

USER'S GUIDE FOR IASI STAND ALONE RETRIEVAL SOFTWARE

*(ATMOSPHERIC TEMPERATURE AND MOISTURE, SURFACE TEMPERATURE AND EMISSIVITY,
CLOUD-FREE CONDITIONS)*

Written by: G. MONNIER

MF/DP/CMS/R&D

Verified by:

Approved by:

Table of contents

Contents	Page
1. Software introduction and principles	3
1.1 First guess computation	3
1.2 1D-Var retrieval	5
1.3 Treatment of surface emissivity	6
2. Main tasks description and general remarks	6
3. Test dataset	7
4. Programs brief descriptions	9
IASI_RTBASEPROF	9
IASI_REGRESSION	10
IASI_SIMUL	13
5. User-files descriptions	14
baseprof.dat	14
trueprof.dat	14
obs_simul.dat	14
selection_1dvar.dat, selection_guess.dat	15
emiss_3cm-1.dat, emiss_3cm-1.info	16
model_emiss.txt	16
iasi_par.txt	17
res_guess.dat	20
res_1dvar.dat	20
res_emiss.dat	21
6. Table of directories contents	22

1. Software introduction and principles

The present code is the first version of a stand-alone retrieval software for IASI data. The current version applies to cloud-free situations only. Basically, a set of IASI brightness temperatures is supplied as the input by the user, and the code outputs a corresponding set of first-guess and retrieved atmospheric temperature and water-vapor profiles together with surface temperature and emissivity.

The retrieval process includes two distinct steps:

- I) a set of linear-regression coefficients is applied to the observations, to produce a first-guess set of atmospheric and surface parameters.
- II) A non-linear 1D-Var algorithm improves iteratively this first estimate to produce the final analyzed state.

The software includes programs which compute the regression coefficients from a user-supplied profile library (IASI_RTBASEPROF, IASI_REGRESSIONS, IASI_COVGUESS), and a program which applies the regression coefficients and performs the 1D-Var retrieval, IASI_INVERSION. The fast radiative transfer code RTIASI (Matricardi and Saunders, 1999) ensures all forward computations, which are required for both steps.

1.1 First guess computation

Profile library

The computation of regression coefficients to produce a first-guess of the atmospheric states implies the use of a sufficiently large profile library. Regression coefficients are computed to best fit the relationship between synthetic brightness temperatures computed from the profiles of the library, and the profiles themselves. As the only external source of information provided to the retrieval process, the profile library strongly affects the final accuracy. Particularly, a “rolling library”, provided with recent atmospheric analyses, proved to achieve much better results than any “climatological”, static library. A key-point is that profiles used to build the library and those considered as “the truth” when computing the retrieval and first-guess errors should come from the same analyses source. Otherwise, very large bias can arise, probably mainly because of discrepancies in extrapolation methods near the surface and in the upper stratosphere.

For those reasons, the profile library is considered here as a user-supplied dataset. It is entirely stored in the **baseprof.dat** ascii file.

Forward radiative transfer computations

Before any linear regression computation, forward radiative transfer calculations must be achieved in order to simulate IASI observations for each profile of the library, and for a set of surface pressures and satellite angle values. The stored quantities are not the observations themselves (i.e. the brightness temperatures) but the corresponding atmospheric transmittances and radiances (upward and downward). This allows to quickly compute the brightness temperature later, for any surface conditions (temperature, emissivity). At each RTIASI call, the radiances and transmittances are interpolated at all required surface pressures. Thus, the total number of RTIASI calls is (number of profiles in the library)X(number of required satellite angles).

Channel selection for the first guess computations

Profiles predictors are computed from observations in a limited number of IASI channels only. This channel selection can be either supplied by the user, or computed by the software (see IASI_RTBASEPROF and **selection_guess.dat**). In the latter case, a selection algorithm based on c. Rodgers' iterative method is applied (Rodgers ???) . This method basically performs successive linear 1D-Var analyses, taking into account an increasing number of channels. The channel that produce the best increase of information (Degree of Freedom for Signal in the current code, or, with a slight change, analysis error entropy) is selected at each step. Though very computationally expensive, this approach proved to be the most efficient one (Rabier et al. 2001). One of its advantage is that it fully accounts for the background knowledge of the parameters to be retrieved, through the background error covariance matrix which initialize the iterative process. In the present case, this knowledge is the profile library variability, and the "error" is its covariance matrix itself. In order to keep regressions over temperature parameter as insensitive as possible to water vapor variables, temperature and water vapor channels are selected and stored separately.

Though already expensive, one selection process provide a set of channel optimized for one given profile only. In order to build a more general selection, a loop must be performed over several profiles. Currently, those are profiles averaged over user-specified integrated water content intervals. However, it seems that those profiles do not need to be nor numerous, nor fully representatives, as the successive selections quickly wrap each-other.

In order to limit the computing time, the selection itself is performed over limited spectral bands of interest, which can be easily modified in the code (`iasi_rtbaseprof.f90`). Presently, all selectable channels have wave numbers smaller than 2200 cm^{-1} , in order to avoid solar radiations.

Regression coefficient computation

Computing the regression coefficient involve two distinct steps:

- computing the principal components: regressions are not performed on the observations, but on a limited number of their principal components. This can be a quite expensive computation step, but principal components do not necessary need to be recomputed at each regression coefficients update. For this reason, they are stored separately (see **comp_t.dat**, **comp_q.dat**, **comp_iwc.dat**, **pca.info**), and are computed only if the corresponding files are absent. The "optimal" number of components is detemined by minimizing the average noise of the spectra reconstructed with only an increasing number among them.
- Computing the regression coefficients themselves: this is a much faster computing step, where least-square linear regressions are performed over the observations principal components as independent variables, and the corresponding profile and surface parameters as dependant variables.

The main drawback of such regressions is linearity, especially when water vapor parameters are implied. In order to reduce non-linearity, separate regression coefficients are computed over the profiles belonging to a few integrated water vapor content (iwc) intervals (those are user-defined, see **iasi_par.txt**). Thus, an additional regression over the iwc is needed. It will be later apply to the observations prior to the first-guess computation, in order to produce a coarse iwc estimate and to classify the situation in the right iwc interval.

First guess profiles computation

They are simply generated by applying the pre-computed regression coefficients to the observations, given the satellite zenith angle, surface pressure, and surface emissivity type. Large

errors in the iwc first estimate can lead to situations misclassifications, and thus obviously bad first-guess profiles. Such cases can be identified and rejected thanks to the resulting discrepancy between the regression iwc first estimate, and the iwc computed from the first guess temperature and water-vapor profiles

Before any retrieval is performed, the first guess profiles corresponding to every true profiles included in the library are extracted, in order to compute the first-guess error covariance.

1.2 1D-Var retrieval

Channel selection for the 1D-Var retrieval

The same channel selection could be used for both first-guess and 1D-Var computations. However, as the 1D-Var algorithm is intended to improve the first-guess error, the channel selection should ideally account for this error. In other words, channels bringing information on parameters with large error in the first guess should be selected first. This is achieved applying Rodgers' iterative selection method again, accounting this time for the first-guess error rather than the profile library covariance. Here, while the 1D-Var algorithm treats temperature and water vapor parameters altogether, no distinction is made between temperature and water-vapor channels. Like the previous one, this channel selection is performed only if the relevant file is not found.

1D-Var retrieval

This algorithm iteratively improves the first-guess profile and surface parameters, by successive addition of analysis increments. Those increments are computed given the first-guess and its error covariance matrix, IASI observations and instrumental noise, and using the forward radiative transfer code RTIASI and its linear tangent (or jacobian). The whole 1D-Var process for a given situation can be very expensive, mostly because of the jacobian computations. A way to limit this computation time is not to update the jacobian at each iteration. The minimum and maximum iteration number and the jacobian update frequency are all user-specified parameters (see **iasi_par.txt**).

1D-Var control variables are:

- temperatures at all model pressure levels (currently 43 levels)
- surface temperature
- water vapor at the 25th first pressure levels above the surface (easily modified). While humidity as positive values only, but possibly very low, it was found more convenient to introduce its logarithm $q_{\text{scale}} \cdot \ln(Q_{\text{ppm}})$ as a control variable. q_{scale} is an arbitrary scaling factor.
- Surface emissivity, averaged within several spectral bands (see next paragraph). Those variables are lower than, and possibly nearly equal to unity, thus the chosen corresponding control variables are $\epsilon_{\text{scale}} \cdot \ln(1 - \epsilon_i)$, here again with a scaling constant ϵ_{scale} .

1D-Var efficiency strongly depends on the first guess error statistics and the actual situation, including the surface properties (temperature and emissivity). Iterating can be useful, quickly reducing the analysis variance and bias, or unnecessary when the first guess is already not far from the truth, or even nearly inefficient in the worst cases. Thus, all user-specified parameters should be empirically adapted to the profile library and the situations actually treated.

1.3 Treatment of surface emissivity

As mentioned above, the surface emissivity can have important effects on the final retrieval. Over lands, the surface emissivity is often supposed to have a constant value slightly lower than unity, typically 0.98. On the one hand, this is not the truth, and some terrain can have an emissivity which varies over the spectrum, and reaches values significantly lower than unity, therefore degrading the atmospheric retrieved profile. On the other hand, a low emissivity generally contributes to enhance the measurements sensitivity to low-level water vapor, like does a large temperature difference between the ground and the surrounding moist air layer. Taking advantage of low-emissivity surfaces is thus a way to reduce the first-guess water vapor errors, which are generally large.

For those reasons, emissivity parameters have been introduced as control variables in the 1D-Var algorithm. This requires to simulate high-resolution emissivity spectra for the considered surface types, and the corresponding average and covariance, to be used as the first-guess emissivity and its error in the retrieval process. This is achieved by mixing emissivity spectra for individual materials from the **MODIS UCSB Emissivity Library** (data set collected by Dr. Zhengming Wan's Group at ICESSE). The user defines surface emissivity types by defining, for each among them, the constitutive materials and corresponding average, minimum and maximum fractions. 3cm^{-1} resolution random spectra can then be generated for a surface with a given type. Those spectra are to be used to simulate IASI observations.

They also allow to compute the emissivity first-guess – simply the average emissivity for each type, considering there is no type error -, and the corresponding covariance matrix. At this step, the emissivity spectra are averaged over bands with a constant width (currently 100 cm^{-1} , can be easily changed) covering the entire IASI spectral range. The average emissivity within those bands are the variables introduced in the 1D-Var retrieval. More precisely, only bands where channels with sufficiently large mean transmission are lying are introduced, other bands being insensitive to surface emissivity.

2. Main tasks description and general remarks

The retrieval software is composed of four distinct program units, which exchange data through several files stored on the disk. Their roles can be summarized as follows:

- IASI_RTBASEPROF computes and stores atmospheric radiances and transmittances for all the profiles contained in the user-supplied library. If necessary, selects the channels to be used for the first-guess computations.
- IASI_REGRESSION computes and stores the principal components and the regressions coefficients for the first guess profiles (atmospheric temperature and water-vapor, surface temperature), from the profile library and the corresponding radiances and transmittances.
- IASI_COVGUESS performs the regressions on the whole profile library, and stores the resulting first-guess error covariance matrices. If necessary, selects channels to be used by the 1D-Var retrieval.
- IASI_INVERSION computes first-guess profiles from the regression coefficients and the user-supplied brightness temperatures. Performs a 1D-Var retrieval of atmospheric temperature and water-vapor and surface temperature and emissivity, from the observations, the first-guess and the covariance matrix.

A full run, including the generation of all the data files except those supplied by the user, requires one execution of each program, in the order of the previous description. The schematic hereafter shows the way data are exchanged between the programs. A few files are not mentioned here:

- those related to the simulation of surface emissivity, gathered by all four programs
- RTIASI data files
- `iasi_par.txt`, a file of control parameters which can be changed by the user without entering the code.

Though updating the channel selection files (**`selection_guess.dat`** and **`selection_1dvar.dat`**) and the principal components files (**`comp.info`**, **`comp_iwc.dat`**, **`comp_t.dat`**, **`comp_q.dat`**) is particularly time-demanding, this may not be necessary at each run. For this reason, those files are only restored when not found in the corresponding directories, so that the user can force the updates or maintain the older data.

The final retrieval results are stored in three files:

- **`res_guess.dat`** is written on the disk as soon as the regression-computed first guess profiles are available, before the start of any 1D-Var computation. It contains the first-guess profiles together with the true profiles, to facilitate any further error computation and plot.
- **`res_1dvar.dat`** contains the final 1D-Var retrieval (atmospheric and surface temperatures, atmospheric water-vapor), as well as the corresponding first-guess and true profiles. Each new retrieval is written immediately after it has been computed.
- **`res_emiss.dat`** contains the retrieved emissivity delivered by the same 1D-Var process. While the user may not be interested in using those results, they've been stored apart from the other parameters

In order to make things easier, those result files contains the retrievals as well as the corresponding truth, which is therefore an input of the retrieval program `iasi_inversion`. This can be easily canceled in the code to make the software explicitly independent from the true atmospheric state.

3. Test dataset

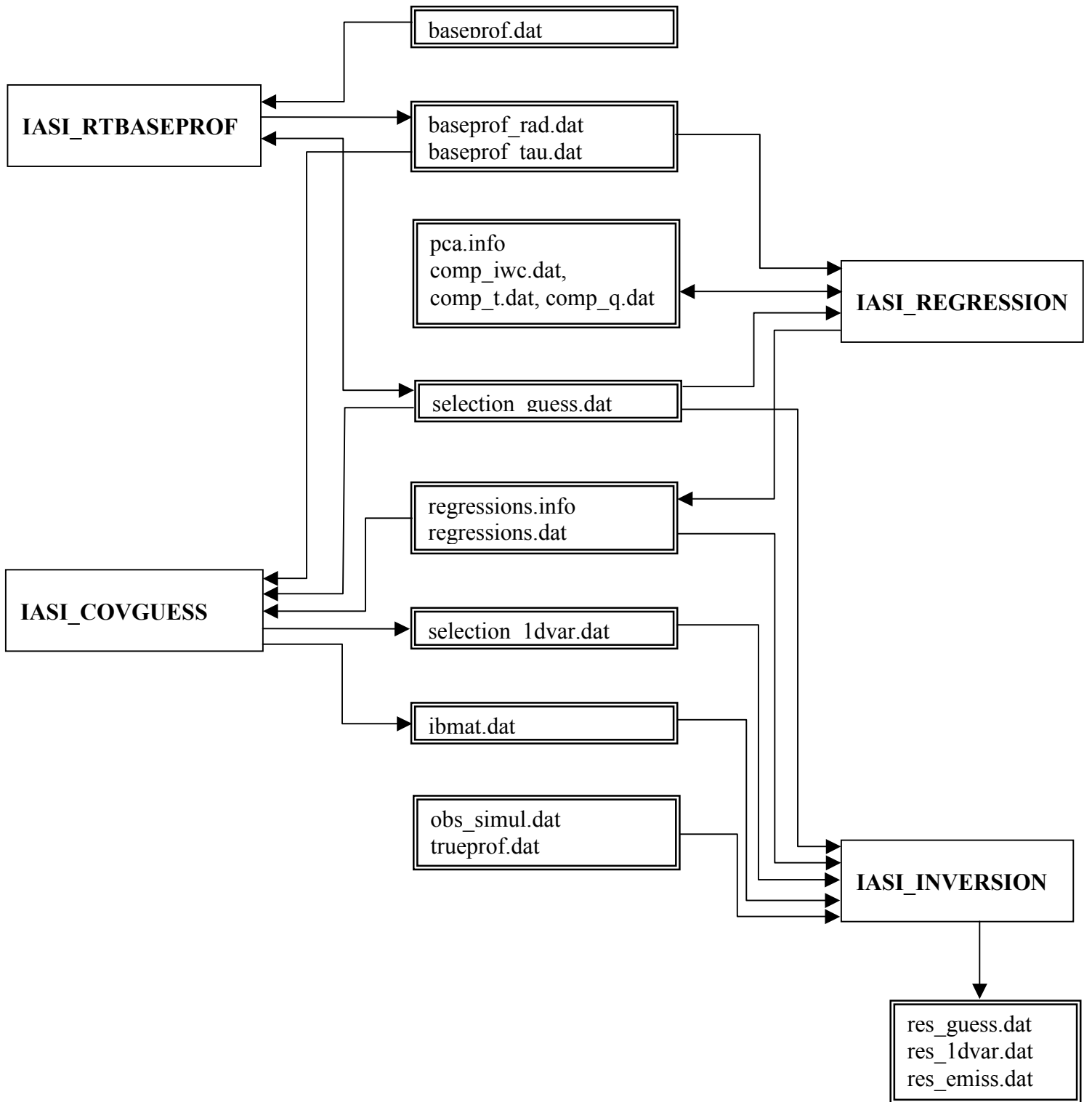
The software is provided with all files present at the end of a complete run. This includes:

- **`iasi_par.txt`**, the running parameters file
- the true profiles file **`trueprof.dat`**, and the corresponding simulated observations **`obs_simul.dat`**. Those 153 profiles were sampled ($10^\circ \times 10^\circ$) in the 18/08/2001 00H Arpege analysis. Their lat-lon range is $100^\circ\text{W}-65^\circ\text{E}$, $0^\circ-80^\circ\text{N}$.
- the first-guess and retrieved profiles files **`res_guess.dat`** and **`res_1dvar.dat`**, and the corresponding retrieved emissivity file **`res_emiss.dat`**.
- a profiles library **`baseprof.dat`**, formed by accumulating profiles sampled in the analyses from the ten previous days ($10 \times 153 = 1530$ profiles), and the corresponding radiances and transmittances files **`baseprof_rad.dat`** and **`baseprof_tau.dat`**.
- channels selection files **`selection_guess.dat`** and **`selection_1dvar.dat`**, computed for the current profiles library.
- all files related to linear regressions performed over this library: **`pca.info`**, **`comp_iwc.dat`**, **`comp_t.dat`**, **`comp_q.dat`**, **`regression.info`**, **`regression.dat`**.
- emissivity models file **`model_emiss.txt`**, filled with four arbitrary example models.

To run the full retrieval process for the same 153 situations. Simply invoke `IASI_INVERSION`. Previous result files **`res_guess.dat`**, **`res_1dvar.dat`**, **`res_emiss.dat`** will then be replaced.

To run the retrieval for the same profiles but with different surface conditions or geometry, first check `iasi_simul.f90`, recompile and run it to update the observations file **`obs_simul.dat`** before invoking `IASI_INVERSION`.

To introduce a new set of situations to be retrieved and a new profiles library, execute successively the four program units, as described in the previous paragraph.



4. Programs brief descriptions

IASI_RTBASEPROF

Task: direct radiative transfer computations over all the user-supplied profile library. If necessary, performs a channel selection for the first-guess computations.

Input:

- **baseprof.dat**, profile library supplied by the user
- **iasi_par.txt**, execution parameters file (modified by the user)
- **selection_guess.dat** (if present), channels selection for first-guess computations
- **transmissions.dat** (if **selection_guess.dat** absent) climatological channels transmissions

Output:

- **baseprof.rad**, upward and downward atmospheric radiances, for all profiles, secant values and surface pressure values.
- **baseprof.tau**, atmospheric transmittances for all profiles, secant values and surface pressure values.
- **selection_guess.dat** (if initially absent), channels selection for first-guess computations

Methods and main subroutines:

The main program loops over the profiles in **baseprof.dat** and the secant values read in **iasi_par.txt**, and writes down the resulting radiances and transmittances in the output binary files.

iasi_rt4psol_lib runs RTIASI for one profile and one secant, and outputs the corresponding radiances and transmissions for all pressure levels specified in **ici_par.txt**. The user-required number of channels is also read in **ici_par.txt**.

selec_rodgers is called prior to any radiative transfer computation when **selection_guess.dat** is not found in the relevant directory. It performs a channel selection following the method proposed by C. Rodgers, and stores the results in **selection_guess.dat**. The resulting set of channels is adapted to the profile library variability.

IASI_REGRESSION

Task: performs regressions computations over all the profiles library (dependant variables) and all the corresponding brightness temperatures (independent variables). Stores the results in the regressions and pca files.

Input:

- **iasi_par.txt**, execution parameters file (modified by the user)
- **baseprof.dat**, profile library supplied by the user
- **baseprof.rad**, upward and downward atmospheric radiances, for all profiles, secant values and surface pressure values.
- **baseprof.tau**, atmospheric transmittances for all profiles, secant values and surface pressure values.
- **selection_guess.dat** (if present), channels selection for first-guess computations
- **model_emiss.txt**, **stat_emiss.dat** surface emissivity models files (through emissivity module)

Output:

- **pca.info**, **comp_iwc.dat**, **comp_t.dat**, **comp_q.dat** files storing de principal components analyses results. They are re-computed only when initially absent in the relevant directory.
- **regression.info** ascii file containing regressions parameters.
- **regression.dat** binary files storing all regression coefficients

Methods and main subroutines

The main program loops over the surface pressure, integrated water-vapor content and secant values specified in **iasi_par.txt**, invoking the following main subroutines. It fills the output files with the final results. Calls to the computation routines follow this scheme:

For each surface pressure value

Brightness temperatures are computed for all secant values (**cal_tbaseprof**).

Pca (if required) and regressions (loop over secants) are performed over iwc

For each iwc class

Pca (if required) and regressions (loop over secants) are performed over temperature and water vapor profiles, and surface temperature.

End for

End for

A (secant, surface pressure) couple of values leads to one record in the **regression.dat** result binary file. Hence, (number of secants)x(number of surface pressures) is the number of regression processes to be performed.

cal_tbaseprof computes the brightness temperatures in the selected channels for all the profiles in the library, one secant and one surface pressure value. Reads the corresponding radiances and transmittances, applies some surface conditions (emissivity and temperature) to complete the forward radiative transfer computations and adds iasi instrumental noise.

regression_iwc computes regression coefficients for the integrated water-vapor content, for all secants and one surface pressure. If any of the principal components storing-files is not found, a PCA is performed over the iwc predictors (including all secants). **Optim_ncomp** is invoked to define the number of components to be retained, and therefore the number of coefficients to be computed. However, those coefficients are subsequently expressed in the predictor's frame.

regression computes regression coefficients relative to the atmospheric temperature and water-vapor profiles and to the surface temperature, for all secants, one surface pressure, and one iwc class. PCA and regressions are computed following the same scheme than in `regression_iwc`.

optim_ncomp evaluates the optimal number of principal components to be taken into account after each PCA computation. Predictors are reconstructed with an increasing number of principal components until the difference (rms averaged over all the profiles) between reconstructed and true predictors reaches a minimum, given the noise that has been added to the observations.

IASI_COVGUESS

Task: computes and stores the first-guess error covariance matrices over the whole profile library. If necessary, performs a channel selection for the 1D-Var computations.

Input:

- **iasi_par.txt**, execution parameters file (modified by the user)
- **baseprof.dat**, profile library supplied by the user
- **baseprof.rad**, upward and downward atmospheric radiances, for all profiles, secant values and surface pressure values.
- **baseprof.tau**, atmospheric transmittances for all profiles, secant values and surface pressure values.
- **selection_guess.dat**, channels selection for first-guess computations
- **regression.info**, **regression.dat** regressions parameters and coefficients (through `iasi_guess` module)
- **model_emiss.txt**, **stat_emiss.dat** surface emissivity models files (through emissivity module)

Output:

- **ibmat.dat**, binary file containing the inverses of first-guess error covariance matrices
- **selection_1dvar.dat**, channel selection file for the 1D-Var computations.
- **selection_guess.dat** (if initially absent), channels selection for first-guess computations

Methods and main subroutines

This program basically applies the pre-computed regressions to obtain the first-guess profiles corresponding to the profile library itself. The resulting covariance matrices are estimates of the regressions accuracy, over the set of profiles which has been used to compute them.

The first-guess and covariance computations loop over the surface pressure and the iwc, each couple of value leading to one covariance matrix (all secant values are mixed up). The inverse matrices are computed, and stored in the **ibmat.dat** binary file. When **selection_1dvar.dat** is not found in the relevant directory, a channel selection procedure is finally invoked.

cal_tbbaseprof computes the brightness temperatures in the selected channels for one profile from the library, and one surface pressure value. Reads the corresponding radiances and transmittances, applies some surface conditions (emissivity and temperature) to complete the forward radiative transfer computations and adds `iasi` instrumental noise.

iasi_cal_guess is the first-guess profile computation subroutine, part of the `iasi_guess` module, also invoked by `IASI_INVERSION`.

cal_bmat computes a covariance matrix from a set of library profiles and the corresponding first-guess profiles. The boundaries of the iwc range to be considered are parameters of the subroutine.

For control purpose, the true and guess profiles are written in a temporary file, **cov_tmp.dat** , and then read by the present subroutine.

selec_rodgers performs a channel selection according to the iterative method proposed by C. Rodgers. It is invoked only when **selection_1dvar.dat** is not found in the relevant directory. The resulting set of channels is adapted to the first-guess profile error covariance.

IASI_INVERSION

Task: computes first guess profiles from the user-supplied brightness temperatures, and then performs a 1D-Var retrieval of atmospheric temperature and water-vapor, as well as surface temperature and emissivity.

Input:

- **iasi_par.txt**, execution parameters file (modified by the user)
- **selection_guess.dat** , channels selection for the first-guess computations
- **selection_1dvar.dat**, channels selection for the 1D-Var retrieval
- **regression.info**, **regression.dat** regressions parameters and coefficients (through **iasi_guess** module)
- **ibmat.dat**, binary file containing the inverses of first-guess error covariance matrices
- **model_emiss.txt**, **stat_emiss.dat** surface emissivity models files (through emissivity module)

Output:

- **res_guess.dat** first guess profiles, stored together with true profiles
- **res_1dvar.dat** 1D-Var retrieved profiles, stored together with first guess and true profiles
- **res_emiss.dat** true, first-guess and 1D-Var retrieved surface emissivity

Methods and main subroutines

The programs reads the user-supplied brightness temperatures and related parameters (geometry, surface type...) and then computes all the corresponding first guess profiles, which are stored in **res_guess.dat**. Therefore, those results can be used prior to subsequent retrieval time-demanding computations. Then, the 1D-Dvar algorithm loops over all input situations and leads to improved estimates of atmospheric variables (temperature and water-vapor profiles) and surface variables (temperature, emissivity).

The core of the program is the 1D-Var iterative algorithm, which can be parameterized by the user checking **iasi_par.txt**.

iasi_cal_guess is the first-guess profile computation subroutine, part of the **iasi_guess** module, also invoked by **IASI_COVGUESS**.

tb_ascii2bin stores the brightness temperatures supplied by the user in a temporary binary file.

damping is a function used to artificially ensure that the matrix to be inverted during the 1D-Var process has no severely small eigenvalues. This is necessary to limit oscillations when iterations are required.

IASI_SIMUL

Task: simulate brightness temperatures for required iasi channels.

Input:

- **trueprof.dat**, user-supplied “true profiles” file
- **iasi_par.txt**, execution parameters file (modified by the user)
- **model_emiss.txt**, **stat_emiss.dat** surface emissivity models files (through emissivity module)
- **selection_guess.dat** , channels selection for the first-guess computations
- **selection_1dvar.dat**, channels selection for the 1D-Var retrieval

Output:

- **obs_simul.dat**, simulated observations file

method: forward radiative transfer computations achieved by calls to RTIASI.

Comment: this is mainly a exemple file, and should be adapted by the user. Particularly, conditions for each simulated situations (surface temperature, surface pressure, surface emissivity type, satellite zenith angle) are defined in the code itself, and should be adapted to users data

Methods

Surface conditions (pressure, temperature, emissivity) and geometry (satellite and solar zenith angles) are defined for each input profile. RTIASI is then invoked for each situation, and the resulting brightness temperatures and emissivity are stored in the output ascii file. Forward radiative transfer computations are performed only for channels selected either for first-guess extraction or for the 1D-Var retrieval.

Here, surface and geometry parameters are defined in the code itself. Therefore, this program should be considered as an example, providing the user with guidelines useful to write his own brightness temperatures simulation, code meeting his own requirements.

5. User-files description

Files that have to be either supplied or modified by the user are described in this section, together with some relevant precision on the programs operations.

baseprof.dat, profile library supplied by the user

format:

ascii file

```
1530  number of profiles in the library (integer)
0.23183E+03  0.25614E+03  ...  profile #1 (3NLEV+9 real values, free format)
0.22949E+03  0.25350E+03  ...  profile #2
...
0.22280E+03  0.24684E+03  ...  profile #1530
```

comments:

The profile library is used as the unique database for all first guess regressions. It can be either a climatological, static, set of representative profiles, or a “rolling” library, supplied with recently analyzed profiles. The accuracy and, moreover, the bias of the final first guess and retrieved profiles largely depends on the relevancy of this library.

This file is an input to IASI_RTBASEPROF (radiative transfer computations), IASI_REGRESSION (regression coefficients computations) and IASI_COVGUESS (computation of background error covariance).

trueprof.dat, true profiles to be retrieved

format:

identical to **baseprof.dat** (see above)

comments:

This file stores the profiles used to simulate iasi observations (brightness temperatures). It can be used as in input to IASI_SIMUL, an interface to RTIASI, to generate the observation file **obs_simul.dat** (see hereafter).

However, in the current release, it is also used as an input to IASI_INVERSION, so that the final result files (**res_guess.dat** and **res_1dvar.dat**) already contains the retrieval and the corresponding true state. This could be easily canceled in the code.

obs_simul.dat, simulated iasi observations

format:

ascii file. For each situation:

```
40.0E+0 100.0E+0  lat, lon (real free format)
2 3 1013.0 276.0659  surf. type, emiss. type, surf. pressure, surf. temperature (free format)
100.0 9.76  solzen, satzen (free format)
620  number of simulated channels (free format)
1 232.23428 0.99313354  chan. index, brightness temp., surf. emiss.(free format)
```

7 233.08354 0.99432873
9 234.98093 0.99472713
...

lat, lon : currently not used in programs.

surf. type: 1=sea, 2=land

emiss. type: meaningful only if surf. type=2 (land). Refers to the surface emissivity model used, as defined by the user in **model_emiss.txt** (see hereafter).

surf. temperature: only used to be copied in the result files, together with the first-guess and retrieved surface temperatures.

solzen, satzen: only the satellite zenith angle is currently used. However, the solar zenith angle should be greater than 90°.

chan. index: index of each simulated channel in the iasi spectrum. Only selected channels have to be present. If a selected channel lacks, the retrieval program will stop with a relevant error message.

surf. emiss. : surface emissivity in the corresponding channel. Only used to be stored together with the first-guess and retrieved emissivity in the result file **res_emiss.dat**.

comments

The IASI_SIMUL program generates this file from the profiles stored in **trueprof.dat**. However, the user may have to modify the code in order to conveniently define the relevant parameters: surface temperature, pressure, emissivity type, satellite zenith angle.

Only the required observations are computed by IASI_SIMUL. However, a user-supplied observation file could contain any more channel than those selected for first-guess and 1D-Var computations.

selection_1dvar.dat, selection_guess.dat: channel selection files

format:

ascii files.

Free formats.

selection_1dvar.dat:

500 *total number of selectable channels*
0.00000E+00 *channel #1 is not selectable*
0.33774E+00 *channel #2 is selectable, with a 0.33774 figure of merit*
0.00000E+00
0.00000E+00
0.45241E+00 *channel #5 is selectable, with a 0.45241 figure of merit*
0.00000E+00
...
0.00000E+00 *last iasi channel (#8461) is not selectable*

selection_guess.dat:

270 404 *total number of selectable channels for temp. and water vapor regressions*
0.19757E+01 0.00000E+00 *channel #1 is selectable for temp., not for water vapor*
...

Comments

Channel selection files are normally generated by IASI_RTBASEPROF and IASI_COVGUESS when not found in the relevant directory. However, they also can be supplied by the user to force any other channel selection.

The final number of selected channels can't be determined before the end of the selecting process (see **iasi_par.txt**). A “figure of merit” is computed for each among those channels, which expresses the mean information content it supplies (here, more precisely, its average contribution to the degree of freedom for signal). In the selection files, all channels that have been selected appear through their figure of merit, others appear through zero values. While the first-guess and 1D-Var computations do not necessarily make use of all of them, they are only “selectable” channels. A user-defined number (see **iasi_par.txt**), chosen among them so that they bring the largest information content, are actually selected.

To propose his own channel selections, the user only has to set non-selected channels figures of merit to zero, and others to arbitrary positive values, possibly classified by order of crescent interest. **iasi_par.txt** has to be subsequently modified.

Temperature and water vapor channels appear separately in **selection_guess.exe**, as they have been selected separately. This aims to keep temperature regressions as insensitive as possible to water vapor profile.

emiss_3cm-1.dat, emiss_3cm-1.info: materials emissivity database

format:

ascii files.

emiss_3cm-1.dat: materials ascii spectra at 3cm^{-1} resolution. 706 values for each spectrum, from 645.0 cm^{-1} to 2760 cm^{-1} (real, free format)

emiss_3cm-1.info: index and comments related to each material, sorted in the same order than the corresponding spectra in the previous file.

comments:

All the emissivity data stored in **emiss_3cm-1.dat** are from the **MODIS UCSB Emissivity Library**. This data set was collected by Dr. Zhengming Wan's Group at ICES (Institute for Computational Earth System Science). They are currently available on the following website: <http://www.ices.ucsb.edu/modis/EMIS/html/em.html>

All the spectra, excepted those of “manmade materials”, have been interpolated at 3cm^{-1} resolution over iasi wavenumber range, and the corresponding sample identification comments stored in **emiss_3cm-1.info**. This is the file to check whenever the user wants to refer to a particular material (see hereafter **model_emiss.txt**).

Any new spectrum, on the same wavenumber grid, and corresponding comment line can obviously be added.

model_emiss.txt: emissivity models definition

format:

#1	27:oak, 31:pine, 35: grass, 41:soil	<i>comment line</i>
4		<i>number of materials to be mixed</i>
27	31 35 41	<i>materials indexes</i>
30.	30. 20. 20.	<i>materials average coverage fraction (%)</i>

10. 10. 10. 10.

maximum departures from average fractions

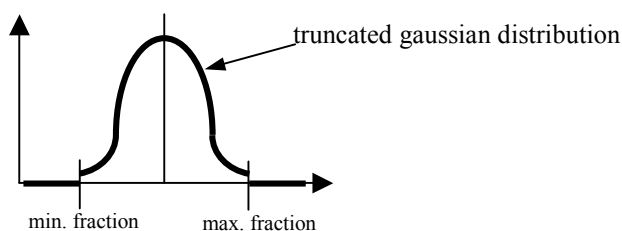
Comments:

The purpose of this file is to store the parameters of user-defined surface emissivity models. Each model is defined by:

- a set of materials (up to ten) whose spectra are found in **emiss_3cm-1.dat**
- their average fractions in the iasi field of view
- the maximum departure from the average

Those data are inputs to emissivity-simulation routines (iasi_lib.f90, emissivity module). They are used to produce synthetic random spectra for each of the user-supplied model, and corresponding covariance in the emissivity bands implied in the 1D-Var retrieval.

The random coverage fractions are computed so that their sum is 100%, with the following truncated gaussian density probability:



Roughly, the fraction range boundaries stand for the classification criteria within a given emissivity model, and the gaussian distribution accounts for the natural variability within this model.

The emissivity models statistics are computed and stored in a binary file, **stat_emiss.dat**. Note those data are recomputed only when not found in the relevant directory. However, as statistics and models must be coherent, **stat_emiss.dat must be destroyed after any change in model_emiss.txt.**

iasi_par.txt: general execution parameters

format:

example of a full parameters file:

all data in free format.

```
*****1D-VAR PARAMETERS*****
# Minimum iterations number nitermin=
1
# Maximum iteration number nitermax=
10
# Jacobian update frequency
4
# Nb. of channels requested for 1d-var retrieval
500
# Nb. of selected channels per profile in selection process
200
# Nb. of emissivity types to be looped-over during selection
3
# indices of emissivity types to be looped-over
```

```

1 2 3
*****REGRESSION PARAMETERS*****
# number of secants
6
#secant values
1.0 1.1 1.2 1.3 1.4 1.5
#number of surface pressures
5
#surface pressure values
1030.0 1013.25 900.0 800.0 700.00
# number of iwc classes
5
# iwc classes boundaries (previous number+1)
0. 1. 2. 3. 4. 8.
# Nb. of channels requested for temperature and moisture regressions
200 100
# Nb. of selected channels for temperature and moisture per profile
200 200
*****PROFILE LIBRARY PARAMETERS*****
# number of secants
6
#secant values
1.0 1.1 1.2 1.3 1.4 1.5
#number of surface pressures
5
#surface pressure values
1030.0 1013.25 900.0 800.0 700.00
#secant values
1.0 1.1 1.2 1.3 1.4 1.5
#number of surface pressures
5
#surface pressure values
1030.0 1013.25 900.0 800.0 700.00

```

comments

Those user-defined parameters have effects on the accuracy and the computation time of the corresponding programs. They should be set empirically, in accordance with the user actual requirements and computing resources.

1D-VAR PARAMETERS

Minimum iterations number, maximum iteration number :

defines the min and max number of iterations to be performed during the non-linear 1D-Var retrieval.

Jacobian update frequency:

Allows the jacobian not to be recomputed at each 1D-Var iteration. For a N update frequency, atmospheric state and its jacobian are recomputed each $i^{\text{th}} = kN+1$ ($k=0,1,2,\dots$) iteration. Other iterations only update the atmospheric states.

Nb. of channels requested for 1d-var retrieval

This is the number of channel to be actually used for 1D-Var retrieval, if enough preselected channels are available in **selection_1dvar.dat**. Otherwise, all preselected channels are taken into account.

Nb. of selected channels per profile in selection process

This channels number parameterizes the channel selection performed by IASI_COVGUESS when **selection_1dvar.dat** is not found. Due to the fact that this process accumulates channels selected for several atmospheric and surface conditions, the final size of the selection can't be anticipated. Thus, only the number of channels to be selected through each individual selection process can be exactly defined. Typically, a 200 channels/profile requirement can lead to more or less 450 selectable channels stored in **selection_1dvar**.

Nb. of emissivity types to be looped-over during selection

The selection process loops over a set of surface emissivity types, for each given profiles. This parameter provides the number of emissivity types to be considered, which must be smaller or equal to the total number of surface emissivity models defined by the user (see **model_emiss.txt**).

indices of emissivity types to be looped-over

On this line are written the corresponding emissivity models indices. While each added model significantly increases the channel selection computation time, it is advisable to choose a small number of representative and contrasted surface types.

REGRESSION PARAMETERS

number of secants, secant values:

Regression coefficients are computed for a discrete set of secant values. For a first guess computation, the real secant is categorized as it's nearest available value.

number of surface pressures, surface pressure values:

Surface pressure values, following the same principle than secant values.

number of iwc classes, iwc classes boundaries (previous number+1)

All regression coefficients computations relates to a given range of the integrated water vapor content. Those variables define the number of iwc classes considered, and the corresponding range boundaries (g/cm^2)

Nb. of channels requested for temperature and moisture regressions

Those are the number of channels to be actually used , in temperature and water vapor regression computations respectively. They are automatically limited to the available amount of preselected channels in **selection_guess.dat**.

Nb. of selected channels for temperature and moisture per profile

As for the **1d_var** channels selection, those numbers specify how many channels would be selected for each individual profile by IASI_RTBASEPROF, if **selection_guess** was not found.

PROFILE LIBRARY PARAMETERS

number of secants, secant values, number of surface pressures, surface pressure values

As the analogous parameters controlling the required number of regressions, those controls the required number of radiative transfer computations over the profile library. The surface pressure and secant sampling must at least cover the corresponding regression sampling. Practically, they may be chosen to be equal.

res_guess.dat: first-guess profiles file

format:

ascii file

number of stored guess profiles

profile nb, latitude, longitude, true iwc, iwc from firts regression, true surf. temp.,...
... first guess surf. temp.

pressure at level 1 | true temp. | true wv | first guess temp | first guess wv |

...

pressure at level nlev | true temp. | true wv | first guess temp | first guess wv |
latitude, longitude

...

comments

This is an intermediate results file, which only contains the first-guess regression-computed atmospheric temperature and water-vapor parameters, together with there corresponding true values. It's written as soon as the first-guess computations ends, before the computation-time demanding 1D-Var retrieval.

The first line contains, in addition to profile number, latitude and longitude (0.,0. if not provided by the user), the true integrated water vapor content, the corresponding regression-computed iwc, the true and first-guess surface temperature. Note that the regression-computed iwc differs from the iwc computed from the temperature and water vapor first-guess profiles. The former value is only used to classify the situation, and to select the regression coefficients set to be applied.

When there is a too large discrepancy between the regression-computed iwc and the first-guess iwc, the profile is rejected. It will neither appear in **res_guess.dat**, nor be treated at the 1D-Var stage. For this reason, both **res_guess.dat** and **res_1dvar.dat** files generally contain fewer profiles than provided by the user, which can be identified thanks to the profile number, referring to their position in the user-supplied dataset.

res_1dvar.dat: final profiles result file

format:

ascii file

profile nb, lat, lon, average residual, max. residual, final cost function value, ...

... true surf. temp., guess. surf. temp., retrieved surf. temp.

pressure at level 1 | true temp. | true wv | guess temp | guess wv | ret. Temp. | ret. wv |

...

pressure at level nlev | true temp. | true wv | guess temp | guess wv | ret. Temp. | ret. wv |

...

comments

This is the file where all retrieval results are stored, except retrieved emissivity. All situation that have successfully passed the first-guess step appear here (identified by their profile number in the user-supplied dataset), whatever the 1D-Var retrieval quality. This quality can be checked by the user by looking to the following values, stored in the first-line of each profile results:

Average residuals: rms of the brightness temperature departures normalized by the noise.
Maximum residual: maximum absolute value of the brightness temperature departures normalized by the noise.
Final cost function value: final value of the cost-function, after all 1D-Var iterations.

In order to filter the final retrievals, a threshold may be applied to one or more among those quality-control values. Typically, 5% to 20% retrieved profiles should be rejected because of convergence failure, depending on the quality of the profile library, the considered surface type etc...

res_emiss.dat: final profiles result file

format:
ascii file

6 number of spectral bands
745.25 1145.25 1345.25 1445.25 1945.25 2045.25 2145.25 *bands boundaries (cm⁻¹)*
3 surface type (emissivity model) for the situation #1
0.98089 0.97798 0.93081 0.90893 0.85539 0.84145 *true emissivity in spectral bands*
0.98556 0.97571 0.92351 0.90341 0.86109 0.86200 *first guess emissivity*
0.98712 0.97872 0.93283 0.89583 0.85769 0.85217 *retrieved emissivity*
2 surface type (emissivity model) for the situation #2
...

comments

While retrieved emissivity may not be considered as a useful output, it is stored apart from profiles results, in the present file, together with true and first-guess emissivity. Only the average emissivity over several spectral bands is retrieved. Those bands number and boundaries depends on the channels selected for the 1D-Var analysis: the spectrum is divided in bands with a constant width (currently 100 cm⁻¹, but this can be easily changed in the code), but only bands where window channels lie are considered for the retrieval. The adjacent bands where the spectrum is unaffected by the surface are given the same emissivity value (as the real value is not needed). The bands parameters (number and boundaries), are the same for all retrieved situations, and stored in the two first lines of **res_emiss.dat**.

For each retrieved situation, the surface emissivity model number referring to the model file **model_emiss.txt**, and the true, first-guess and retrieved emissivity in all selected spectral bands are stored. Note that the first-guess emissivity are the average values corresponding to the selected emissivity model. The emissivity error covariance is then the model covariance itself, that is that no surface misclassification is expected.

