

# FLEXINVERT User's Guide

## **FLEXINVERT Version 3.0**

Date: January 2016

R. L. Thompson  
Norwegian Institute for Air Research (NILU)

## **Introduction**

FLEXINVERT is an open source code developed for calculating land/ocean surface to atmosphere fluxes of any species that is inert (or if it has chemical reactions that these are linear to a good approximation) and has measurable atmospheric concentrations. There may be other practical considerations that restrict its use, but it is up to the user to consider these restrictions and to decide whether FLEXINVERT may be useful to solve it or not. FLEXINVERT is not meant to be used as a black box, but rather requires the user to carefully consider the problem they are addressing by choosing appropriate prior information, observations and uncertainty estimates, and by evaluating the output.

This code was developed principally for use with greenhouse gases, e.g. methane and halocarbon species, and the terminology used in the code and this documentation is based on that commonly used in the greenhouse gas modelling community. We refer to the 'state vector' as the vector of the parameters to optimize, such as the surface-atmosphere fluxes, and the 'observation vector' as the vector of atmospheric measurements, that is, concentrations measured at a given point or 'receptor'. We also use the term 'background', by which we mean the atmospheric concentration that would be measured at a receptor if there were no surface-atmosphere fluxes in the domain of interest. The 'background' can also be thought of as the historical concentration that is not directly accounted for by the transport model. The model relating the state vector parameters to the observations is referred to as the 'kernel' matrix and should not be confused with 'averaging kernel'. The kernel matrix is built from so-called 'source-receptor relationships' (SRRs), which relate the change in flux in a given grid cell to the change in concentration at a receptor. The SRRs used in FLEXINVERT are calculated using the Lagrangian Particle Dispersion Model (LPDM), FLEXPART, but with some modifications FLEXINVERT could also be used with SRRs from any other LPDM.

The philosophy behind FLEXINVERT was to create an inversion framework that is adaptable to many different problems, coping with different species, domain sizes etc. A caveat of this is that it is not tailored to specific requirements of some species. For example, FLEXINVERT in its current form cannot solve for CO<sub>2</sub> fluxes, as this requires resolving the diurnal cycle in fluxes. However, the code can be developed to treat such specific cases, and a CO<sub>2</sub> version of FLEXINVERT is planned in 2016.

The aim of this guide is to give the user an overview of the code, the input data requirements and directions on how to use it. For more details refer to Thompson and Stohl, 2014. If you use FLEXINVERT in a publication, please also cite Thompson and Stohl, 2014.

# 1. Theoretical aspects

## 1.1 Forward model

For a species with no chemistry (or chemistry that can be described as linear) the change in its concentration at any number of receptors,  $\mathbf{y}^{\text{nest}}$  can be related to its surface-atmosphere fluxes,  $\mathbf{f}^{\text{nest}}$  in a nested domain (i.e. nested in the global domain) by the kernel matrix,  $\mathbf{H}^{\text{nest}}$ :

$$\mathbf{y}^{\text{nest}} = \mathbf{H}^{\text{nest}} \mathbf{f}^{\text{nest}} \quad (1)$$

The absolute concentration at the receptors,  $\mathbf{y}^{\text{mod}}$  is then the change in concentration,  $\mathbf{y}^{\text{nest}}$  plus the background concentration, i.e., the contribution to the concentration from fluxes outside the domain,  $\mathbf{H}^{\text{out}} \mathbf{f}^{\text{out}}$  and the ‘historical’ concentration,  $\mathbf{H}^{\text{bg}} \mathbf{y}^{\text{bg}}$  (the ‘historical’ concentration,  $\mathbf{y}^{\text{bg}}$  is that prior to the start of the model run, e.g., if the LPDM is used to run 10-day back trajectories, it is the concentration prior to these 10 days):

$$\mathbf{y}^{\text{mod}} = \mathbf{H}^{\text{nest}} \mathbf{f}^{\text{nest}} + \mathbf{H}^{\text{out}} \mathbf{f}^{\text{out}} + \mathbf{H}^{\text{bg}} \mathbf{y}^{\text{bg}} \quad (2)$$

The matrices  $\mathbf{H}^{\text{nest}}$  and  $\mathbf{H}^{\text{out}}$  contain the SRRs for inside and outside the nested domain, respectively (they do not overlap), and are calculated by the LPDM (FLEXINVERT uses FLEXPART output by default). The matrix  $\mathbf{H}^{\text{bg}}$  contains the sensitivity to the ‘historical’ concentrations (i.e.,  $\mathbf{y}^{\text{bg}}$ ) and is a weighted averaging operator.  $\mathbf{H}^{\text{bg}}$  is also derived from the LPDM (in FLEXPART this is built-in). The concentrations,  $\mathbf{y}^{\text{bg}}$  can be taken from a global model or a 3D field of interpolated measured concentrations and must be provided by the user if using option ‘calcbg’ is ‘true’ (see section 2.4). Alternatively, the user can subtract his/her own background estimates from the observations prior to running FLEXINVERT, i.e., the option ‘calcbg’ is ‘false’, in which case  $\mathbf{H}^{\text{out}} \mathbf{f}^{\text{out}} + \mathbf{H}^{\text{bg}} \mathbf{y}^{\text{bg}}$  will be zero.

The fluxes,  $\mathbf{f}^{\text{nest}}$  are interpolated/aggregated from their original resolution onto a regular grid to match the FLEXPART output (e.g.  $1^\circ \times 1^\circ$ ) and have a given temporal resolution (e.g., weekly or monthly). However, this may not be the best spatial resolution at which to optimize the fluxes owing to: 1) computation restrictions on the number of variables in the inversion, and 2) there will likely be regions where there is little constraint from the atmospheric observations and, hence, insufficient information to constrain the fluxes at this spatial resolution. Therefore, the fluxes are aggregated in space onto a new grid of variable resolution (vg = variable grid). The variable grid is defined based on the atmospheric constraint given by the SRRs and optionally convolved with the prior fluxes. It is also possible to choose to aggregate over water bodies and ice. To convert from the regular to the variable grid, a projection operator is used,  $\mathbf{\Gamma}$  (effectively a summation operator of the area weighted fluxes). Thus, the fluxes and kernel matrix on the variable grid are:

$$\mathbf{f}_{\text{vg}}^{\text{nest}} = \mathbf{\Gamma} \mathbf{f}^{\text{nest}} \quad \text{and} \quad \mathbf{H}_{\text{vg}}^{\text{nest}} = \mathbf{H}^{\text{nest}} \mathbf{\Gamma}^T \quad (3)$$

where ‘T’ indicates the matrix transpose. The concentration at the receptors is then:

$$\mathbf{y}^{\text{mod}} = \mathbf{H}_{\text{vg}}^{\text{nest}} \mathbf{f}_{\text{vg}}^{\text{nest}} + \mathbf{H}^{\text{out}} \mathbf{f}^{\text{out}} + \mathbf{H}^{\text{bg}} \mathbf{y}^{\text{bg}} + \boldsymbol{\varepsilon}^{\text{agg}} \quad (4)$$

where  $\boldsymbol{\varepsilon}^{\text{agg}}$  is the model representation error from having reduced the resolution of the model and is calculated in FLEXINVERT (for details see Thompson and Stohl, 2014).

In the case that the background concentrations are also optimized (this is set by the option ‘optimizbg’ and can only be used if ‘calcbg’ is ‘true’), then the contributions from fluxes outside the nested domain ( $\mathbf{H}^{\text{out}} \mathbf{f}^{\text{out}}$ ) and from the historical concentrations ( $\mathbf{H}^{\text{bg}} \mathbf{y}^{\text{bg}}$ ) are combined:

$$\mathbf{M}_{cg} = (\mathbf{H}^{out} \circ \mathbf{F}^{out} + \mathbf{H}^{bg} \circ \mathbf{Y}^{bg}) \Gamma_{cg}^T \quad (5)$$

where  $\circ$  indicates the Hadamard matrix product,  $\mathbf{F}^{out}$  has rows of  $(\mathbf{f}^{out})^T$ , and  $\mathbf{Y}^{bg}$  has rows of  $(\mathbf{y}^{bg})^T$ . The matrix,  $\Gamma_{cg}$ , is a projection operator from the regular grid to a coarse grid (note  $\Gamma_{cg} \neq \Gamma$ ). The matrix  $\mathbf{M}_{cg}$  has rows corresponding to each observation (the sum of each row gives the background concentration for each observation) and columns corresponding to the coarse resolution grid on which the background is optimized. A new kernel matrix,  $\mathbf{H}$  which includes the background influence, and state vector,  $\mathbf{x}$ , which includes the fluxes in the nested domain as well as scalars of the background influence, can be defined as follows:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{vg}^{nest} \mathbf{M}_{cg} \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} \mathbf{f}_{vg}^{nest} \mathbf{a}_{cg} \end{bmatrix} \quad (6)$$

where  $\mathbf{a}_{cg}$  are the background scalars and have a prior value of 1.

In the case that the background is not optimized, the background concentrations ( $\mathbf{H}^{out} \mathbf{f}^{out} + \mathbf{H}^{bg} \mathbf{y}^{bg}$ ) are subtracted from the observed concentrations and  $\mathbf{H} = \mathbf{H}^{nest}$  and  $\mathbf{x} = \mathbf{f}_{vg}^{nest}$ .

## 1.2 Atmospheric observations

FLEXINVERT can assimilate three different types of atmospheric observations: 1) discrete measurements at a site, such as flask samples, 2) quasi-continuous measurements at a site, and 3) measurements from a moving platform (discrete or quasi-continuous), such as a ship or an aircraft (currently FLEXINVERT cannot handle satellite data but an update for this is planned in the future). Measurements at stationary receptors (i.e., sites) are treated slightly differently from those from moving receptors (i.e, ships and aircraft). For stationary receptors, back trajectories can be run at any given frequency (e.g. hourly) regardless of the observation timestamp (note that in this case, the same frequency should be used for all stationary receptors). The user can set the temporal resolution at which the observations are assimilated (any multiple of the trajectory frequency) and the observations and corresponding trajectories are averaged to this frequency. For moving receptors, it is assumed that back trajectories have been run for each observation, and no averaging is applied since each observation is made from a different location. If there are very many observations from the moving receptors (order of 10000) then it may be necessary to pre-average the observations to an appropriate spatio-temporal resolution. FLEXINVERT recognises the stationary or moving receptor and matches it to the corresponding trajectories by using a 7-character identifier (RECNAME). A list of receptors with their measurement type (FM for Flask Measurement, CM for Continuous Measurement and SM/AM for Ship/Aircraft Measurement) is required as input to FLEXINVERT (see sections 2.1, 2.3 and 2.6).

## 1.3 Bayesian inversion

Based on Bayes' theorem, the most probable solution for  $\mathbf{x}$  is the one that minimises the difference between the observed and modelled mixing ratios while also depending on the prior state variables,  $\mathbf{x}_b$  and their uncertainties (for details on Bayes' theorem see e.g. Tarantola 2005). Assuming that the uncertainties have a Gaussian probability density function this can be described by the cost function:

$$J(\mathbf{x}) = \frac{1}{2} (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + \frac{1}{2} (\mathbf{H}\mathbf{x} - \mathbf{y}^{obs})^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{x} - \mathbf{y}^{obs}) \quad (7)$$

where  $\mathbf{B}$  is the prior flux error covariance matrix,  $\mathbf{R}$  is the observation error covariance matrix, and  $\mathbf{y}^{\text{obs}}$  is the vector of observed concentrations. FLEXINVERT finds the optimal  $\mathbf{x}$  using an analytical solution, i.e., finding  $\mathbf{x}$  for which the first order derivative of  $J(\mathbf{x})$  is zero.

The choice of uncertainty for the prior fluxes and of the observations, used to calculate  $\mathbf{B}$  and  $\mathbf{R}$ , respectively, will strongly influence the solution, therefore, the user needs to carefully consider the uncertainties for the problem at hand. There are two options for defining the prior flux uncertainty. Either the user can supply these as a gridded 2D field in a NetCDF file (this is the option ‘precalcerror’ is ‘true’) or they can be calculated within FLEXINVERT. In the latter case, FLEXINVERT calculates the uncertainty for each grid cell as a fraction of the maximum prior flux estimate in that cell and the 8 neighbouring cells (the fraction is set by the user, see ‘control.def’). This is done in order to smooth the prior uncertainty fields and allow some more degrees of freedom for departure from the fine spatial structure of the prior fluxes. In addition, minimum and maximum limits for the prior flux uncertainty are given by the user. The full error covariance matrix is then calculated as:

$$\mathbf{B} = \mathbf{C}_T \otimes (\mathbf{C}_S \circ \boldsymbol{\sigma} \boldsymbol{\sigma}^T) \quad (8)$$

where  $\mathbf{C}_T$  and  $\mathbf{C}_S$  are the temporal and spatial correlation matrices, respectively,  $\boldsymbol{\sigma}$  is the vector of uncertainties, and  $\otimes$  the Kronecker product. The elements of  $\mathbf{C}_S$  and  $\mathbf{C}_T$  are calculated as an exponential decay over space and time, respectively with a user defined correlation length (one for land, ocean and time, see ‘control.def’). In FLEXINVERT, the spatial and temporal error covariance matrices are stored as Eigen-values and Eigen-vectors where only the Eigen-values (and their corresponding Eigen-vectors) above a certain threshold are kept. The threshold is set as a fraction of the maximum Eigen-value (value of ‘trunc’ in ‘mod\_var.f90’).

The user also defines the uncertainties for the observation and background concentrations (see ‘control.def’). The total error in the observation space is the quadratic sum of the measurement and background errors, as well as, the standard deviation of the observations over the averaging interval, which is an estimate of the temporal aggregation error. Currently no correlations are assumed between observations.

The posterior error covariance matrix is equivalent to the inverse of the second derivative of the cost function  $J''(\mathbf{x})$  and the posterior flux uncertainties are calculated as the square root of the sum of the variance and covariances for each grid cell. Note that although mathematically correct,  $(J''(\mathbf{x}))^{-1}$  tends to underestimate the posterior errors, therefore, we recommend estimating the posterior error from a Monte Carlo ensemble (for details see section 1.8).

#### 1.4 Re-gridding of the optimized fluxes

FLEXINVERT re-grids the optimized fluxes onto the original regular grid. The projection of the fluxes from the variable (and partially coarser) grid to the regular grid is an under-determined problem. FLEXINVERT provides an option of including the information on the spatial distribution of the prior fluxes within each aggregated grid cell in the re-gridding (the option ‘scaleflx’). By default, if ‘scaleflx’ is “1”, then FLEXINVERT will use a simple weighting approach to re-grid the fluxes. There is also an option to use a Bayesian approach to the re-gridding, ‘scaleflx’ is “2”. This is like a second inversion using the posterior fluxes as the ‘observations’ and the prior fluxes as the prior information. It should be noted, however, that this option requires more computational time and that the output should be carefully evaluated. Also, if the observational constraint is weak, the algorithm may fail.

## 1.5 Inequality constraints

In many cases, there is a physical constraint on the sign of the flux, i.e., many species can only have positive surface fluxes, e.g., halocarbons, however, the *pdf* allows for negative fluxes in the solution. To get around this problem, an inequality constraint can be applied, and effectively is the same as truncating the Gaussian distribution at the inequality value (which then becomes an equality constraint). The code for applying an inequality constraint can be downloaded from <http://flexinvert.nilu.no> and is applied to the already optimized fluxes (i.e., after having already run FLEXINVERT). For details see Thacker et al. (2007).

## 1.6 Memory efficient mode

The storage of the transport matrix on the fine grid,  $\mathbf{H}^{\text{nest}}$  and the corresponding error-covariance matrix (stored as Eigen-values and Eigen-vectors) uses a lot of memory. Therefore, for testing or for large problems, it is recommended to run FLEXINVERT in a memory efficient mode in which the fine grid transport matrix and its error covariance matrix are not calculated (this is the option ‘lowmemory’ in ‘control.def’). In this case, there is no calculation of the aggregation error and the redistribution of the posterior fluxes to the fine grid is based on the prior distribution (i.e. ‘scaleflx’ equal to “1” is used).

## 1.7 Restart mode

The “restart” mode (in ‘control.def’) can be used once a FLEXINVERT run has completed and the user wants to re-run the inversion with e.g., different prior fluxes or different uncertainties. In the first run of FLEXINVERT, the transport matrix is calculated according to the releases in the grid\_time files and saved. In the second run, if “restart” is “true” the transport matrix is not recalculated but simply read from the output directory, which saves considerable computational time. Note, if the output directory is different in the second run, the user will need to manually copy the following files from the first run into the output directory of the second run:

```
knest.nc  
kout.nc  
indexobs.txt  
knest_fin.nc (only if not using the memory efficient mode)
```

## 1.8 Testing modes

FLEXINVERT has two testing modes that can be activated in ‘control.def’. The first simply simulates concentrations without performing any inversion and can be used to check how well the prior information reproduces the observations. This is the ‘checkprior’ mode. It is recommended to first run FLEXINVERT in ‘checkprior’ mode before performing an inversion. The second mode is the ‘perturb’ mode and can be used for synthetic data tests. In this mode, the prior information is perturbed with random noise with error characteristics consistent with the prior error covariance matrix. The perturbed prior becomes the new prior and the original prior becomes the “true” fluxes and the concentrations simulated with the “true fluxes” become the observations. This mode can be used to check how well the actual number of observations and transport can constrain the fluxes. Alternatively, this mode can be used by modifying the subroutine ‘calcperturb.f90’ slightly to run a Monte Carlo ensemble to estimate the posterior uncertainty. Although the posterior uncertainty is calculated as the reciprocal of the second order derivative of the cost function, it is recommended to calculate the uncertainty using a Monte Carlo ensemble as the default method tends to underestimate the posterior uncertainty.

## 2. Input data

### 2.1 FLEXPART files

FLEXINVERT requires FLEXPART backwards mode runs to be made for each receptor (for stationary receptors, these can be at any frequency so long as the same frequency is used for all receptors, but for moving receptors these must be according to the timestamp of the observations). FLEXINVERT looks for the output files for each receptor in separate directories named after the receptor. Naming consistency is required for the FLEXPART directories, the receptor names given in the station list file (see e.g. stations.def), and the names of the observation files. FLEXPART output must also be separated by year and month.

Example:

```
/flexpart_output/RECNAME/YYYYMM/
```

Each sub-directory must contain the following files:

```
header_grid_time and/or header
```

These are binary format files containing information about the FLEXPART run, including details of the output grid. Note, if you are using FLEXPART-9.2 (or a more recent version), there is an option in the 'COMMAND' file ('surf\_only') to save only the surface layer to the 'grid\_time' files but retain full vertical resolution in the 'grid\_initial' files. It is strongly recommended to run FLEXPART with this option as this will greatly increase the rate at which the 'grid\_time' files are read and reduce the storage space, while on the other hand, full vertical resolution is needed in the 'grid\_initial' files if calculating the background. If 'surf\_only' is true, then the 'header\_grid\_time' file will be written and corresponds to the 'grid\_time' files, while the 'header' file corresponds to the 'grid\_initial' files.

```
grid_time_YYYYMMDD000000_001
```

These are binary format files containing the surface influence for every release of particles (i.e., back trajectory) and written in sparse matrix form. If 'surf\_only' is true, then these files contain only contain the surface layer. FLEXINVERT assumes these files have units of  $\text{s m}^3 \text{ kg}^{-1}$ . Note that the last 3 digits of the file name specify the species number – if only one species is run in FLEXPART (the usual case) then the species number will be '001'. If more than one species is run in FLEXPART, then the user will need to specify which number to use in FLEXINVERT – this is defined by the variable 'spec' set in 'control.def'.

If 'calcbg' is true you will also need:

```
grid_initial_001
```

A binary format file containing the sensitivity to the mixing ratio at the termination of the back trajectory. Written in sparse matrix form. These files must have full vertical resolution (with an uppermost layer of at least 20 km altitude).

### 2.2 Prior emission files

FLEXINVERT includes a subroutine to read emission files in NetCDF format (reademissions.f90). It is assumed that this variable has the dimensions 'longitude', 'latitude', and 'time' (also with these names). The default flux unit is  $\text{kg m}^{-2} \text{ h}^{-1}$ . If the emissions grid is different to that used in the FLEXPART runs, then the emissions will be interpolated/aggregated onto the FLEXPART grid resolution.

### 2.3 Observation files

The default subroutine to read the observation files (readobsfile.f90) assumes one file per receptor and that these are in ascii format. The naming convention of the observation files is:

```
RECNAME_SPEC_RECTYP_{YYYY}.dat
```

where RECNAME is the receptor name and must correspond to the FLEXPART output directory, SPEC is the species name and must correspond to the species simulated in FLEXPART, and RECTYP is the 2-character description of the measurement type: FM for Flask Measurement, CM for Continuous Measurement and SM/AM for Ship/Aircraft Measurement, and YYYY is the year and is only specified for annual data files (see also sections 1.2, 2.1 and 2.6).

The default file format is the following (other formats will require modification to 'readobs.f90'):

```
21 header rows
```

```
YYYY MM DD HH MI SS LON LAT ALT CONC
```

where the concentration ('CONC') is in units of mole fraction e.g. ppb, and the time stamp is given in UTC. The variable 'coef' in 'control.def' is the coefficient to convert from the units of 'ppt' (determined by the FLEXPART output unit) to whatever unit the observations are in.

### 2.4 Background or 'historical' concentrations

For the background calculation option 'calcbg', input files of background concentrations are required. Currently, subroutines to read output files from the models LMDZ and TM5, as well as 3D concentration fields based on NOAA data, are provided and can be used as templates for other file formats. The background concentrations from LMDZ and TM5 are time resolved 3D fields in NetCDF format.

### 2.5 Land-cover and land-sea mask files

For calculating the variable grid, information on land-cover (namely, land, sea or ice) is needed. By default, FLEXINVERT uses the IGBP-DIS 1 km land-cover dataset from Belward et al. (2009), which is provided with the FLEXINVERT tar archive. In addition, FLEXINVERT uses a land-sea mask to calculate the spatial correlations (these are separate for land and ocean. Again, the default land-sea mask file is provided in the tar archive.

### 2.6 Receptor definition file

The receptor definition file tells FLEXINVERT which receptors/observations to include in the inversion. The receptor file is specified in 'files.def' as the variable 'receptfile'. The 'receptfile' is structured as follows:

```
REC    TYPE    FREQ  
RECNAME RECTYP ANNUAL
```

where the first line is a header line and the subsequent lines specify the receptor names 'RECNAME' (which must correspond to the FLEXPART output directory and observation files), the receptor types 'RECTYP' (one of 'FM', 'CM', 'SM' or 'AM') and the observation file frequencies (either 'ANNUAL' for annual files or 'NULL' if the observations are not separated by time). (See also sections 2.1 and 2.3).

### 3. Output data

#### 3.1 Standard output

`modout.txt`

Contains the simulated mole fractions using the original temporal and spatial resolution of the FLEXPART runs (units mole fraction,  $\text{nmol mol}^{-1}$ , abbreviated as ppb).

1. receptor name (RECNAME, 7 characters)
2. number given to each modelled concentration (for moving receptors this is equivalent to the number in 'obsread.txt')
3. date (yyyymmdd)
4. model timestep (hour since start of inversion period)
5. prior concentration (with the background is subtracted)
6. posterior concentration (with the background is subtracted)
7. prior background concentration (zero if 'calcbg' is false)
8. posterior background concentration (zero if 'calcbg' or 'optimizbg' is false)

`analysis.nc`

Contains 3D variables (longitude, latitude, time) of fluxes on the fine grid (units  $\text{kg m}^{-2} \text{s}^{-1}$ ).

longitude:	longitude (eastern boundary)
latitude:	latitude (southern boundary)
time:	index of emission time step
emis_post:	posterior fluxes (on regular grid)
emis_prior:	prior fluxes (on regular grid)
error_post:	posterior error (on regular grid)
error_prior:	prior error (on regular grid)

Note: error\_post is calculated from  $(J''(\mathbf{x}))^{-1}$  which tends to underestimate the posterior error (see section 1.3).

`monitor.txt`

Observed mole fractions (ppb) used in the inversion and the corresponding prior and posterior model simulations (using the coarse grid), and the prior and posterior background concentrations and the observation error.

1. receptor name (RECNAME, 7 characters)
2. date (yyyymmddhh)
3. prior concentration (if 'optimizbg' is false the background is subtracted)
4. posterior concentration (if 'optimizbg' is false the background is subtracted)
5. prior background (if 'calcbg' is false this is zero)
6. posterior background (if 'calcbg' or 'optimizbg' is false this is zero)
7. observation uncertainty (quadratic sum of all uncertainties in the observation space)

`obsread.txt`

Observations used in the inversion

`nbox_xy.nc`

Contains the 2D matrix relating the fine grid to the variable grid and may be used alternatively to `grid_operator.nc`. The variable `nbox_xy` has the same dimensions (longitude, latitude) as the nested domain at the fine resolution and has elements from 1 to `nbox` (i.e., the number of grid cells on the variable grid).



grid\_operator.nc

The 2D projection (summation) operator,  $\Gamma$  from the fine to the variable grid with rows corresponding to each of the variable size grid cells and columns corresponding to each of the fine size grid cells.

grid\_coarse.nc

Only used if ‘calcbg’ is true. The 2D projection operator,  $\Gamma_{cg}$  from the fine to the coarse grid with rows corresponding to each of the grid cells on the coarse grid (i.e., used for the background optimization) and columns corresponding to each of the grid cells on the fine grid.

knest.nc

Contains the 2D transport operator,  $\mathbf{H}_{vg}^{nest}$  for the variable grid with rows corresponding to all releases, i.e., back trajectories (units  $s\ m^3\ kg^{-1}$ ).

kout.nc

Contains the 2D transport operator,  $\mathbf{H}^{out}$  for the coarse grid with rows corresponding to all releases, i.e., back trajectories (unit is mixing ratio).

knest\_fin.nc

Contains the 2D transport operator,  $\mathbf{H}^{nest}$  for the fine grid (averaged over all emission time steps) with rows corresponding to all releases, i.e., back trajectories (units  $s\ m^3\ kg^{-1}$ ).

bgscalars.txt

Contains a vector of the scalars of the background concentrations in each of the coarse grid cells (only written if ‘calcbg’ and ‘optimizbg’ are both ‘true’).

### 3.2 Additional files for testing purposes

true.txt

Only used if ‘perturb’ is true. Contains the prior state vector (coarse grid) before perturbation and hence is equivalent to the *true* fluxes (units  $kg\ m^{-2}\ s^{-1}$ ).

prior.txt

The prior state vector,  $\mathbf{x}_{b,vg}$  with the fluxes in the nested domain (units  $kg\ m^{-2}\ s^{-1}$ ) at coarse resolution and, if ‘optimizbg’ is true, the scalars of the background mole fractions.

posterior.txt

The posterior state vector,  $\mathbf{x}_{vg}$  with the fluxes in the nested domain (units  $kg\ m^{-2}\ s^{-1}$ ) at coarse resolution and, if ‘optimizbg’ is true, the scalars of the background mole fractions.

area\_box.txt

Contains a vector of the areas of the variable grid cells in the same order as in the above 3 files (units  $m^2$ ).

knest\_trim.nc

Contains the 2D transport operator,  $[\mathbf{H}_{vg}^{nest}\ \mathbf{M}]$  for the variable grid with  $M$  rows corresponding to the observations used in the inversion (units  $s\ m^3\ kg^{-1}$ ).

knest\_fintrim.nc

Contains the 2D transport operator,  $\mathbf{H}^{\text{nest}}$  for the fine grid (averaged over all emission time steps) with rows corresponding to all releases averaged to the observation frequency.

emisflex.nc

The prior fluxes (longitude, latitude, time) interpolated to the fine grid (units  $\text{kg m}^{-2} \text{s}^{-1}$ )

gain.nc

Contains the Gain matrix,  $\mathbf{G}$  (see ‘bayesian1.f90’ and ‘bayesian2.f90’).

covb.nc

Contains the Eigen-values and vectors of the prior error covariance matrix,  $\mathbf{B}_{\text{vg}}$  for the variable grid (note that these are scaled by the numerical scaling factor – see mod\_var.f90).

covbfin.nc

Contains the Eigen-values and vectors for the prior error covariance matrix,  $\mathbf{B}^{\text{flux}}$  for the fine grid at only 1 time-step (note that these are scaled by the numerical scaling factor).

covbfinaw.nc

Contains the Eigen-values and vectors for the area-weighted prior error covariance matrix,  $\mathbf{B}^{\text{flux, aw}}$  for the regular grid at only 1 time-step (note that these are scaled by the numerical scaling factor).

covagg.nc

Contains the aggregation error covariance matrix in the observation space.

cova.nc

Contains the posterior error covariance matrix,  $\mathbf{A}_{\text{vg}}$  for the variable grid (note that this is scaled by the numerical scaling factor).

covr.nc

Contains the observation error covariance matrix (including the aggregation error),  $\mathbf{R}$ .

## 4. Running FLEXINVERT

### 4.1 System requirements

FLEXINVERT is written in Fortran90 and comes with a makefile for use with the gfortran compiler. It requires the following libraries for Fortran90: LAPACK and NetCDF.

### 4.2 Making the first run

To compile the code using the gfortran compiler:

```
make -f makefile_gfortran
```

Edit the files for the inversion settings and input/output files:

```
control.def
```

```
files.def
```

Run the code in background mode:

```
nohup ./FLEXINVERT &
```

## 5. References

Belward, A. S., Estes, J. E., and Kline, K. D. (1999), The IGBP-DIS global 1-km land-cover data set DISCover: A project overview, *Photogrammetric Engineering and Remote Sensing*, 65(9), 1013-1020

Tarantola, A. (2005), *Inverse problem theory and methods for model parameter estimation*, Society for Industrial and Applied Mathematics, Philadelphia

Thacker, W. C. (2007), Data assimilation with inequality constraints, *Ocean Modelling*, 16(3-4), 264-276, doi:10.1016/j.ocemod.2006.11.001

Thompson, R. L., and Stohl, A. (2014), FLEXINVERT: an atmospheric Bayesian inversion framework for determining surface fluxes of trace species using an optimized grid, *Geosci. Model Devel.*, 7, 2223-2242, doi:10.5194/gmd-7-2223-2014

