SFIT Retrieval process at BIRA-IASB

BIRA·IASB

aeronomie·be

Setup: from measurement to HDF in 3 steps

∗ SFIT2_x: retrieval program

∗ SFIT2_e: error computation

∗ SFIT2_h: HDF creation

All programs are compiled matlab scripts

SFIT2_{x,e,h} <instrument name> <strategy file> ...

SFIT2_x

Inputs are

∗ a list of spectra

∗ zpt from NCEP (zpt data is available 6hourly)

∗ apriori profiles (e.g. WACCM)

∗ template binput file

∗ a strategy file: this lists the different retrieval steps that should be
   performed on the spectra (e.g. first retrieve H2O, update apriori profile,
   retrieve CH4,...)

Strategy file

Strat file defines retrieval steps
In chronological order:

NCEP>.... = copy NCEP water profiles to these molecules in apriori profiles (labelled ncep)

H2O<ncep= retrieve H2O using ncep apriori profiles

...

h2o=ncep .. = update apriori profile with H2O retrieval results (now with label h2o)

...

CH4!<h2o = retrieve CH4 using h2o apriori profiles (!=for these molecules we wish to create HDF)

R/retrievals/working/bavol/sfit2/strategies/ch4

stratfinal.txt – KWrite

File   Edit   View   Tools   Settings

New   Open   Save   Save As   Clo

```
 1  ch4n2ofinal: hh hl
 2  ncep=ALLWACCM.v6
 3  NCEP>H2O HDO H217O H218O
 4  H2O<ncep
 5  HDO<ncep
 6  h2o=ncep HDO H2O
 7  H2O>H217O H218O
 8  CH4!<h2o
 9  N2O!<h2o
10  /bira-iasb/projects/FTIR/
```

Optimization for SNR factor

SFIT2$_x$ can be run in 'optimisation mode': retrievals are performed 2x

After first run the contribution matrix is used to estimate the optimal SNR factor: optimal means

* minimal total error on total column of target molecule (L-curve)

* minimal mean error on retrieval profile of target molecule (L-curve)

* profile contains no negative vmr's

* profile does not oscillate more than apriori profile of target molecule

Second run is done with this optimal SNR factor.   per spectrum!
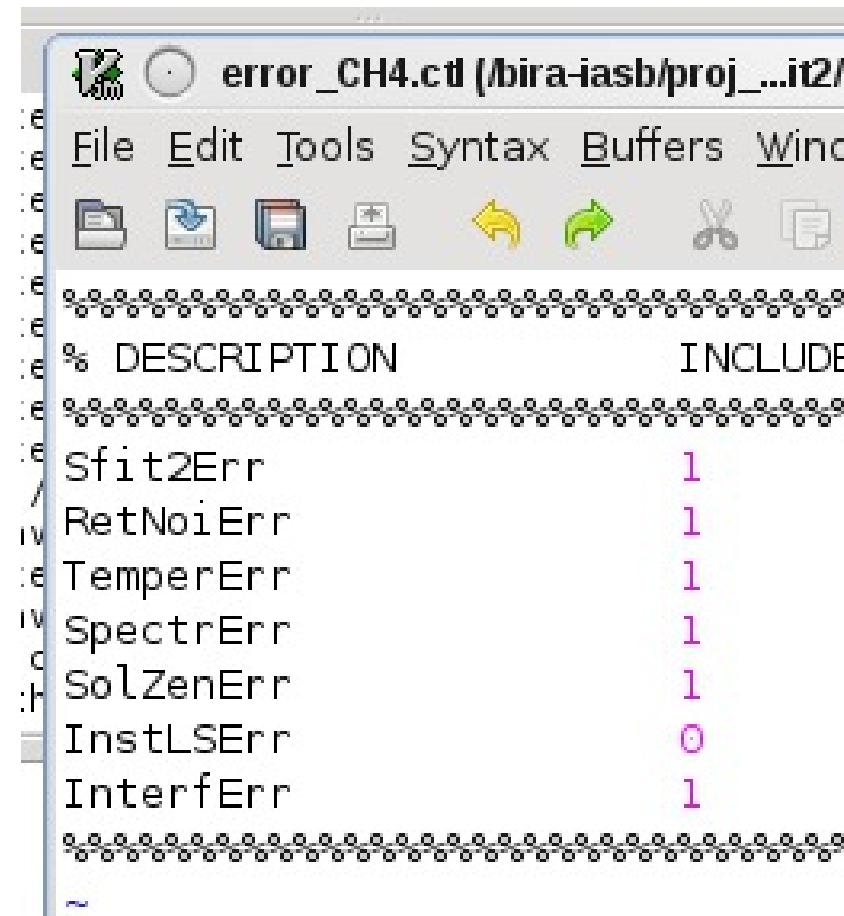
SFIT2_e

After retrieval, filter the results on

∗ converge flag is 'T' in sfit2

∗ rms values $<$ threshold value

∗ DOF $>$ threshold value

Purpose: we do error computation per spectrum so we wish to avoid "bad" spectra in this process

On the filtered lists, the error analysis is executed:

1. Error on spectrum (noise)

2. Temperature (pt file from fastcode)

3. Solar Zenith Angle (parameter in binput file)

4. Spectroscopic Database error

5. Apriori profile of interfering molecule (e.g. H2O)

6. Instrument line shape (apodization from linefit)

error_CH4.ctl (/bira-iasb/proj_...it2/

File Edit Tools Syntax Buffers Wind

```
% DESCRIPTION              INCLUDE
Sfit2Err                   1
RetNoiErr                  1
TemperErr                  1
SpectrErr                  1
SolZenErr                  1
InstLSErr                  0
InterfErr                  1
```

All errors (except measurement noise) are calculated by

∗ perturbation of the above parameters

∗ call sfit2 in forward model mode $\mathbf{F}$ (no retrieval, i.e. calculate the synthetic spectrum for the perturbed parameters)

∗ construct K matrix $K_{\mathbf{b}} = \partial_{\mathbf{b}}\mathbf{F} \approx \dfrac{\mathbf{F}(\text{perturbed}) - \mathbf{F}(\text{retrieved})}{\text{perturbation}}$

($\dim K$ =spectrum×parameter)

∗ transfer covariance matrices from
parameter space $\rightarrow$ spectrum space (using $\mathbf{K}_{\mathbf{b}}$)
$\rightarrow$ profile space (using contribution function $\mathbf{G}$)

1. Noise on spectrum is calculated from the spectrum directly

2. Temperature: if n is the number of layers in the retrieval, sfit2 is called n times in forward model mode

3. SZA 1 perturbation

4. Spectroscopic database

   * The HITRAN database is used to get the variance on the line intensity and broadening factors.
   * Only lines with intensity above certain threshold are taken into account ($\rightarrow$ allows to control the number of lines or the number of calls to sfit2)
   * Lines for isotopes can also be taken into account

5. Interfering molecule: per molecule n calls to sfit2.

6. Instrument line shape: perturbs the apodizaton and phase values.

SFIT2_h

This routine

* ∗ gets 'SURFACE.TEMPERATURE' and 'SURFACE.PRESSURE' from meteo files (contain information from local weather station: measures rain, wind, t, p,...)

* ∗ groups errors contributions in SYSTEMATIC or RANDOM variables

* ∗ creates HDF files (splits the list of retrievals by year)

* ∗ optional: add RD flag, version number (for NDACC database)

A Python script for automization

A Python script

* Determines the recent measurements (not yet retrieved for the given strategy)

* Divide the list of new spectra in sublists (for parallel retrievals, up to 60 lists can be processed simultaneously)

* Call SFIT2_x, SFIT2_e, SFIT2_h for each sublist: distribute these jobs over available compute servers

* FTP HDF file(s) to NDACC

* Does administration: what spectra were used for which HDF files

* Generates plots for RMS, DOF, mean errors, . . .

Filtering on spectra quality before retrievals

∗ Variance in solar tracking

∗ Variance in DC signal

∗ Variance in solar intensity

Filtering on spectra before NDACC submission

∗ Extended error budget calculations: total error on total column (systematic + random) should not exceed threshold