NWP SAF

IRSPP v2 Top Level Design

Version 1.0

16th July 2024

IRSPP 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 September 2021, 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.

Change record			
Version	Date	Author / changed by	Remarks
1.0	16/07/2024	Nigel Atkinson	

Table of Contents

1. INTRODUCTION

This document presents the Top Level Design for version 2 of the Infrared Sounder Pre-processor, IRSPP, a software deliverable of the NWP SAF. Information on the package can be found at https://nwp-saf.eumetsat.int/site/software/irspp/.

It is an update from the Top Level Design document issued with IRSPP v1 (RD-2) and is intended to accompany the IRSPP Version 2 Product Specification (RD-5).

RD-2 also included the theory of PC compression (i.e. it functioned as an Algorithm Theoretical Basis Document), but now that this information has been incorporated into a User Manual [RD-4], with more detail added, the ATBD section will only contain details about things that have changed since IRSPP v1.

Please note that the main difference between IRSPP v2 and IRSPP v1 is that v2 will be released after validation with post-launch data. The review of user requirements (see RD-5) did not identify any significant changes in requirements, or in the required functionality of IRSPP.

1.1 Reference documents

[RD-1] IRSPP Product Specification, NWPSAF-MO-DS-037, v1.0, April 2018, available on the NWP SAF web site.
[RD-2] IRSPP Top Level Design, NWPSAF-MO-DS-043, v1.4, Jan 2022
[RD-3] MTG IRS Level 0 & 1 Format Specification [IRSL1FS], EUM/MTG/SPE/10/0449
[RD-4] IRSPP User manual, NWPSAF-MO-UD-053, v1.3, Feb 2023
[RD-5] IRSPP Version 2 Product Specification, NWPSAF-MO-DS-051, v0.1, April 2024

2. INSTALLATION

2.1 Delivery

IRSPP 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

IRSPP is intended to be built on a Linux platform, using a ksh or bash environment. A Fortran90 compiler (e.g. gfortran) is required. IRSPP 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.

IRSPP 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.

2.3 Building IRSPP 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 IRSPP

The IRSPP 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 IRSPP are shown in Figure 1 to Figure 5. For operational use, it is expected that only the first of these workflows 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 IRSPP v2 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.

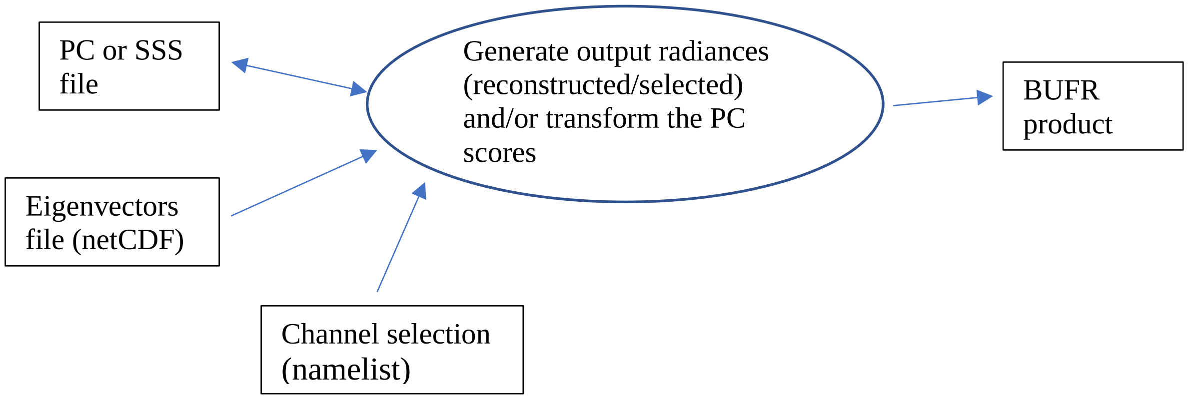

Figure 1: Workflow for the core task of processing IRS data for use in NWP

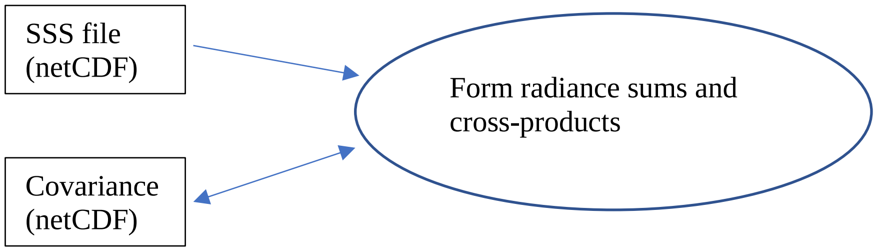

Figure 2: Workflow for generating covariance matrix from many SSS files

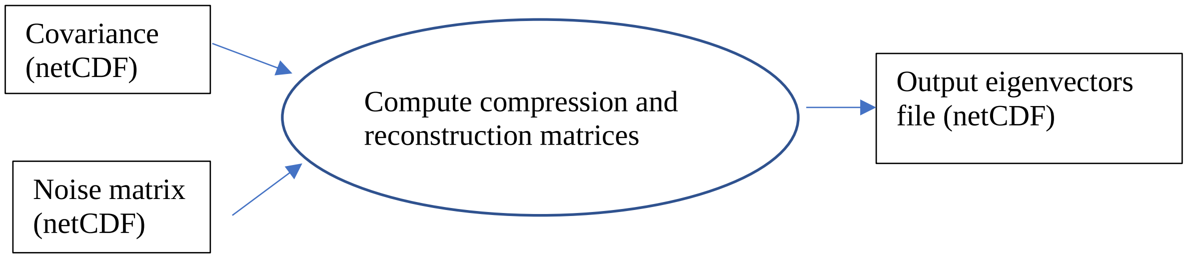

Figure 3: Workflow for generating user-defined eigenvectors

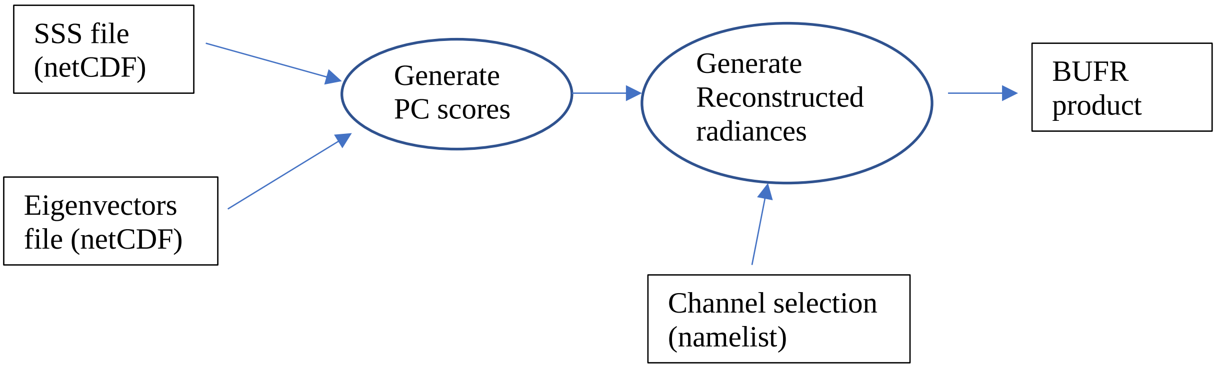

Figure 4: workflow for performing spectral filtering using the user's own eigenvectors

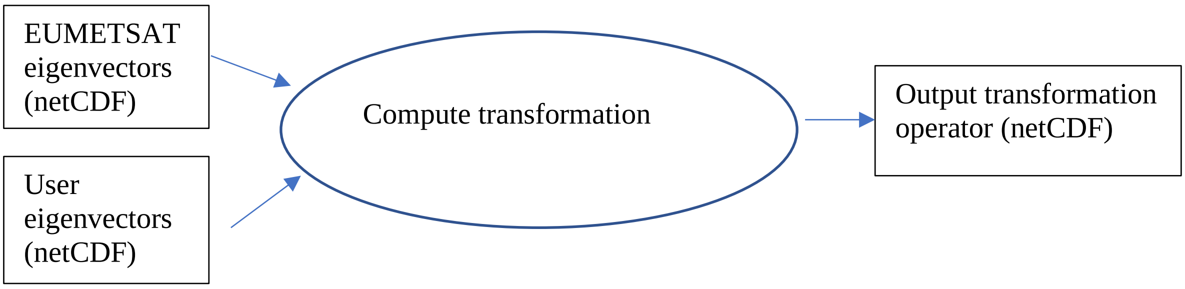

Figure 5: workflow for generating a transformation operator, for converting from one basis function to another

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 [RD-4]).

3.2 Relationships with RTTOV and assimilation systems

The NWP SAF deliverable RTTOV (Saunders et al., 2018; version 14 planned for release in late 2024) is a fast model that 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
2. PC-RTTOV (Matricardi, 2010) simulates Principal Component scores, from which reconstructed radiances can be generated. RTTOV v14 supports both clear-sky and cloudy/scattering PC simulations.

Previous versions of RTTOV supported the Havemann-Taylor Fast Radiative Transfer Code (HT-FRTC, Havemann et al., 2018), which is based on principal components. This model was capable of simulating lightly-apodised IRS radiances, but it is not supported in RTTOV v14.

The PC-RTTOV eigenvectors are Hamming-apodised, i.e. reconstructed radiances would also be strongly apodised. In principle, the PC regression (section 6) could be trained with lightly-apodised line-by-line spectra, but this has not been demonstrated. Currently, RTTOV only supports strongly apodised IRS spectra.

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. One way of doing this is to use the NWP SAF Radiance Simulator package 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 PC-RTTOV basis function set could be computed, see section 6.

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.

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

A BUFR sequence is available, see section 7. Note that local descriptors are used to hold IRS-specific quality flags.

6. PRINCIPAL COMPONENTS – THEORY

See section 8 of the IRSPP User Manual

For IRSPP v2, the basic theory related to generating reconstructed radiances from the IRS PC product is unchanged. However, an important difference is that HT-FRTC is no longer supported in RTTOV v14. Users that need a PC-based RT model are encouraged to use PC-RTTOV instead (Matricardi, 2010). As mentioned in section 3.2, the PC simulation is based on Hamming-apodised spectra.

The technical differences between HT-FRTC and PC-RTTOV are summarised below:

HT-FRTC (text from RD-4)

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. *htfrtc_coef_sensor_mtg_irs.nc*)

<u>PC-RTOV</u>

1. For clear sky, in the "training" stage, a line-by-line radiative transfer model (LBLRTM) is used to generate high-resolution (0.001 cm^{-1}) spectra covering a wide frequency range, for 195000 model profiles.
2. For each supported instrument, the training spectra are convolved with the instrument line shape to produce polychromatic radiances.
3. For the scattering simulations (aerosols, clouds), instead of LBL the fast Chou-scaling solver is used with RTTOV v14, for the given instrument. There are 63135 training profiles for aerosols and 15000 training profiles for clouds, giving a total of 273135 training profiles for the RTTOV v14 PC coefficients.
4. 400 Principal Components are computed based on the polychromatic radiances.
5. A regression is computed to link PC scores to polychromatic radiances, for a specifically chosen set of "classical" RTTOV radiances. The result is stored in a PC coefficient file (e.g. *pccoef_mtg_2_irs-hamming-2mopd_v13p_landsea_6gas_clr_nlte_aer_hydro.nc*).
6. At run time, the user uses normal RTTOV (based on optical depth files) to simulate radiances for a set of channels specified in the PC coefficient file (the actual number of channels is user-configurable: a larger subset is slower but yields more accurate simulations), then the regression coefficients from step 5 are applied to compute the PC scores. Optionally, the PC scores can be used to compute reconstructed radiances for the user's choice of channels.

The PC coefficient file (step 5) contains a 1-D noise normalisation vector and the eigenvectors (noting that the number of channels will be updated in the near future):

```
DATASET "/pccoef/noise" {
   DATATYPE  H5T_IEEE_F64LE
   DATASPACE  SIMPLE { ( 1960 ) / ( 1960 ) }

DATASET "pccoef/eigen/01/coefficients" {
   DATATYPE  H5T_IEEE_F32LE
   DATASPACE  SIMPLE { ( 400, 1960 ) / ( 400, 1960 ) }
```

It should therefore be possible to create a transformation matrix in a similar way to that described in RD-4, allowing EUMETSAT PC scores for IRS to be transformed to PC-RTTOV scores. The capability will be updated in IRSPP v2: functions specific to HT-FRTC will be removed, and others based on PC-RTTOV will be added.

7. BUFR SEQUENCE

This section details the BUFR sequence proposed for IRSPP v2. There are a few minor changes compared with that used in IRSPP v1.2. 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-3.

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, 72=Meteosat-13=MTG-S1, 75=Meteosat-16=MTG-S2)
001033 centre (CCT C-1)
001034 subCentre (CCT C-12)
002019 satelliteInstruments (CCT C-8, 212 = IRS)
002020 satelliteClassification (334 = MTG)

301011 (year, month, day: 004001, 004002, 004003)
301012 (hour, minute: 004004, 004005)
207003 (increase scale and width)
004006 second
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)
005066 spacecraftYaw (0 or 180 degrees)

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:
 spatial_sample_quality, 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
  11   cloudy
  12   limb_view
  13   space_view

 detector_sample_quality, 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. CAPABILITY TRACEABILITY

Table 2 provides a cross-reference between high-level capabilities, as stated in section 3.1 of RD-5, and the relevant section of this document.

Table 2: Capability traceability matrix

Capability from [RD-5]	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. Also the facility to select a "warmest spectrum" based on radiance of a specific channel. 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 [RD-4]
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, incorporating the relevant functionality of the IASI PCA-based Compression Package.	Workflows of Figure 2 and Figure 3, implemented in irs_covariance.exe and irs_generate_pcs.exe. To use these eigenvectors in conjunction with full-spectrum datasets, see Figure 4 and irs_sss_filtering.exe.
Validated prior to release using in-orbit data	To be demonstrated once real data are available.

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

Matricardi, M., 2010: A principal component based version of the RTTOV fast radiative transfer model, https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/qj.680

Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., Vidot, J., Roquet, P., Matricardi, M., Geer, A., Bormann, N., and Lupu, C., 2018: An update on the RTTOV fast radiative transfer model (currently at version 12), Geosci. Model Dev., 11, 2717–2737, https://doi.org/10.5194/gmd-11-2717-2018