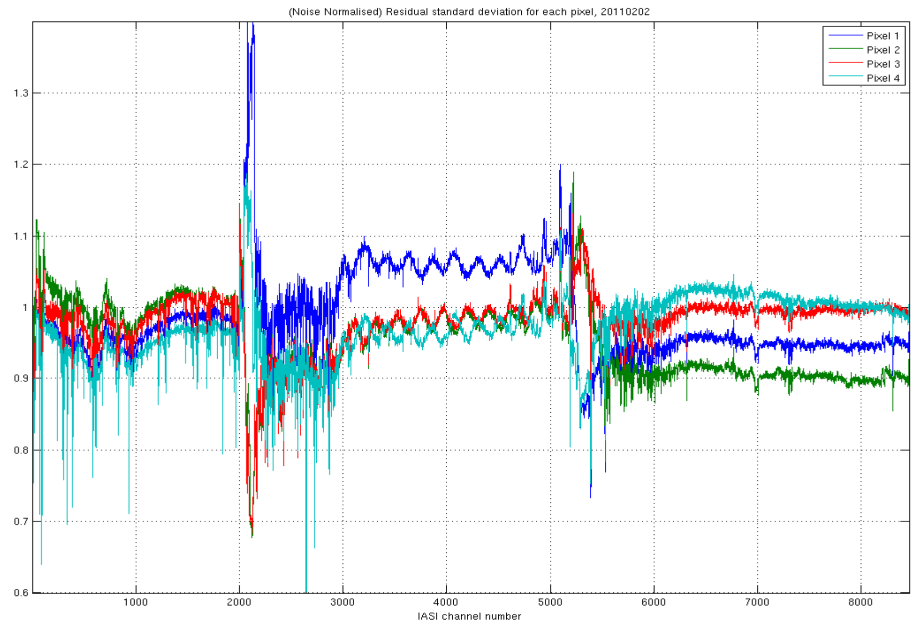


Figure 5: Residual standard deviation for each detector

The differences are biggest in the band overlap regions. Outside the band overlap regions, the varying noise properties from pixel to pixel can be observed. The 'out of sync' wave shapes seen for channel numbers 3000 to 5000 look like they could be related to a L1C processing artefact, but have not been satisfactorily explained.

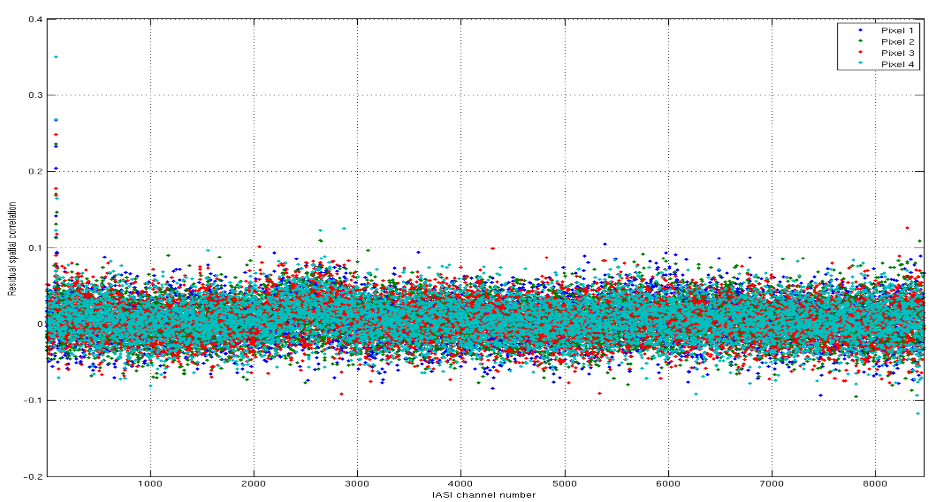


Figure 6: Residual spatial correlation

As we expect the atmospheric signal to be well represented by the truncated set of PC scores, and therefore expect to find only a negligible amount of 'signal' left in the residuals, we don't expect to observe any spatial correlation of the residuals. To check this we computed and plotted (Figure 6) the residual spatial correlation within each IASI channel separately for each of the detectors. We see that this is indeed the case, except for the channels 90 to 95.

6.2 Examination of the Scores

In this section we look at the PC scores. We have computed various statistical properties of the individual scores and present them in three plots (one for each band). The plots show the PC score spatial correlation as well as normalised standard deviation and mean as a function of PC rank. The four upper curves (with values around 1) show the PC score standard deviation for each of the four pixels divided by the average of the standard deviation over all four pixels. Each colour corresponds to a different pixel as follows: Pixel 1, blue; Pixel 2, green; Pixel 3, red; Pixel 4, cyan. The four lower curves (with values around 0) show the mean of the PC score for each of the four pixels minus the average mean over all four pixels divided by the average of the standard deviation over all four pixels. The four coloured curves between one and zero show the spatial correlation of the PC score for each of the four pixels and the black curve shows the inter-EFOV spatial correlation, i.e. the correlation computed from all six pairs of IFOVs within each EFOV.

From these plots we can identify certain eigenvectors which carry most of the radiance inter-pixel differences. We also see that for a few eigenvectors, the spatial correlation is remarkably low (even negative in one case) for some or all of the pixels; this could be a sign that the information carried by these eigenvectors mostly arises from undesired instrument or processing effects.

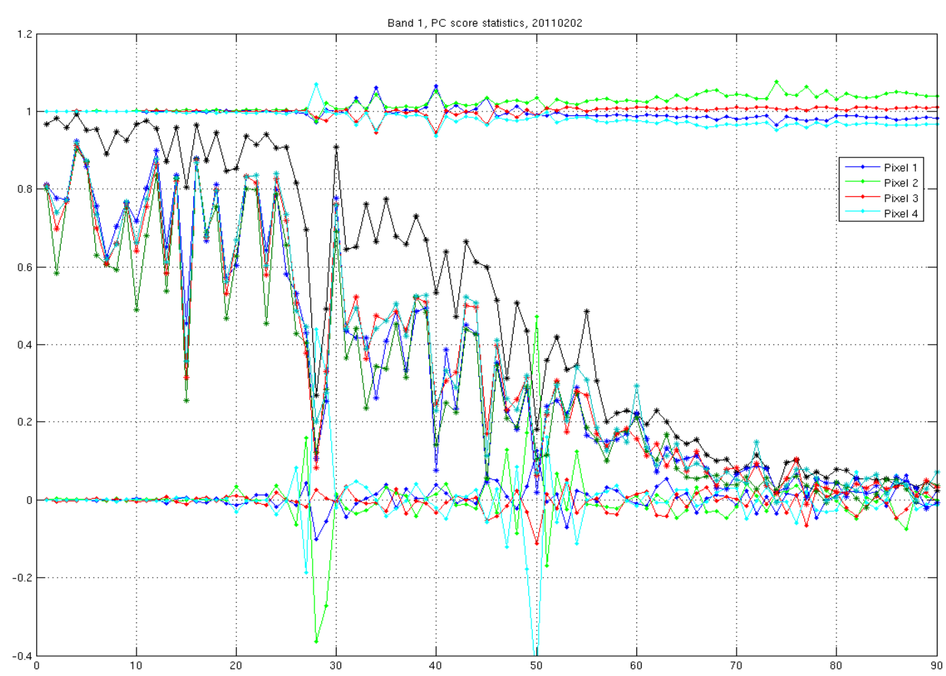


Figure 7: Band 1 PC score statistics

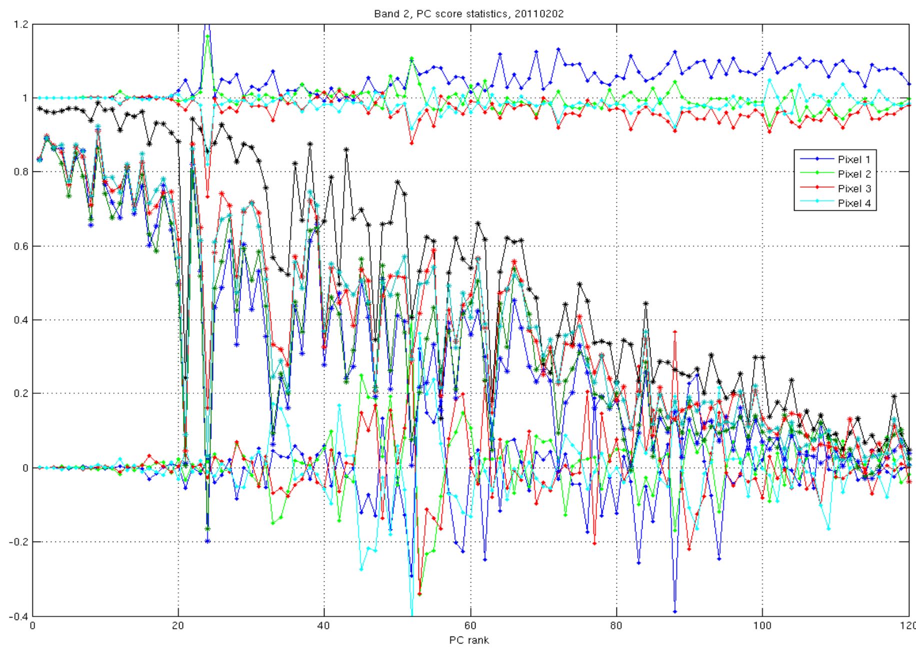


Figure 8: Band 2 PC score statistics

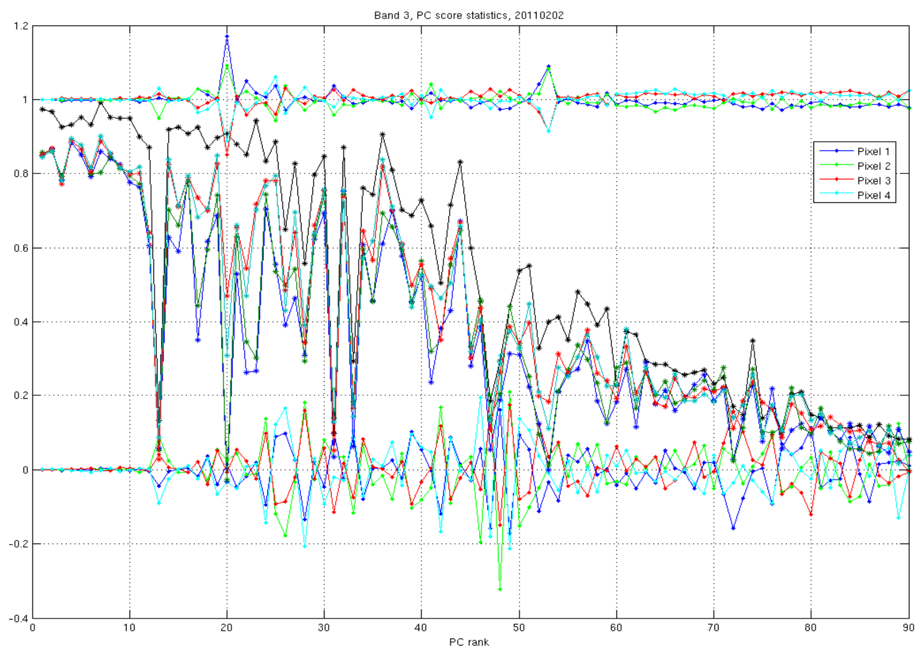
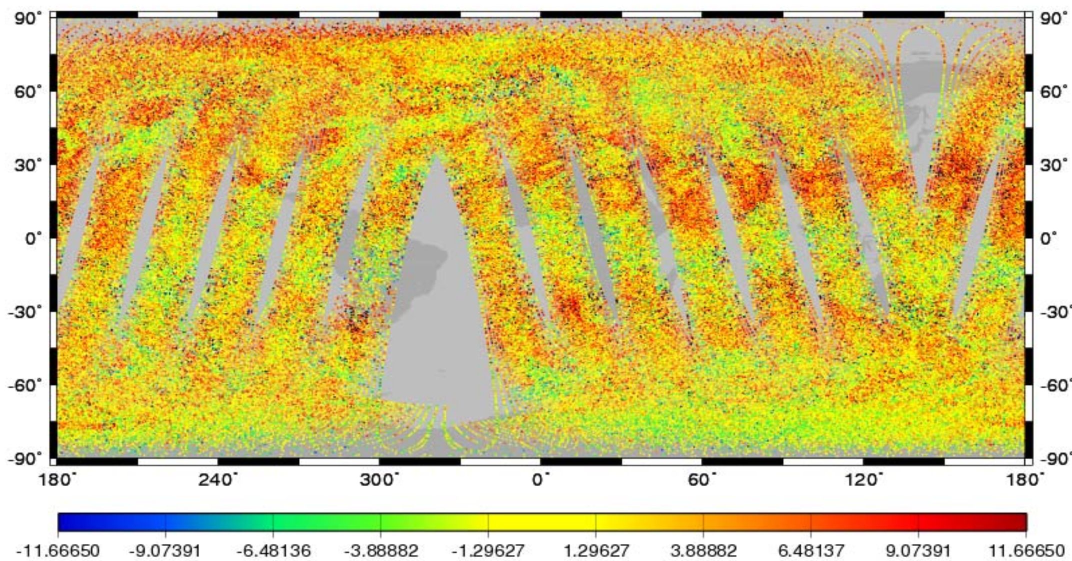


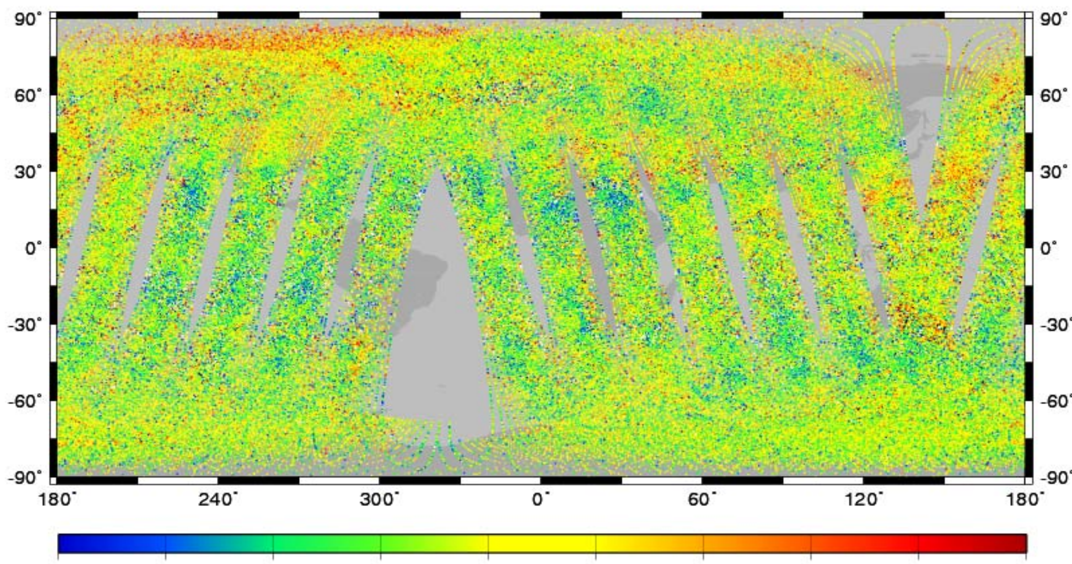
Figure 9: Band 3 PC score statistics

In the following we show a number of plots of some interesting PC scores separated per pixel. Only the last 12 hours of 2 February 2011 have been plotted in order to show ascending and descending orbits side by side in a single plot. The colour scale for a particular score is the same for each pixel and has been chosen to cover the range +/–2.7 times the overall standard deviation.

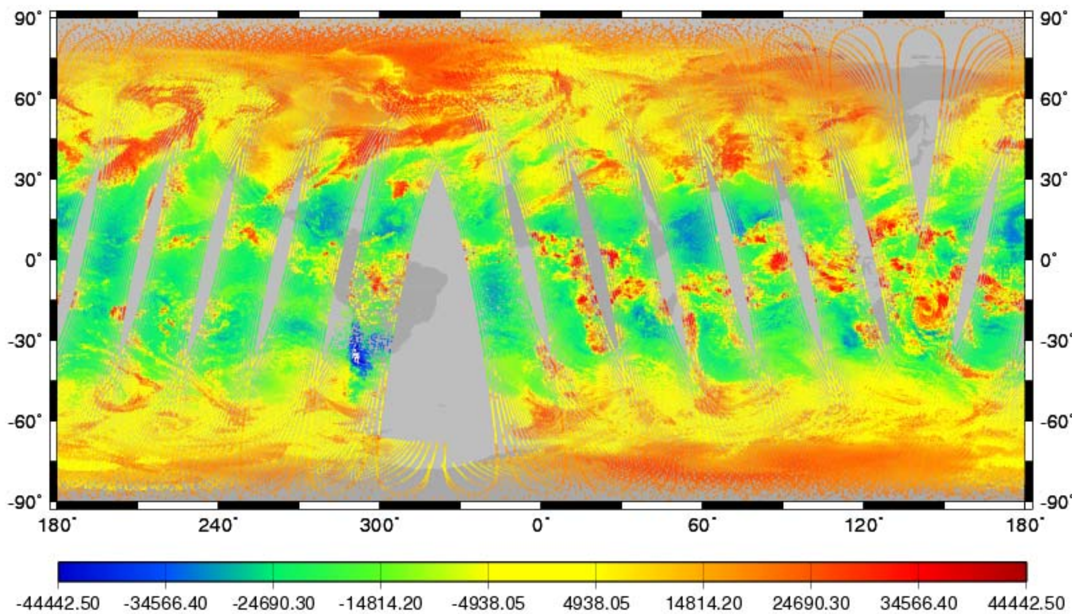
Band 1, PC score 50, Pixel 2 (20110202 12-24)



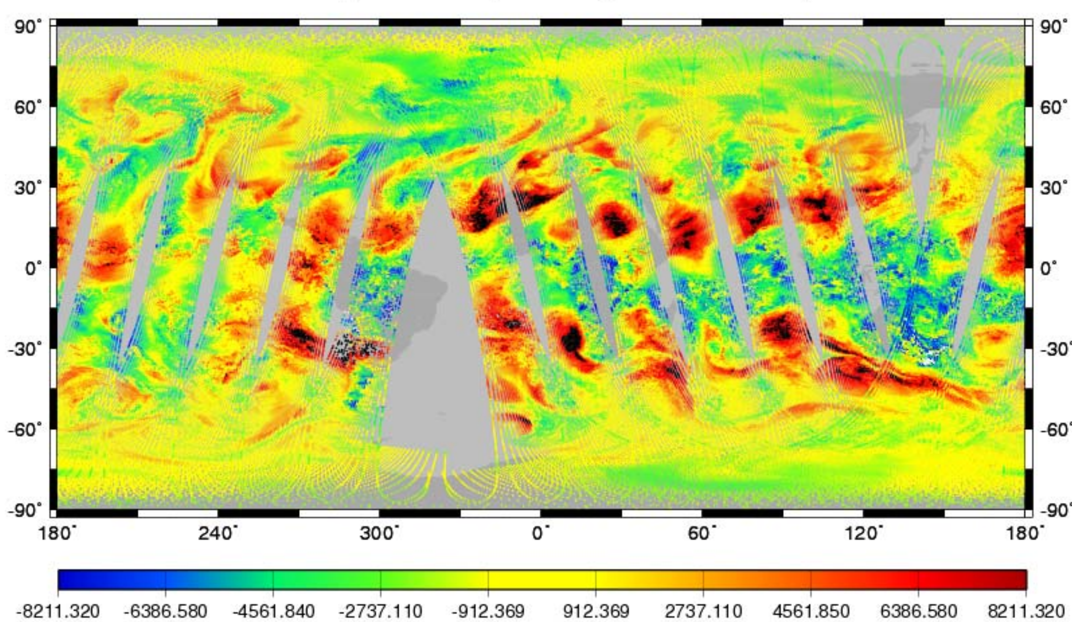
Band 1, PC score 50, Pixel 4 (20110202 12-24)



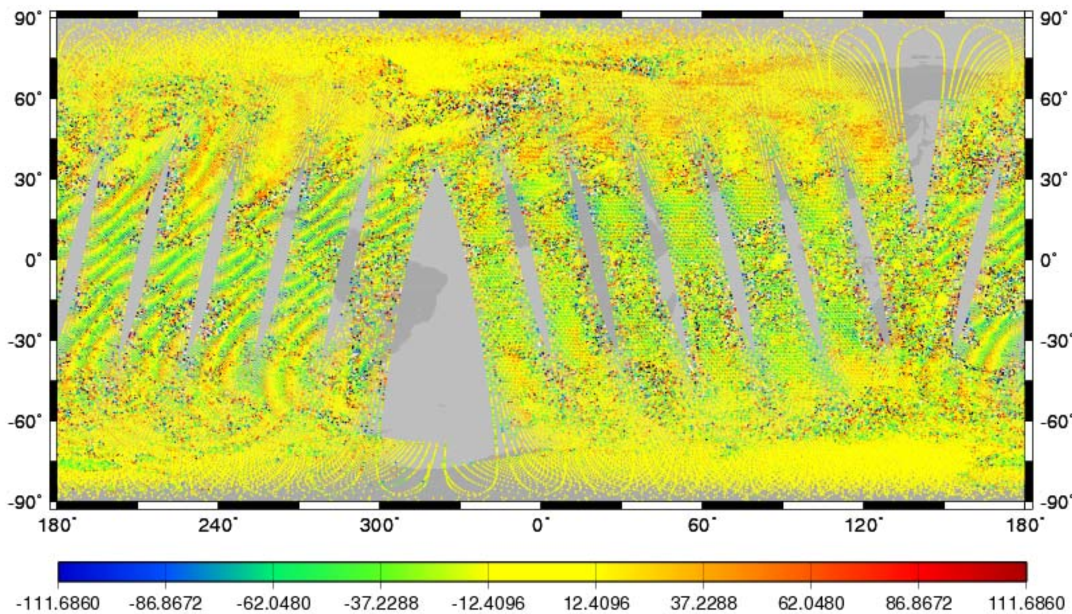
Band 2, PC score 1, Pixel 1 (20110202 12-24)



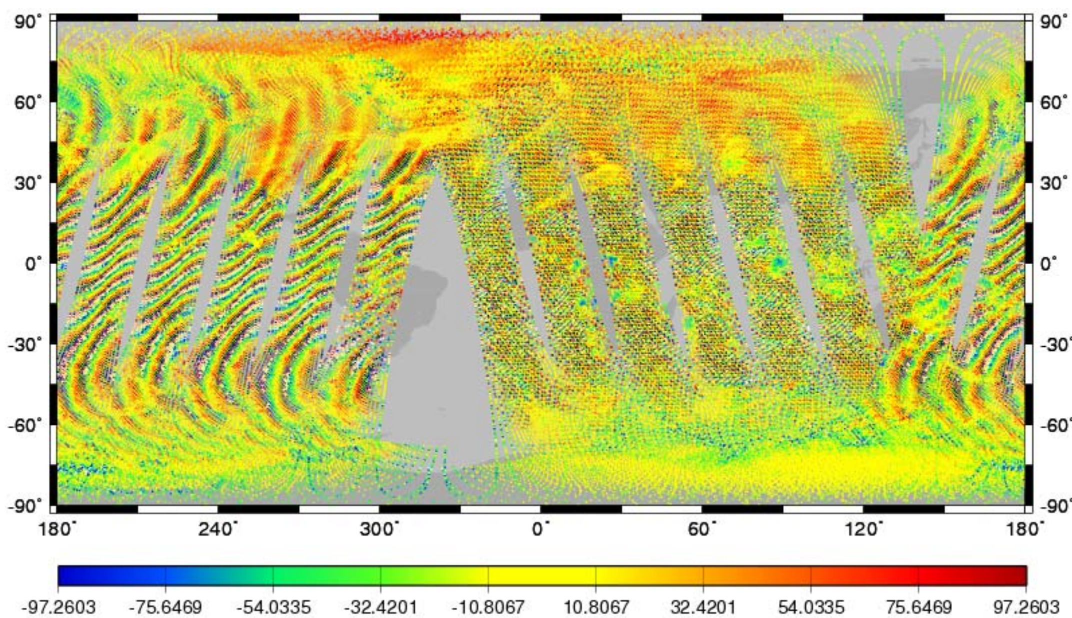
Band 2, PC score 2, Pixel 1 (20110202 12-24)

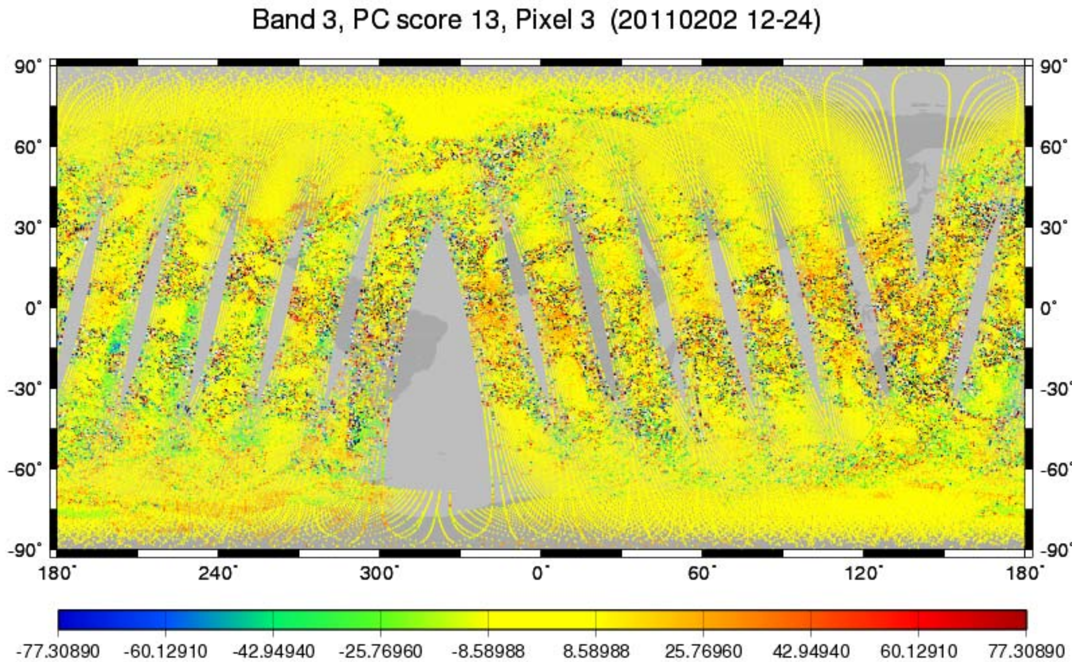


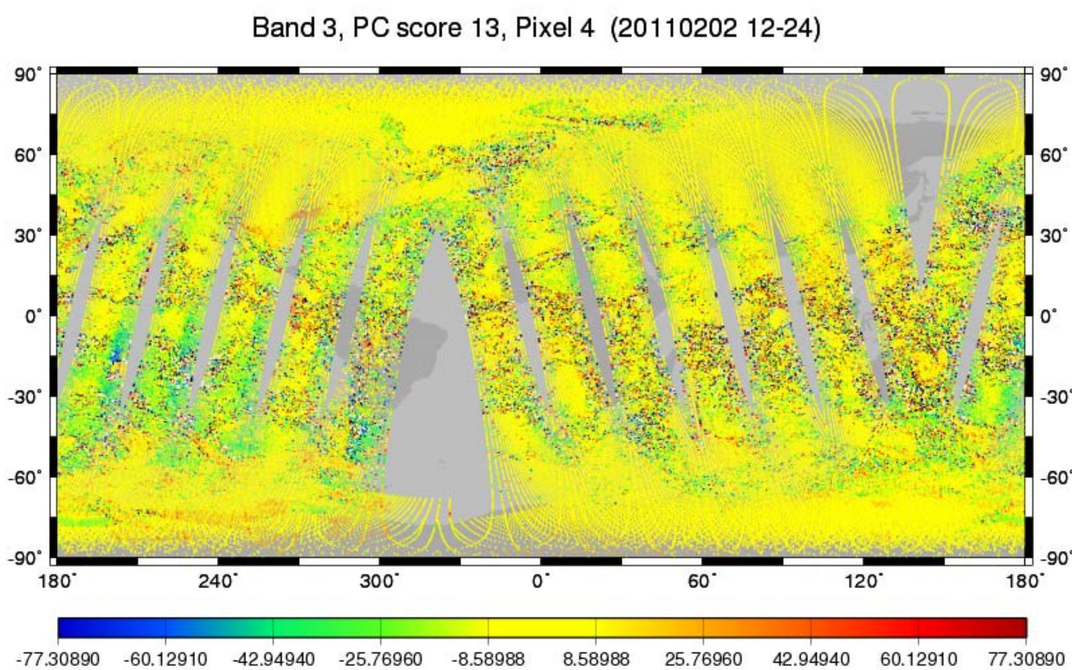
Band 2, PC score 21, Pixel 1 (20110202 12-24)



Band 2, PC score 24, Pixel 1 (20110202 12-24)




Band 3, PC score 13, Pixel 3 (20110202 12-24)
90°
60°
30°
0°
-30°
-60°
-90°
180°
240°
300°
0°
60°
120°
180°
-77.30890
-60.12910
-42.94940
-25.76960
-8.58988
8.58988
25.76960
42.94940
60.12910
77.30890


Band 3, PC score 13, Pixel 4 (20110202 12-24)
90°
60°
30°
0°
-30°
-60°
-90°
180°
240°
300°
0°
60°
120°
180°
-77.30890
-60.12910
-42.94940
-25.76960
-8.58988
8.58988
25.76960
42.94940
60.12910
77.30890

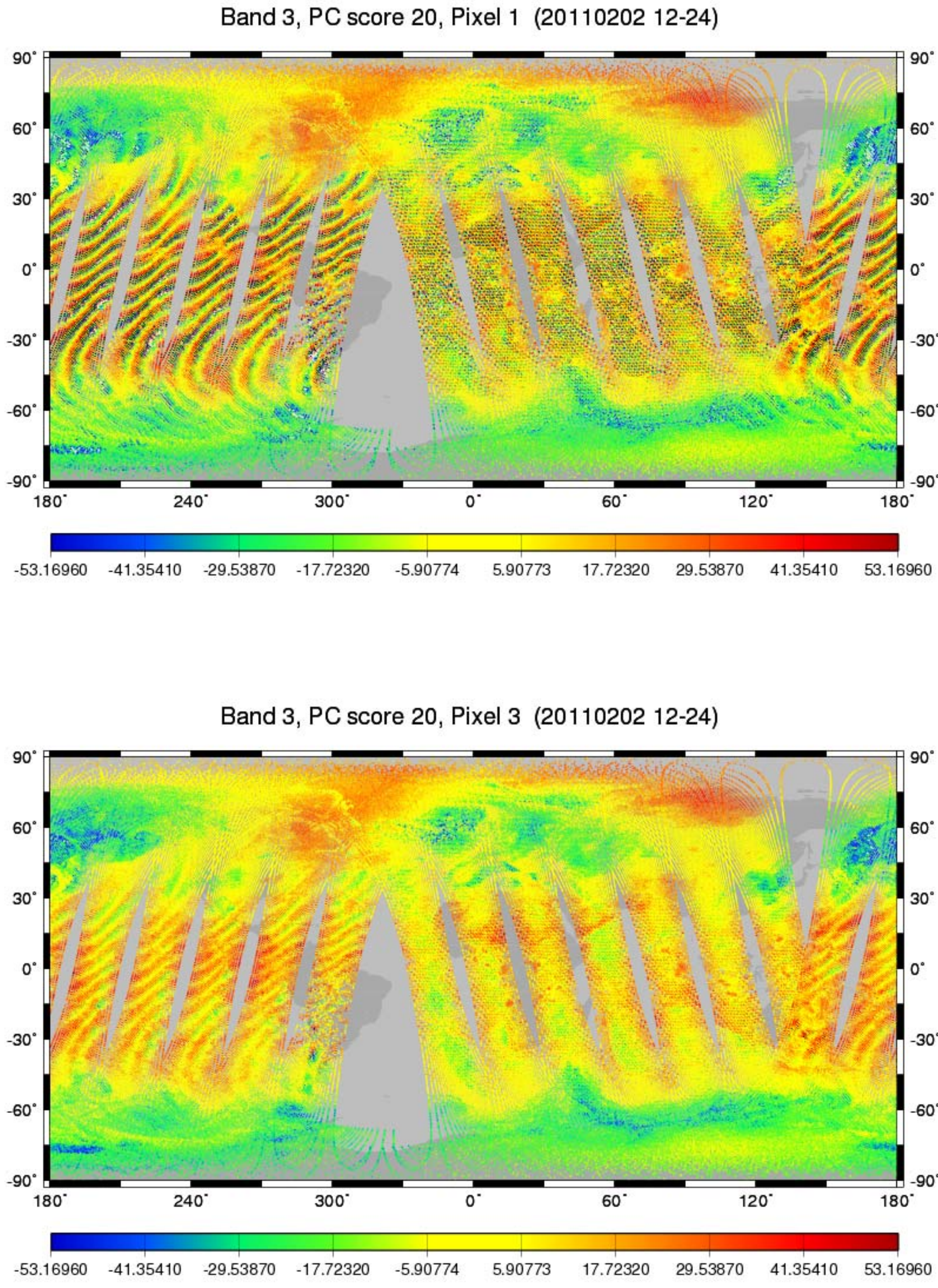


Figure 10: PC scores for selected bands, ranks and pixels

An interesting pattern is observed for PC score #20 in Band 3 (and others). At daytime a beautiful zebra-stripe pattern appears, whereas the night-time pattern looks more like a snakeskin. It is likely that these eigenvectors are responsible for the chessboard patterns in the OBS-CALC spatial covariances observed by N. Bormann et al. ([RD17]), which were tentatively attributed to micro-vibrations of IASI's beam splitter. The fact that the pattern is very different at day and at night indicates that stray light might also play a role.

It is worth recalling ([RD14]) that these patterns can also be observed directly, with an amplitude of at least 5 K, in some of the IASI Band 3 channels after noise filtering (without noise filtering the pattern is hidden by the noise).

We have seen that some of the PC scores are characterised by very small spatial correlation and at the same time exhibit significantly different statistics (mean and/or standard deviation) for each of the four instrument detectors. It is tempting to try excluding such scores from the radiance reconstruction in order to harmonise the spectra measured by each of the four detectors, and to suppress spectral features which do not correspond to atmospheric signal, as for example the non-uniform scene ILS effects ([RD18]). (The non-uniform scene effect shows up nicely in Band 2 score 21 and Band 3 score 13.) A full investigation of such an approach is however beyond the scope of this validation report.

6.3 Examination of the Outliers

Of the 1,296,000 spectra from 2 February 2011, a total of 400 were classified as outliers (only spectra flagged with good quality are eligible). Of these, 114 were Band 2 outliers and 286 Band 3 outliers – no Band 1 outliers were encountered. Their locations are shown in the map below (Band 2 in red and Band 3 in black).

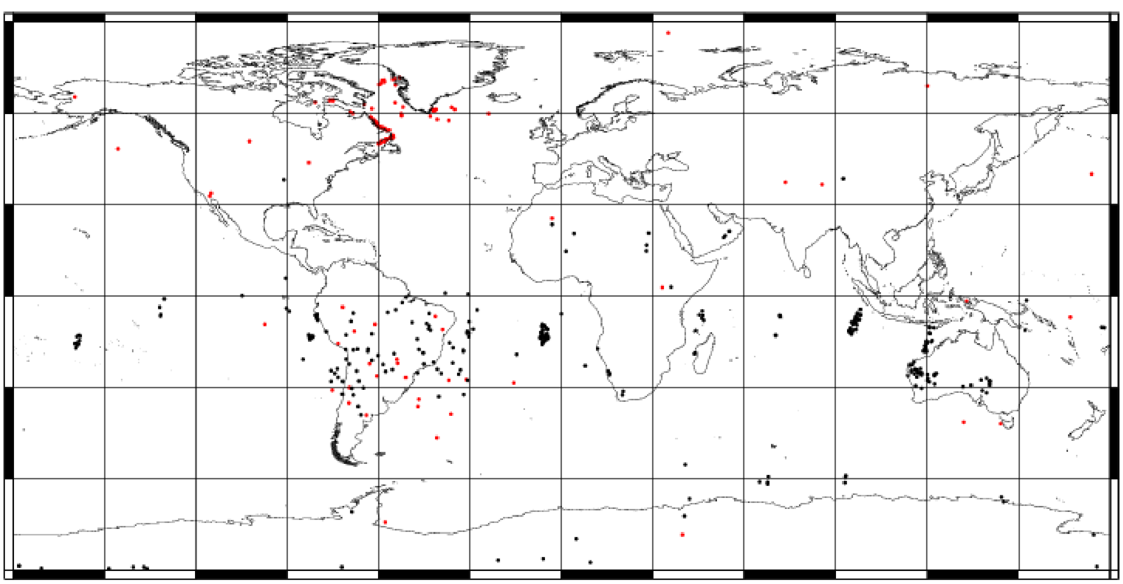


Figure 11: Locations of outliers for 2 February 2011

It can be seen that outliers (both Band 2 and 3) occur over the SAA and can likely be attributed to undetected spikes. The clusters of Band 3 outliers a little south of the equator coincide with regions of sun glint. No particular spectral features can be observed in the residuals for these two groups of outliers. But the outliers at the coast of Newfoundland and in the ocean around southern Greenland are hard to explain. They were isolated by choosing all Band 2 outliers with longitude between 90°W and 30°W and latitude higher than 30°N, resulting in 72 cases, which were analysed separately. The plot below shows the mean and standard deviation of the residuals for these cases within the IASI channels of Band 2. It is believed that this structure is caused by spectral shifts resulting from spatially inhomogeneous scenes and that it would not be beneficial to add these spectra to the training set. This should however be further investigated.

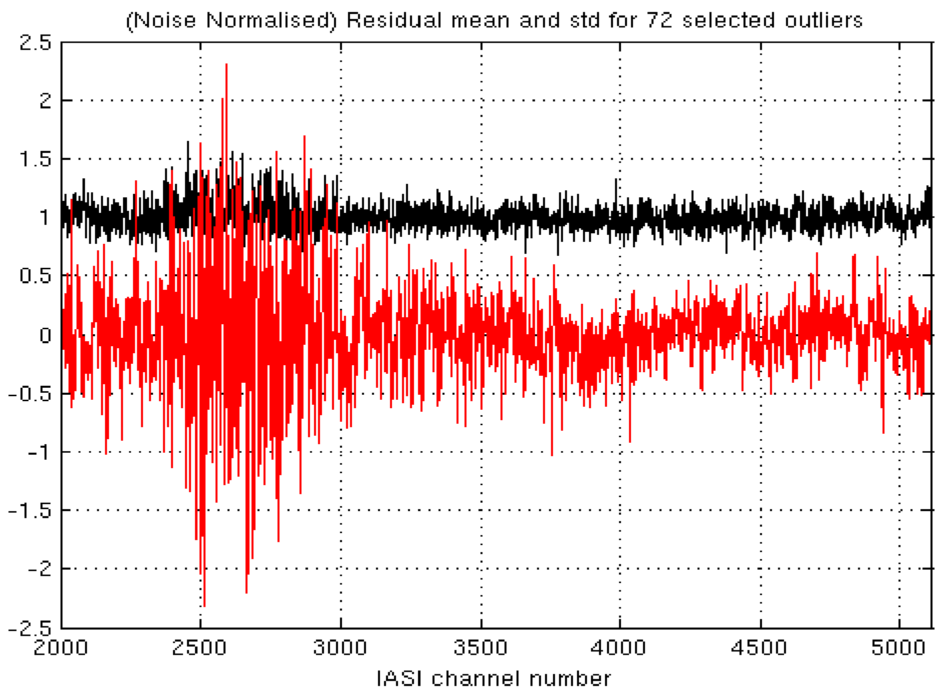


Figure 12: (Noise normalised) residual mean and standard deviation for 72 selected outliers

7 CONCLUSION AND RECOMMENDATION

The eigenvector files have been updated to remove the small bias in the residuals computed with the previous eigenvectors, as required by the previous PVRB ([RD15]), in order to reach operational status. It is therefore recommended to:

1. Inform users of the planned date and time of the change of eigenvectors.
2. Make the eigenvector files IASI_EV[1|2|3]_xx_M02_20110125000000Z_xxxxxxxxxxxxxxZ_20110119000104Z_xxxx_xxxxxxxxxx available to users and ingest them in GS1.
3. Install and activate the IASI PCC PPF V1.2.1 on GS1.
4. Declare the IASI Level 1 PCS products operational.

On a longer timescale it is furthermore recommended to:

1. Further investigate outlier spectra and continue to seek improving the representation of rare situations in the training set.
2. Examine the validity and shortcomings of tailor-made radiance reconstruction to suppress undesired L1C effects (as discussed in Section 6.2), including an investigation of its impact on retrievals.
3. Investigate the benefits of improving the noise normalisation by using the matrix square root of the noise covariance matrix as the noise normalisation matrix (this is essentially equivalent to de-apodisation followed by diagonal noise normalisation). This would make the tail of the eigenvalue plot flatter (N. Atkinson, personal communication) and is expected to improve the ability to distinguish signal from noise (i.e. to get less noise in the reconstructed radiances and get (even) less signal in the residuals).

APPENDIX A CONTENTS OF CONFIGURATION FILE (IPCC_PPF.CONF)


```
<IpccPpfConfig>
  <McsConfig>
    <FacilityId>PC</FacilityId>
    <SendStatusInterval>15</SendStatusInterval>
  </McsConfig>

  <Processing>
    <nbrScoresB1P1>   1  </nbrScoresB1P1>
    <nbrScoresB1P2>  41  </nbrScoresB1P2>
    <nbrScoresB1P3>  48  </nbrScoresB1P3>

    <nbrScoresB2P1>   2  </nbrScoresB2P1>
    <nbrScoresB2P2>  61  </nbrScoresB2P2>
    <nbrScoresB2P3>  57  </nbrScoresB2P3>

    <nbrScoresB3P1>   1  </nbrScoresB3P1>
    <nbrScoresB3P2>  44  </nbrScoresB3P2>
    <nbrScoresB3P3>  45  </nbrScoresB3P3>

    <outlierThresholdB1D1> 1.0265 </outlierThresholdB1D1>
    <outlierThresholdB1D2> 1.0702 </outlierThresholdB1D2>
    <outlierThresholdB1D3> 1.0521 </outlierThresholdB1D3>
    <outlierThresholdB1D4> 1.0099 </outlierThresholdB1D4>

    <outlierThresholdB2D1> 1.0703 </outlierThresholdB2D1>
    <outlierThresholdB2D2> 0.9721 </outlierThresholdB2D2>
    <outlierThresholdB2D3> 0.9745 </outlierThresholdB2D3>
    <outlierThresholdB2D4> 0.9752 </outlierThresholdB2D4>

    <outlierThresholdB3D1> 0.9335 </outlierThresholdB3D1>
    <outlierThresholdB3D2> 0.9065 </outlierThresholdB3D2>
    <outlierThresholdB3D3> 0.9831 </outlierThresholdB3D3>
    <outlierThresholdB3D4> 0.9868 </outlierThresholdB3D4>

    <outlierSlopeB1> 0.029567 </outlierSlopeB1>
    <outlierSlopeB2> 0.21756 </outlierSlopeB2>
    <outlierSlopeB3> 3.4011 </outlierSlopeB3>

    <scoreQuantisationFactorB1> 0.5   </scoreQuantisationFactorB1>
    <scoreQuantisationFactorB2> 0.5   </scoreQuantisationFactorB2>
    <scoreQuantisationFactorB3> 0.5   </scoreQuantisationFactorB3>

    <residualQuantisationFactorB1> 0.5   </residualQuantisationFactorB1>
    <residualQuantisationFactorB2> 0.5   </residualQuantisationFactorB2>
    <residualQuantisationFactorB3> 0.5   </residualQuantisationFactorB3>

  </Processing>

</IpccPpfConfig>
```