

NWP SAF	IRSPP Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	-------------------------------	---

NWP SAF
IRSPP Top Level Design

Version 1.4

31st January 2022

NWP SAF	IRSP Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	------------------------------	---

IRSP Top Level Design

This documentation was developed within the context of the EUMETSAT Satellite Application Facility on Numerical Weather Prediction (NWP SAF), under the Cooperation Agreement dated 7 December 2016, between EUMETSAT and the Met Office, UK, by one or more partners within the NWP SAF. The partners in the NWP SAF are the Met Office, ECMWF, DWD and Météo France.

COPYRIGHT 2022, EUMETSAT, ALL RIGHTS RESERVED.

Change record			
Version	Date	Author / changed by	Remarks
1.0	09/02/2021	Nigel Atkinson	For user consultation
1.1	09/03/2021	Nigel Atkinson	Incorporate EUMETSAT feedback
1.2	03/06/2021	Nigel Atkinson	Add HT-FRTC schematic
1.3	06/10/2021	Nigel Atkinson	Add section 3.2 and modify 6.7
1.4	31/01/2022	Nigel Atkinson	Incorporate comments from review

NWP SAF	IRSP Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	------------------------------	---

Table of Contents

1.	INTRODUCTION.....	4
1.1	Reference documents.....	4
2.	INSTALLATION.....	4
2.1	Delivery.....	4
2.2	Prerequisites.....	4
2.3	Building IRSP from source.....	5
3.	FUNCTIONALITY OVERVIEW.....	5
3.1	Functionality within IRSP.....	5
3.2	Relationships with RTTOV and assimilation systems.....	7
4.	SOFTWARE TOP-LEVEL DESCRIPTION.....	7
4.1	Fortran code.....	7
4.2	Namelists.....	8
5.	INPUT AND OUTPUT DATA.....	8
6.	PRINCIPAL COMPONENTS - THEORY.....	9
6.1	Converting between radiances and PC scores.....	9
6.2	Global and local PCs.....	9
6.3	Reconstruction scores.....	10
6.4	Creating the eigenvectors.....	10
6.5	Apodisation.....	11
6.6	Basis function transformation.....	12
6.7	Support for HT-FRTC.....	13
7.	PROTOTYPE BUFR SEQUENCE.....	14
8.	REQUIREMENTS TRACEABILITY.....	16
9.	REFERENCES.....	17

NWP SAF	IRSPF Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	-------------------------------	---

1. INTRODUCTION

This document presents the Top Level Design for the Infrared Sounder Pre-processor, IRSPF, a software deliverable of the NWP SAF. Information on the package can be found at <https://nwp-saf.eumetsat.int/site/software/irspf/>. As well as describing the proposed implementation, at high level, this document also presents a theoretical description of Principal Components techniques to be implemented, i.e. it functions as an Algorithm Theoretical Basis Document.

On release of the package, this document will be expanded to a full description, and will be issued as a User Manual.

1.1 Reference documents

- [RD-1] IRSPF Product Specification, NWPSAF-MO-DS-037, available on the NWP SAF web site.
[RD-2] MTG IRS Level 0 & 1 Format Specification [IRSL1FS], EUM/MTG/SPE/10/0449

2. INSTALLATION

2.1 Delivery

IRSPF will normally be delivered via the NWP SAF web site as a gzipped tar file containing source code, to be compiled by the user. The package may also be made available in the form of compiled executables, for a platform to be specified in the User Manual.

Test cases will be delivered separately, also as gzipped tar files.

2.2 Prerequisites

IRSPF is intended to be built on a Linux platform, using a ksh or bash environment. A Fortran90 compiler (e.g. gfortran) is required. IRSPF makes use of the following libraries, which need to be installed on the user's system:

- hdf5 (with Fortran enabled)
- netcdf-c
- netcdf-f
- ecCodes

and the following optional library (for users who want to generate their own eigenvectors):

- LAPACK (Linear Algebra PACKage). Current versions come with BLAS (Basic Linear Algebra Subprograms) included.

Users will also need cmake (for building ecCodes) and zlib. These are commonly included in Linux distributions.

IRSPF will include a script that can download and build the dependency libraries in a local directory, without the need for administrator privilege.

Note that the use of Fortran90 is consistent with NWPSAF-MO-SW-002 "Development Procedures for Software Deliverables". Other languages could have been considered, but Fortran90 (sometimes with Fortran2003 extensions) is still widely used in the meteorological community.

NWP SAF	IRSP Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	------------------------------	---

2.3 Building IRSP from source

There will be a 2-step process:

1. Run a *configure* script (shell script), which will search for the dependency libraries and create an appropriate Makefile.ARCH containing information about the user's system, including the location of libraries to be linked in.
2. Run "make"

Makefiles will be included in the source and each one will reference the top-level Makefile.ARCH. The structure of the makefiles will be relatively simple, so that users can modify them if desired.

The *configure* script will also generate an environment file which can be sourced by the user at run time in order to set up environment variables such as PATH, LD_LIBRARY_PATH and ECCODES_DEFINITION_PATH.

3. FUNCTIONALITY OVERVIEW

3.1 Functionality within IRSP

The IRSP package is designed for processing data from the MTG-IRS sensor, specifically the Principal Component (PC) and Spectral Sounding Sample (SSS) level 1b products. It also incorporates the functionality that was present in an earlier NWP SAF deliverable, the "IASI PCA-based compression package".

Typical workflows for IRSP are shown in Figure 1 to Figure 5. For operational use, it is expected that only the first of these would be used. However, the ability for the users to generate, and use, their own eigenvectors is also provided, for consistency with the IASI PCA package.

Please note that IRSP v1 does not attempt to make any improvements to the IRS data quality: it is assumed that all necessary corrections have been made in the level 1 processing that is run at EUMETSAT, i.e. all pixels use the same spectral grid and have been harmonised so that their spectral response functions are the same (Coppens et al., 2019). If, after launch, specific corrections are found to be necessary for NWP assimilation, these will be considered for a future software release.

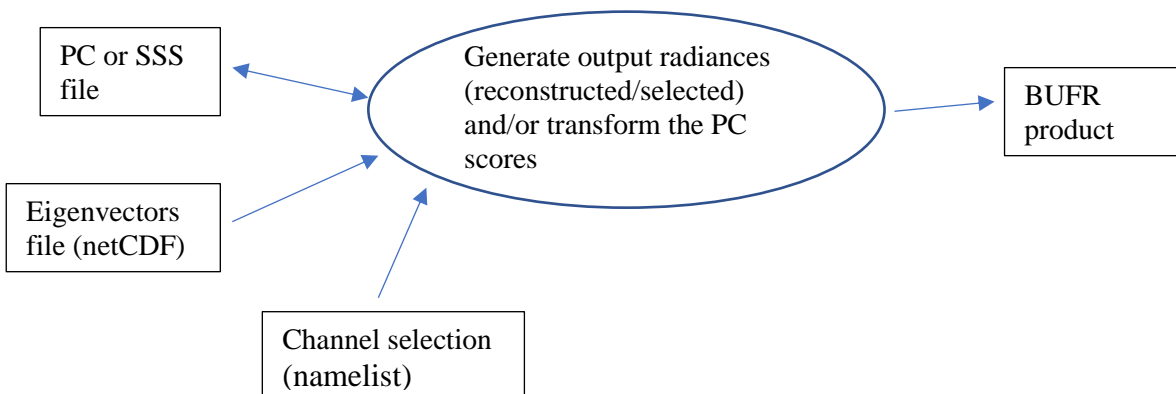


Figure 1: Workflow for the core task of processing IRS data for use in NWP

NWP SAF	IRSPP Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	-------------------------------	---

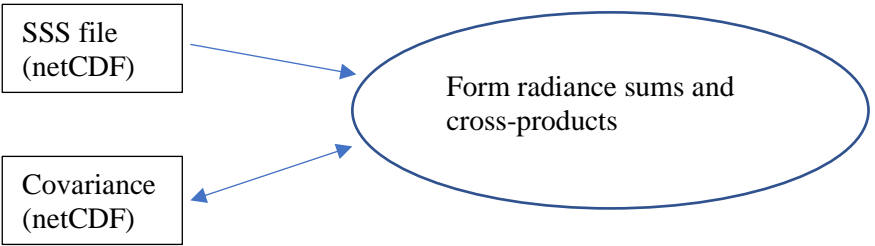


Figure 2: Workflow for generating covariance matrix from many SSS files

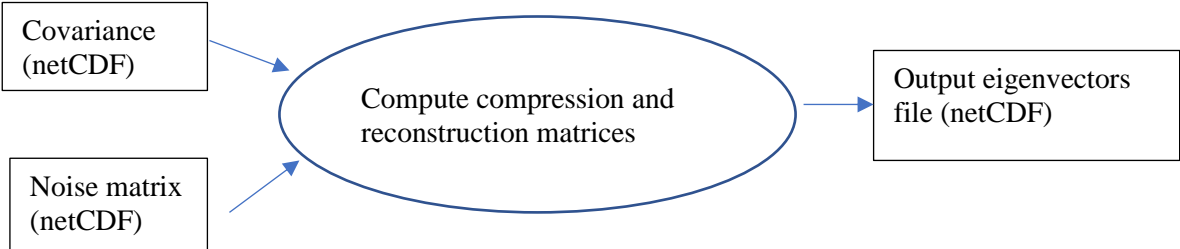


Figure 3: Workflow for generating user-defined eigenvectors

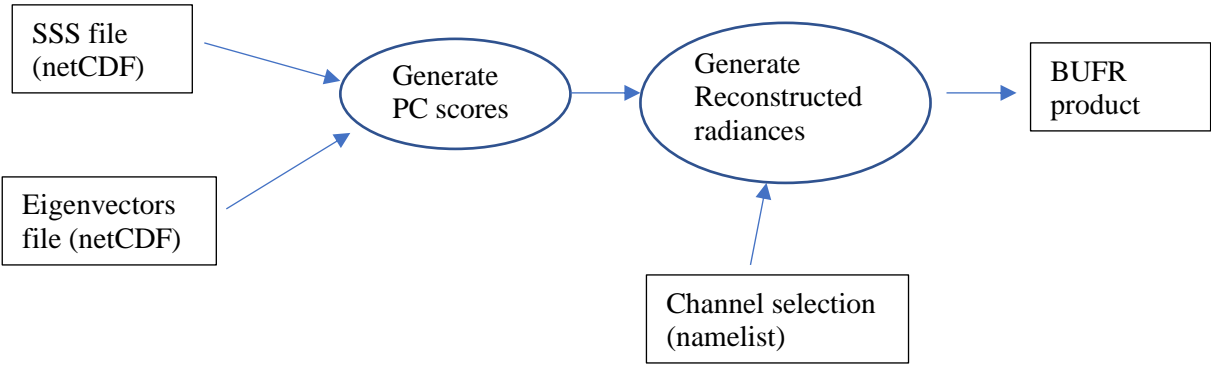


Figure 4: workflow for performing spectral filtering using the user's own eigenvectors

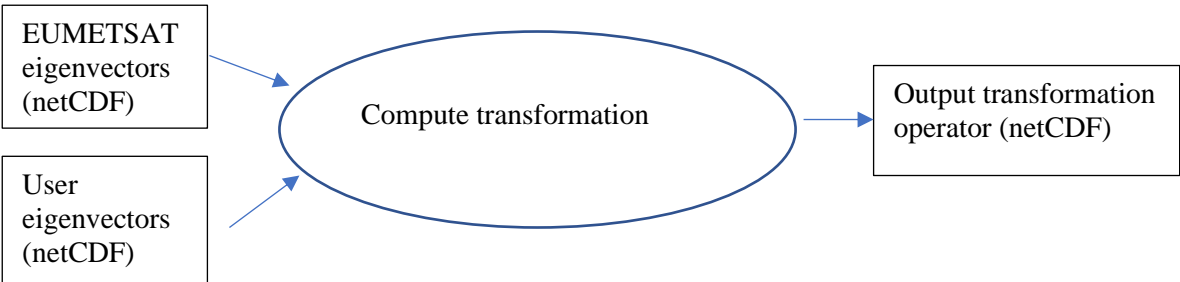


Figure 5: workflow for generating a transformation operator, for converting from one basis function to another

NWP SAF	IRSPP Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	-------------------------------	---

The following points should be noted:

- An input IRS file normally contains 1 dwell, i.e. 160 x 160 spectra, acquired in 10 seconds of observation time.
- There will be an option to thin the data by selecting 1 spot in n spots or 1 line in m lines.
- Processing time is approximately proportional to the number of spots processed. If reconstructed radiances are being generated, processing time is also proportional to the number of channels in the user's channel selection.
- One instance of IRSPP runs on one thread.
- The expectation is that a dwell can be processed in less than 10 seconds (i.e. faster than real-time). But if increased throughput is required, the user can run several instances simultaneously, processing different data files.
- The software will also support processing of simulated IRS spectra from the NWP SAF's Radiance Simulator (RadSim) package (see discussion in section 6.7).

3.2 Relationships with RTTOV and assimilation systems

The RTTOV fast model can be used to simulate IRS radiances in two ways:

1. Classical RTTOV: accurately simulates strongly-apodised radiances (e.g. Hamming-apodised) but does not work well with lightly-apodised radiances, due to the negative sidelobes of the spectral response function (see section 6.5)
2. Using the Havemann-Taylor Fast Radiative Transfer Code (HT-FRTC), which is based on principal components. This is the recommended way to generate lightly-apodised radiances in RTTOV.

Assimilation systems typically generate simulated radiances based on model profiles (this step is outside the scope of IRSPP) and compare with the observed radiances. The function of IRSPP is to generate the observed radiances in a convenient form, either as reconstructed radiance for specified channels or as PC scores for an appropriate set of basis functions (not necessarily the same basis functions that are used for data transmission).

IRSPP would normally be used in near-real-time as a pre-processor, converting the incoming netCDF files (PC scores) to radiances or PC scores in BUFR format. The BUFR files would then be stored in a meteorological database. Some organisations may prefer to embed elements of IRSPP inside their assimilation code, but the details of that would be very dependent on the organisation's particular requirements, so are not discussed in this document.

IRSPP can also be used *off-line* to generate conversion matrices for PC basis function transformation, i.e. to generate a set of PCs that are optimised for model radiances. A convenient way of doing this is to use the NWP SAF Radiance Simulator to generate simulated radiances from model fields, and this has the advantage that there is a well-defined netCDF format for radiances that are presented to IRSPP. An IRSPP test case will be provided that uses such a radiance file. Alternatively, a transformation to the HT-FRTC basis function set can be computed, see section 6.7.

4. SOFTWARE TOP-LEVEL DESCRIPTION

4.1 Fortran code

The software comprises a collection of subroutines, together with main programs that perform different tasks, according to the workflows of the previous section. The main routines that correspond to the workflows of the previous section are listed in Table 1.

NWP SAF	IRSPP Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	-------------------------------	---

Table 1: IRSPP main routines

Main routine	Purpose
irs_main.f90	Processing IRS data for NWP (Figure 1)
irs_covariance.f90	Generating a covariance file (Figure 2)
irs_generate_pcs.f90	Computing eigenvectors from covariance (Figure 3)
irs_sss_filtering.f90	SSS filtering using user eigenvectors (Figure 4)
irs_transform_pcs.f90	Generate matrices that can be used to transform PC scores from one basis function to another (Figure 5)
dummy.f90	Dummy routine, only compiled if LAPACK is not installed

If users wish to write their own main program, to perform something specific, it should be straightforward to edit the Makefile in order to add this program to the build system.

The majority of the subroutines will access IRS data via a data module. This keeps the interfaces simple. (The MWIPP deliverable adopts the same approach). A few subroutines are generic (e.g. the routine that interfaces with LAPACK).

4.2 Namelists

The requirements of the user are specified in two ways: by command arguments (e.g. “-i” or “-o” to give the names of the input/output files) and by namelists (for requirements that are the same from run to run – for example, the file of fixed eigenvectors).

The namelists will follow the standard Fortran structure and details will be given in the User Manual.

5. INPUT AND OUTPUT DATA

The main input data to IRS will normally be netCDF files in the format specified in [RD-2]. The software will accept:

- IRS PC files
- IRS SSS (full-spectrum) files

Each file contains data for 1 dwell (160x160 spectra, gathered in 10 seconds). EUMETSAT plan to make the PC files available in near-real-time, via EUMETCast, but the SSS files will only be available from the Data Archive (see <https://www.eumetsat.int/mtg-data>).

In addition, there will be ancillary data, also in netCDF format, notably the file of fixed eigenvectors.

Some users may want to ingest simulated IRS data that have been generated using a radiative transfer model. IRSPP will be able to ingest netCDF radiance files generated with the NWP SAF’s *Radiance Simulator* package. If a different format is needed (e.g. ASCII) then the user will need to modify the ingest routine.

The main outputs will be:

- BUFR files containing PC scores and/or reconstructed radiances
- There will be an option to write reconstructed radiances back to the input netCDF files

NWP SAF	IRSP Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	------------------------------	---

A draft BUFR sequence is available, see section 7. Note that two new descriptors are proposed, to hold IRS-specific quality flags, which will need to be submitted to WMO.

6. PRINCIPAL COMPONENTS - THEORY

6.1 Converting between radiances and PC scores

Principal Components (PC) scores, \mathbf{p} , are related to the measured radiance spectrum, \mathbf{r} (column vector) and the reconstructed radiance spectrum, \mathbf{r}' , as follows:

$$\mathbf{p} = \mathbf{E}^T \mathbf{N}^{-1} (\mathbf{r} - \mathbf{r}_m) \quad (1)$$

$$\mathbf{r}' = \mathbf{r}_m + \mathbf{N} \mathbf{E} \mathbf{p} \quad (2)$$

where \mathbf{r}_m is a fixed mean radiance spectrum, \mathbf{N} is the noise normalisation matrix (usually the matrix square root of the instrument noise covariance) and \mathbf{E} is a truncated set of eigenvectors ($\mathbf{E}^T \mathbf{E} = \mathbf{I}$, where \mathbf{I} is the identity matrix). Note that \mathbf{E} is of rank $\{nchan \text{ rows}, npc \text{ columns}\}$.

To minimise the number of matrix multiplications being carried out in near-real-time, we define a compression operator $\mathbf{C} = \mathbf{E}^T \mathbf{N}^{-1}$ and a reconstruction operator $\mathbf{R} = \mathbf{N} \mathbf{E}$, so that:

$$\mathbf{p} = \mathbf{C} (\mathbf{r} - \mathbf{r}_m) \quad (3)$$

$$\mathbf{r}' = \mathbf{r}_m + \mathbf{R} \mathbf{p} \quad (4)$$

For self-apodised data, the noise matrix is expected to be diagonal (i.e. white noise in the interferogram), but the ancillary files supplied by EUMETSAT have a full-rank ($nchan \times nchan$) noise matrix.

For computing reconstructed radiances from a set of (global) PC scores, the user only needs to apply equ. (4).

For more information on the principles of principal component analysis, see Antonelli et al., 2004, or Collard et al., 2010.

6.2 Global and local PCs

MTG-IRS PC files contain two types of PC scores: (i) global PCs that relate to a fixed set of eigenvectors held in an external ancillary file, and (ii) a small number of local PCs that are specific to the dwell. (A dwell comprises 160×160 spectra). The purpose of the local PCs is to ensure that any unusual signals (e.g. caused by localised factory emissions or volcanic eruptions) are accurately represented. The local PCs are computed from the residuals, i.e. the differences between raw radiances and those reconstructed using the global PCs. See Hultberg et al., 2017a or 2017b.

IRSP does not provide the facility to *compute* local PCs, as this is the function of the EUMETSAT ground segment. However, it does have the facility to use them when reconstructing radiances. In this case, equation (4) is modified to

$$\mathbf{r}' = \mathbf{r}_m + \mathbf{R}_{\text{global}} \mathbf{p}_{\text{global}} + \mathbf{R}_{\text{local}} \mathbf{p}_{\text{local}} \quad (4a)$$

NWP SAF	IRSP Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	------------------------------	---

6.3 Reconstruction scores

The reconstruction score, or PCR score, for a particular band, is defined as

$$\text{PCR score} = \frac{1}{n_{chan}} \sqrt{\sum_{i=1}^{n_{chan}} \left(\mathbf{N}^{-1}(\mathbf{r}' - \mathbf{r}) \right)_i^2} \quad (5)$$

In the EUMETSAT PC product, the PCR score is computed over all channels in the band. If you are working with SSS data, the PCR score computed by IRSP uses only the channels selected by the user (to save on computing resources).

If the PCR score is significantly greater than 1.0, that is an indication either that there are signals in the observed spectra that are not present in the eigenvectors, or that the instrument noise profile is unrealistic.

6.4 Creating the eigenvectors

The facility to create eigenvectors is an optional part of IRSP and is included for consistency with the earlier NWP SAF "IASI PCA-based compression package". The method implemented in IRSP closely follows that implemented in the earlier package, except that data files are netCDF rather than ASCII.

The process of creating eigenvectors, from a set of reference spectra, comprises (i) the formation of a noise-normalised covariance matrix, **Cov**, and (ii) computation of its eigenvectors and eigenvalues.

To create the covariance matrix, the method implemented in IRSP is first to form the following sums, over all spectra in the training set:

$$\sum \mathbf{r} \quad \text{and} \quad \sum \mathbf{r} \mathbf{r}^T$$

These quantities are held in a netCDF data file. Further spectra can be easily added to the training set, if required.

Then when the training set is complete (with n spectra) we compute the mean spectrum, \mathbf{r}_m , and form the noise-normalised covariance as follows:

$$\mathbf{r}_m = \frac{1}{n} \sum \mathbf{r} \quad (6)$$

$$\begin{aligned} \mathbf{Cov} &= \frac{1}{n} \sum \left(\mathbf{N}^{-1}(\mathbf{r} - \mathbf{r}_m) \right) \left(\mathbf{N}^{-1}(\mathbf{r} - \mathbf{r}_m) \right)^T \\ &= \mathbf{N}^{-1} \left(\frac{1}{n} \sum \mathbf{r} \mathbf{r}^T - \mathbf{r}_m \mathbf{r}_m^T \right) \mathbf{N}^{-1} \end{aligned} \quad (7)$$

where n is the number of spectra. The eigenvectors and eigenvalues are related to each other, and to the covariance matrix, by:

$$\mathbf{Cov} = \mathbf{E} \mathbf{\Lambda} \mathbf{E}^T \quad (8)$$

where $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues. IRSP uses LAPACK routines to perform the eigenvalue decomposition. (Actually, interfaces to two LAPACK routines are provided, but they are functionally equivalent).

NWP SAF	IRSP Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	------------------------------	---

6.5 Apodisation

Apodisation is the process of modifying the instrument’s spectral response function (SRF) in order to control the sidelobes, typically in order to reduce “ringing”. It can be performed as either a multiplication in the interferogram domain or a convolution in the spectral domain.

EUMETSAT products for IRS are based on “light apodisation”. This produces a modest reduction in sidelobes while leaving the main lobe approximately unchanged. Some users prefer a heavier apodisation (e.g. Hamming) – heavy attenuation of the sidelobes at the expense of a broader main lobe. See Figure 6.

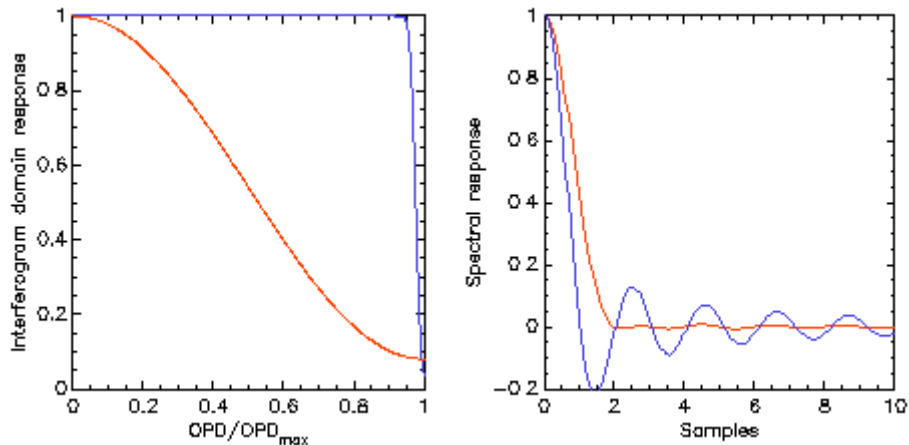


Figure 6: Comparison of light apodisation (blue) with Hamming apodisation (red), in the interferogram (left) and spectral (right) domains. OPD is the optical path difference.

In principle, apodisation is reversible, and does not change the information content of the spectra. However, different types of apodisation can be preferred for different applications.

EUMETSAT’s light apodisation is most easily computed in the interferogram domain: it is the convolution of a top-hat (with edge slightly below OPDmax) with a narrow gaussian. Whereas the Hamming is most easily computed in the spectral domain: simply convolve the spectrum with the 3-element array [0.23,0.54,0.23].

It is clear that users who want a heavy apodisation, and are working with spectra, can easily generate one by applying the Hamming *on top of* the light apodisation supplied by EUMETSAT. This is computationally much easier than first removing the light apodisation (using FFTs) and then applying the Hamming. In practice, the radiance differences between the two approaches are very small (below noise levels – see Figure 7), but it is preferable for radiative transfer to take this into account by simulating the modified Hamming rather than the pure Hamming.

Users who are working with PC scores can alter the apodisation of the reconstructed radiances by apodising the reconstruction operator and mean radiance vector. For the global PCs, this can be a one-off operation.

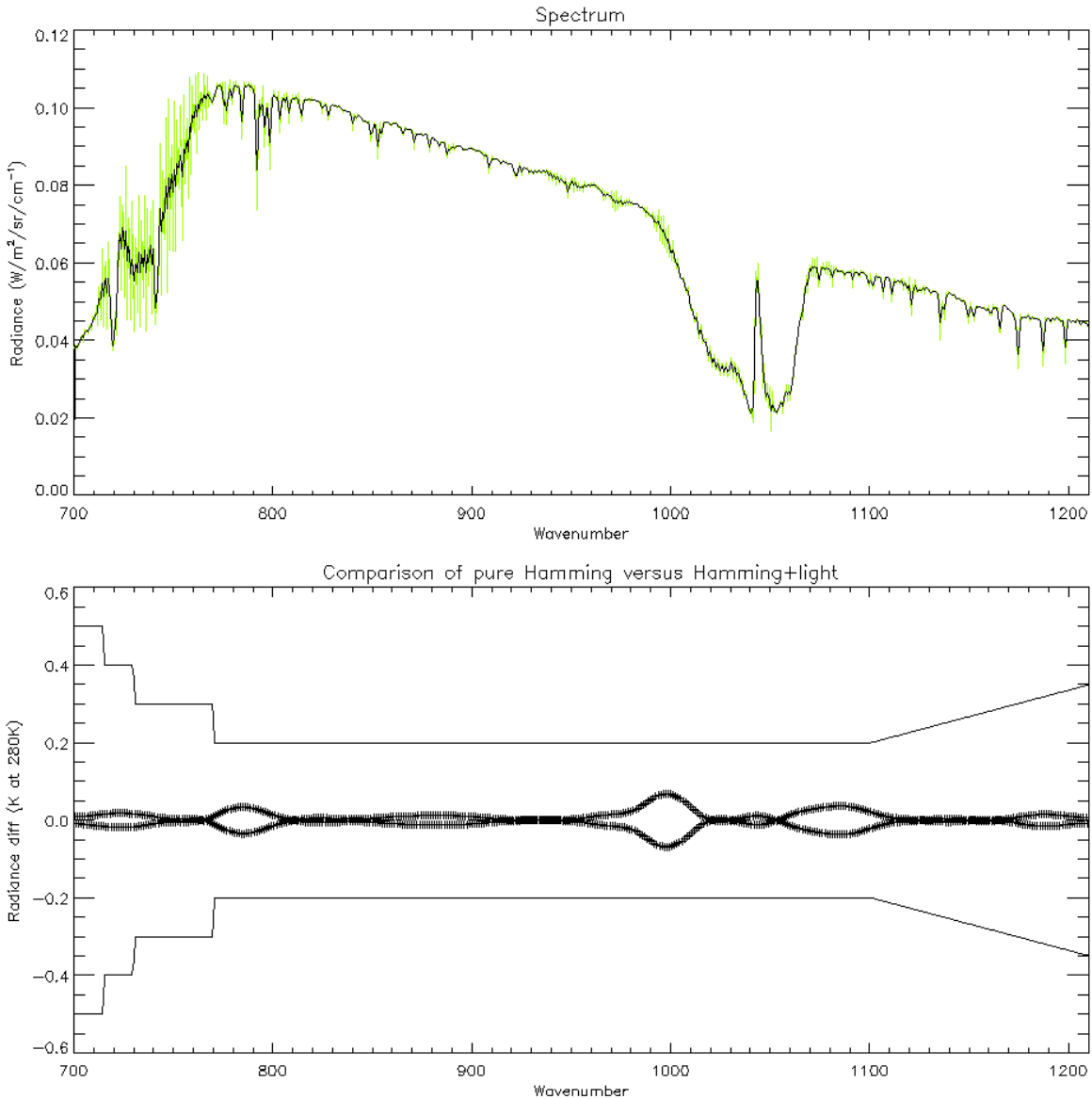


Figure 7: Top: a typical simulated LWIR spectrum with light apodisation (green) and Hamming apodisation (black). Bottom: crosses show the difference between pure Hamming and Hamming applied on top of light. The solid lines at top and bottom show the instrument noise specification.

6.6 Basis function transformation

For some applications it may be required to transform EUMETSAT's PC scores into PC scores appropriate to a different basis function, e.g. based on model-generated spectra. The noise normalisations may also be different (e.g. the second may include forward model uncertainty). If subscript 1 refers to the initial scores and subscript 2 the final scores, then using (1) and (2) we have:

$$\mathbf{p}_1 = \mathbf{C}_1 (\mathbf{r} - \mathbf{r}_{m,1}) \tag{9}$$

$$\begin{aligned} \mathbf{p}_2 &= \mathbf{C}_2 (\mathbf{r} - \mathbf{r}_{m,2}) \\ &\approx \mathbf{C}_2 (\mathbf{r}'_1 - \mathbf{r}_{m,2}) \\ &= \mathbf{C}_2 \mathbf{R}_1 \mathbf{p}_1 + \mathbf{C}_2 (\mathbf{r}_{m,1} - \mathbf{r}_{m,2}) \end{aligned} \tag{10}$$

NWP SAF	IRSPF Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	-------------------------------	---

It is advisable for the number of PCs in the second basis function to be less than or equal to the number in the first, otherwise the PCs will not be independent. The transformation operator $\mathbf{C}_2 \mathbf{R}_1$ has npc_2 rows and npc_1 columns, while the adjustment to the mean $\mathbf{C}_2(\mathbf{r}_{m,1} - \mathbf{r}_{m,2})$ has npc_2 rows.

Having computed the new PC scores, we can then, if required, use \mathbf{R}_2 and $\mathbf{r}_{m,2}$ to compute the corresponding reconstructed radiances.

6.7 Support for HT-FRTC

As noted in section 3.2, the Havemann-Taylor Fast Radiative Transfer Code (HT-FRTC, Havemann et al., 2018) is available in RTTOV and is the recommended method for generating lightly-apodised MTG-IRS simulated spectra. In brief, HT-FRTC works as follows:

1. In the “training” stage, a line-by-line radiative transfer model is used to generate high-resolution (0.001 cm^{-1}) spectra covering a wide frequency range, for typically 1000 diverse model profiles.
2. 300 Principal Components are computed for this high-resolution dataset.
3. A regression is computed to link PC scores to model variables. This allows the PC scores to be estimated for any input model profile. Note that steps 1 to 3 are instrument independent.
4. For each instrument, the high-resolution PCs are convolved with the instrument spectral response function to give the Reconstruction matrix – allowing PC scores to be converted to simulated radiance. This information is held in a data file (e.g. *hffrtc_coef_sensor_mtg_irs.nc*)

For IRSPF, we can ask the question as to whether it is possible to convert EUMETSAT’s PC product into equivalent HT-FRTC scores.

Technically, it is indeed possible. The HT-FRTC Reconstruction matrix, \mathbf{R} , can be inverted to give a compression matrix, \mathbf{C} , via $\mathbf{C} = (\mathbf{R}^T \mathbf{R})^{-1} \mathbf{R}^T$, satisfying $\mathbf{C} \mathbf{R} = \mathbf{I}$. Then we can use basis function transformation (section 6.6) to compute the new PC scores. However, there are some differences in the processing due to the fact that in HT-FRTC the two bands are combined together whereas the EUMETSAT PC product treats them separately. Therefore we would need to merge the EUMETSAT reconstruction operators (\mathbf{R}_{LW} and \mathbf{R}_{MW}) before computing the transformation operator ($\mathbf{C}_2 \mathbf{R}_1$):

$$\mathbf{R}_1 = \begin{pmatrix} \mathbf{R}_{LW} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{MW} \end{pmatrix} \quad (11)$$

Also, the PC scores for the two bands need to be concatenated together before they are transformed.

For the reconstruction operators (to convert from 300 HT-FRTC PC scores to reconstructed radiance), a separate matrix can be used for each band.

Users should also note the following points:

- The PCs in HT-FRTC are optimised for the high-spectral-resolution training set and may not be optimal for MTG-IRS.
- The HT-FRTC Reconstruction matrix is far from orthogonal, i.e. the radiance normalisation (\mathbf{N} in equations (1) and (2)) is full rank. This complicates the interpretation of the PC scores.
- The magnitude of the PC scores is of order 1 – they are not signal to noise ratios. Since PC scores in IRSPF are held as integers (to match EUMETSAT’s PC product), the quantisation (q in *irspp_main_namelist*) must be set to a low value, e.g. 0.0001, to avoid loss of precision.
- When encoding the transformed PC scores into BUFR, *npc_out_lw* should be set to 300 and *npc_out_mw* to 0, to avoid unnecessary duplication.

NWP SAF	IRSP Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	------------------------------	---

IRSP supports the ingest and use of RTTOV's HT-FRTC coefficient files, as described above, and illustrated in Figure 8. It is an open question as to whether there are scientific benefits in such a treatment.

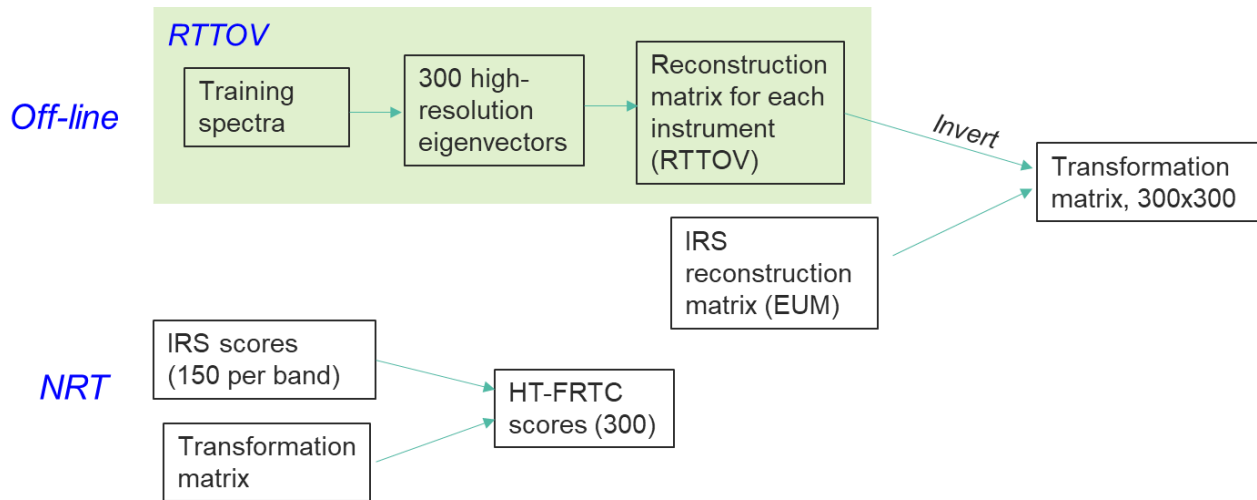


Figure 8: Top: off-line creation of a transformation matrix. Bottom: conversion of disseminated PC scores into equivalent HT-FRTC scores

An alternative way of using HT-FRTC is to use it to generate simulated spectra from a set of model profiles (e.g. using the NWP SAF Radiance Simulator, as mentioned in section 3.2, and noting that IRSP supports the reading of RadSim netCDF output files). The user can then use the workflows of Figure 2 and Figure 3 to generate a covariance dataset and new PCs that are fully consistent with the model variables used in the simulation, and with the supplied noise profile. A transformation operator can then be computed for each band (as in section 6.6) to allow EUMETSAT PC scores to be readily transformed to the new PC scores.

7. PROTOTYPE BUFR SEQUENCE

In this section a prototype BUFR sequence is proposed. It supports inclusion of PC scores, radiances or both. The number of PCs or channels is set by delayed replication. Note that two locally-defined quality flags are used: *irsSpatialSampleQuality* and *irsDetectorSampleQuality*. The bit definitions follow those given in RD-2. It is planned that official WMO descriptors will be provided eventually.

We have included the mean and standard deviation of “imager-mode radiances”, providing a measure of scene inhomogeneity. This reflects the fact that whereas a dwell comprises 160x160 IRS pixels, each of these actually comprises a 3x3 array and is reported as a broad-band radiance. We have re-used IASI BUFR descriptors that refer to AVHRR.

```

001007 satelliteIdentifier (CCT C-5)
001033 centre (CCT C-1)
001034 subCentre (CCT C-12)
002019 satelliteInstruments (CCT C-8)
002020 satelliteClassification (334 = MTG? To be allocated by WMO.)

301011 (year, month, day: 004001, 004002, 004003)
301012 (hour, minute: 004004, 004005)
207003 (increase scale and width)
004006 second

```

NWP SAF	IRSPP Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	-------------------------------	---

207000 (reset scale and width)
201135 (increase width from 8 to 15 bits)
005043 fieldOfViewNumber (1 to 25600 - 160x160)
201000 (reset width)
005041 scanLineNumber
005045 fieldOfRegardNumber (dwell)
002165 radianceTypeFlags (4=apodized, 5=unapodized)
202126 (decrease scale)
201132 (increase width)
007001 heightOfStation (m, to nearest 100m, geostationary height range)
201000 (reset width)
202000 (reset scale)

124002 Replicate 24 descriptors 2 times (for the 2 bands)

008076 band (2=LW, 3=MW)
006029 waveNumber (start)
006029 waveNumber (end)
025140 startChannel
025141 endChannel
301021 (latitude, longitude: 005001, 006001)
007024 satelliteZenithAngle
005021 bearingOrAzimuth
007025 solarZenithAngle
005022 solarAzimuth
025142 channelScaleFactor
014047 scaledMeanAvhrrRadiance (imager mode, 3x3 pixels)
014048 scaledStandardDeviationAvhrrRadiance
033230 irsSpatialSampleQuality (14 bit flag table - local definition)
033231 irsDetectorSampleQuality (5 bit flag table - local definition)
025187 confidenceFlag (0=valid, 1=invalid, 15=missing)
207002 (increase scale and width)
040026 scoreQuantizationFactor
207000 (reset scale and width)
040016 residualRmsInBand
025062 databaseIdentification
101000 (replication)
031002 extendedDelayedDescriptorReplicationFactor
040017 nonNormalizedPrincipalComponentScore

008076 band (set to missing)

104000 (replication)
031002 extendedDelayedDescriptorReplicationFactor
201133 (increase bit width)
005042 channelNumber
201000 (reset bit width)
014044 channelRadiance

Quality flags:

irsSpatialSampleQuality, 14-bit (bit 1 defined here as most significant)

Bit	Meaning
1-4	reserved
5	solar_straylight_correction_warning
6	solar_straylight_warning
7	noisy_detector_sample_warning
8	undersaturated_detector_sample_warning
9	saturated_detector_sample_warning
10	dust

NWP SAF	IRSPP Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	-------------------------------	---

```

11 cloudy
12 limb_view
13 space_view

```

irsDetectorSampleQuality, 5-bit (bit 1 defined here as most significant)

```

Bit Meaning
1 excluded_detector_sample
2 noisy_detector_sample
3 undersaturated_detector_sample
4 saturated_detector_sample

```

It is a BUFR regulation that in a flag table the least significant bit remains unused, to allow “missing” to be uniquely defined (i.e. all bits set).

8. REQUIREMENTS TRACEABILITY

Table 2 provides a cross-reference between requirements, as stated in RD-1, and the relevant section of this document.

Table 2: Requirements traceability matrix

Requirement from [RD-1]	Applicable section(s) of the top level design
Ingest of native-format (netCDF4) PC-score data from MTG-IRS	Workflow of Figure 1, implemented in irs_main.exe
Code to convert PC scores to reconstructed radiances for specific channels, using externally-supplied eigenvectors. These eigenvectors may be a fixed set or there may also be some dwell-dependent eigenvectors.	Workflow of Figure 1, implemented in irs_main.exe
Code to convert PC scores to an alternative basis function, using an externally-supplied transformation matrix	Workflow of Figure 5, implemented in irs_transform_pcs.exe
Conversion of PC scores to BUFR	Workflow of Figure 1, implemented in irs_main.exe. BUFR sequence of Section 7.
Conversion of reconstructed radiances to BUFR	Workflow of Figure 1, implemented in irs_main.exe. BUFR sequence of Section 7
Conversion of reconstructed radiances to NetCDF4	Workflow of Figure 1, implemented in irs_main.exe. The reconstructed radiances are appended to the input file.
Spatial sub-sampling, to a user-defined grid (nearest neighbour approach)	The facility to thin the observations (e.g. 1 in n north-south and east-west) will be provided via namelist control. More sophisticated schemes could be considered if required. Note that the granularity of the input data is a dwell and the filename will indicate the dwell number.
Ability to change the apodisation of the eigenvectors (for example, convert EUMETSAT's “light” apodisation to a Gaussian)	Workflow of Figure 1, implemented in irs_main.exe. The manipulation of the eigenvectors is discussed in Section 6.5
Facility to download ancillary information (e.g. eigenvectors) from EUMETSAT, where required	Not currently implemented as it is assumed that eigenvector files will be made available alongside the spectra. Can be added separately if needed.
Tool for generating eigenvectors from a base set of spectra, via a covariance matrix,	Workflows of Figure 2 and Figure 3, implemented in irs_covariance.exe and irs_generate_pcs.exe. To

NWP SAF	IRSP Top Level Design	Doc ID : NWPSAF-MO-DS-043 Version : 1.4 Date : 31.01.2022
----------------	------------------------------	---

incorporating the relevant functionality of the IASI PCA-based Compression Package.	use these eigenvectors in conjunction with full-spectrum datasets, see Figure 4 and <code>irs_sss_filtering.exe</code> .
---	--

9. REFERENCES

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.*, 109(D23).
<https://doi.org/10.1029/2004JD004862>

Collard, A. D. (1997): Selection of IASI channels for use in numerical weather prediction, *Quarterly Journal of the Royal Meteorological Society*, Volume 133, Issue 629, pages 1977-1991, October 2007 Part B. <https://doi.org/10.1002/qj.178>

Collard, A. D., A. P. McNally, F. I. Hilton, S. B. Healy, N. C. Atkinson (2010): The use of principal component analysis for the assimilation of high-resolution infrared sounder observations for numerical weather prediction, *Quarterly Journal of the Royal Meteorological Society*, Volume 136, Issue 653, pages 2038-2050, October 2010 Part B. <https://doi.org/10.1002/qj.701>

Coppens, D., B. Theodore, T. August, T. Hultberg, C. Goukenleuque, Jochen Grandell (2019): MTG-IRS: Scientific Improvements For a User-Friendly Mission. Proc. ITSC-22,
<http://cimss.ssec.wisc.edu/itwg/itsc/itsc22/presentations/5%20Nov/14.01.coppens.pdf>

Havemann, S., J-C. Thelen, J. P. Taylor, R. C. Harlow, 2018: The Havemann-Taylor Fast Radiative Transfer Code (HT-FRTC): A multipurpose code based on principal components, *Journal of Quantitative Spectroscopy and Radiative Transfer*, Volume 220, Pages 180-192, <https://doi.org/10.1016/j.jqsrt.2018.09.008>

Hultberg, T., August, T. and Lenti, F. (2017a): Local or global? How to choose the training set for principal component compression of hyperspectral satellite measurements: a hybrid approach. In *Sensors, Systems, and Next-Generation Satellites XXI* (Vol. 10423, p. 104231G). International Society for Optics and Photonics, <https://doi.org/10.1117/12.2278349>

Hultberg, T., August, T. and Lenti, F. (2017b): A global-local hybrid approach to retain new signals in hyperspectral PC products. Proc. ITSC-21,
https://cimss.ssec.wisc.edu/itwg/itsc/itsc21/program/1december/0830_8.01_ITSC_2017_Hybrid_PC.pdf