



Supplement of

The Community Inversion Framework v1.0: a unified system for atmospheric inversion studies

Antoine Berchet et al.

Correspondence to: Antoine Berchet (antoine.berchet@lscce.ipsl.fr)

The copyright of individual parts of the supplement might differ from the article licence.

Table S1: Plugin types in pyCIF

★ obsoperator:	<ul style="list-style-type: none"> - purpose: maps data from the control space to the observation space and conversely - inheritance: <ul style="list-style-type: none"> - controlvect - obsvect - model - metadata: none - data: none - methods: <ul style="list-style-type: none"> - obsoper: $\mathbf{x} \rightarrow \mathcal{H}(\mathbf{x})$ <li style="padding-left: 40px;">$\mathcal{H}^*(\mathbf{y}^*) \leftarrow \mathbf{y}^*$ - calls: <ul style="list-style-type: none"> - controlvect.$\Pi_{\mathcal{X}}^{\mathcal{F}}$: $\mathbf{x} \leftrightarrow \mathbf{x}_{\text{model}}$ - model.$\Pi_{\mathcal{F}}^{\mathcal{F}}$: $\mathbf{x}_{\text{model}} \leftrightarrow$ model inputs - model.run: model inputs \leftrightarrow model outputs - model.$\Pi_{\mathcal{C}}^{\mathcal{M}}$: model outputs $\leftrightarrow \mathbf{y}_{\text{model}}$ - obsvect.$\Pi_{\mathcal{M}}^{\mathcal{Y}}$: $\mathbf{y}_{\text{model}} \leftrightarrow \mathbf{y}^0$
★ controlvect:	<ul style="list-style-type: none"> - purpose: initializes the control vector (including metadata) and uncertainties and computes projections from and to the control space - comments: needs metadata from the model (e.g., resolution) and the domain to carry out projections - inheritance: <ul style="list-style-type: none"> - domain - model - metadata: <ul style="list-style-type: none"> - components (e.g., fluxes, initial conditions, etc.) - dimension - correlation patterns - data: <ul style="list-style-type: none"> - \mathbf{x}_b - \mathbf{B} if stored, main components otherwise - methods: <ul style="list-style-type: none"> - $\Pi_{\mathcal{X}}^{\mathcal{F}}$: $\mathbf{x} \leftrightarrow \mathbf{x}_{\text{model}}$ - $\Pi_{\mathcal{X}}^{\mathcal{C}}$: $\mathbf{x} \rightarrow \chi \equiv \mathbf{B}_{1/2}\mathbf{x}$ - $\Pi_{\mathcal{X}}^{\mathcal{Y}}$: $\chi \rightarrow \mathbf{x} \equiv \mathbf{B}_{1/2}\chi$ - init_B: some data $\rightarrow \mathbf{B}$ - calls: <ul style="list-style-type: none"> - domain.resolution
★ obsvect:	<ul style="list-style-type: none"> - purpose: initializes the observation vector (including metadata) from the measurements and computes projections from and to the observation space

- inheritance: - domain
 - measurements
 - metadata - species
 - observation type (in situ, satellite, etc.)
 - correlations if any
 - data - \mathbf{y}^o
 - \mathbf{R} if stored, main components otherwise
 - methods:
 - $\Pi_{\mathcal{M}}^{\mathcal{Y}}$: $\mathbf{y}_{\text{model}} \leftrightarrow \mathbf{y}^o$
 - $\Pi_{\mathcal{O}}^{\mathcal{Y}}$: $\mathbf{y}_{\text{meas}} \leftrightarrow \mathbf{y}^o$
 - $\mathbf{R}^{-1} \cdot ()$: $\mathbf{y} \rightarrow \mathbf{y} = \mathbf{R}^{-1}\mathbf{y}$
 - calls: - domain.resolution
-

- ★ model:
 - purpose: drives the transport model, prepares inputs and extracts outputs to CIF-compatible structures
 - inheritance: none
 - metadata - resolution
 - computation mode
 - sub-periods if any
 - chemistry if any
 - model-specific configuration
 - path to fixed inputs (e.g., meteo data)
 - data none
 - methods:
 - run: model inputs \rightarrow model outputs
 - $\Pi_{\mathcal{F}}^{\mathcal{C}}$: $\mathbf{x}_{\text{model}} \leftrightarrow$ model inputs
 - $\Pi_{\mathcal{C}}^{\mathcal{M}}$: model output \leftrightarrow $\mathbf{y}_{\text{model}}$
 - calls: none
-

- ★ simulator:
 - purpose: computes the cost function and its gradient
 - inheritance: observation operator
 - metadata none
 - data none
 - methods: simul: $\chi \rightarrow (J(\chi), \nabla_{\chi} J)$
 - calls: - obsoperator.obsoper: $\mathbf{x} \rightarrow \mathcal{H}(\mathbf{x})$
 - $\delta\mathbf{y} \rightarrow \mathbf{H}^* \delta\mathbf{y}$
-

- ★ minimizer:
 - purpose: minimizes a function starting from a given point
 - inheritance: observation operator
 - metadata none

- data none
- methods: minimize: $(\chi, J) \rightarrow \chi_{\text{opt}}$
- calls: - simulator.simul: $\chi \rightarrow (J(\chi), \nabla_{\chi} J)$

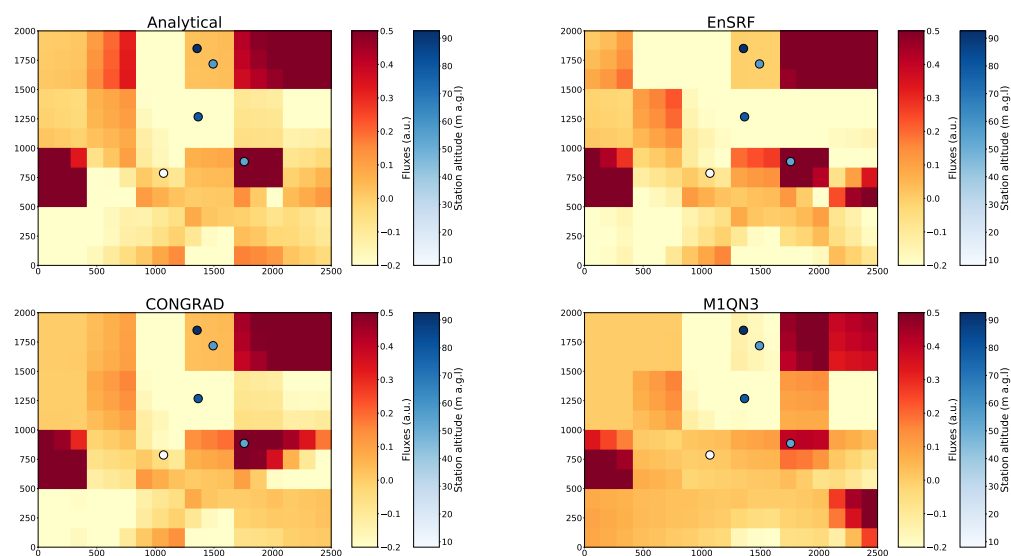


Figure S1. Posterior increments for analytical, EnSRF, variational with CONGRAD and variational with M1QN3 (from top to bottom, left to right) for an inversion set-up with aggregated bands.

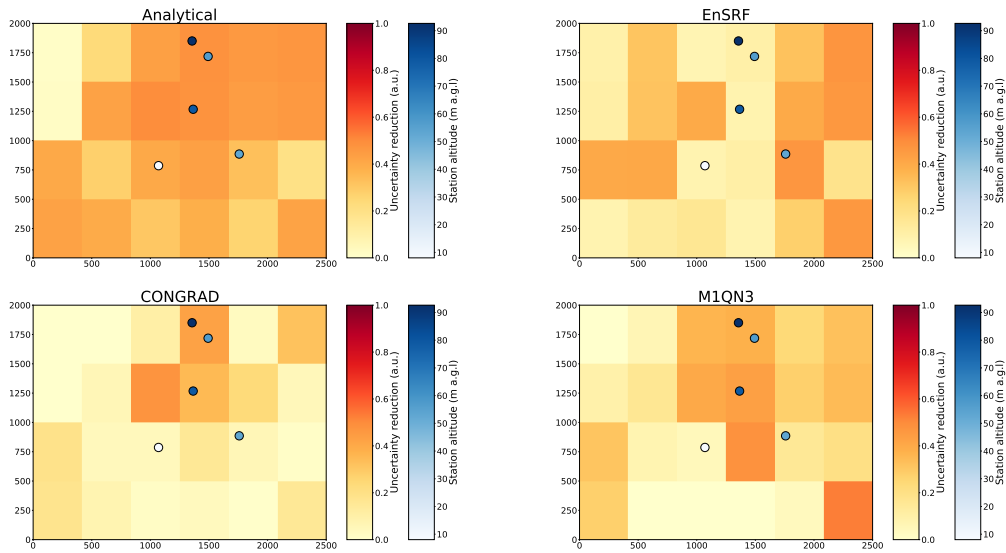


Figure S2. Uncertainty reduction for analytical, EnSRF, variational with CONGRAD and variational with M1QN3 (from top to bottom, left to right) for an inversion set-up with an inversion set-up with aggregated bands.

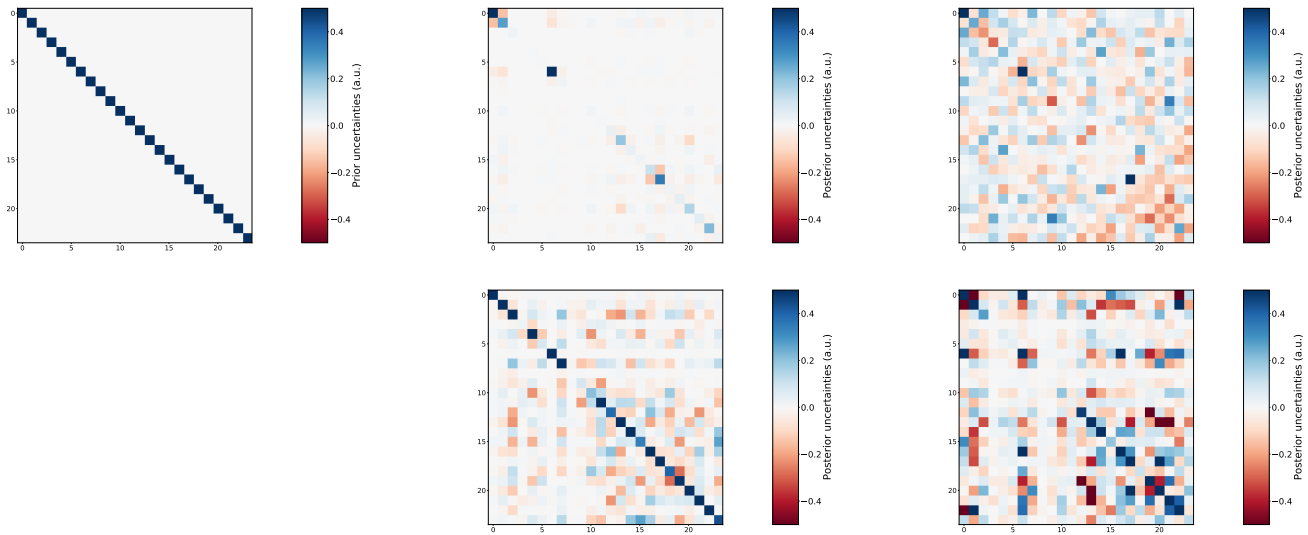


Figure S3. Prior (left) and posterior (middle and right) uncertainty matrices for analytical, EnSRF, variational with CONGRAD and variational with M1QN3 (from top to bottom, middle and right columns) for an inversion set-up with aggregated bands.

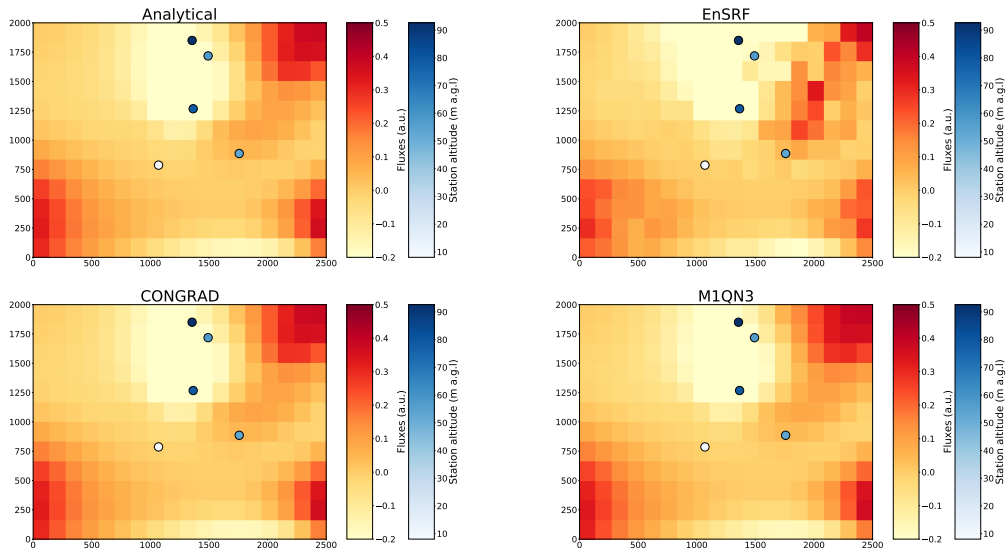


Figure S4. Same as Fig. S1 with an inversion set-up at the pixel resolution with horizontal correlation length of 200000 m.

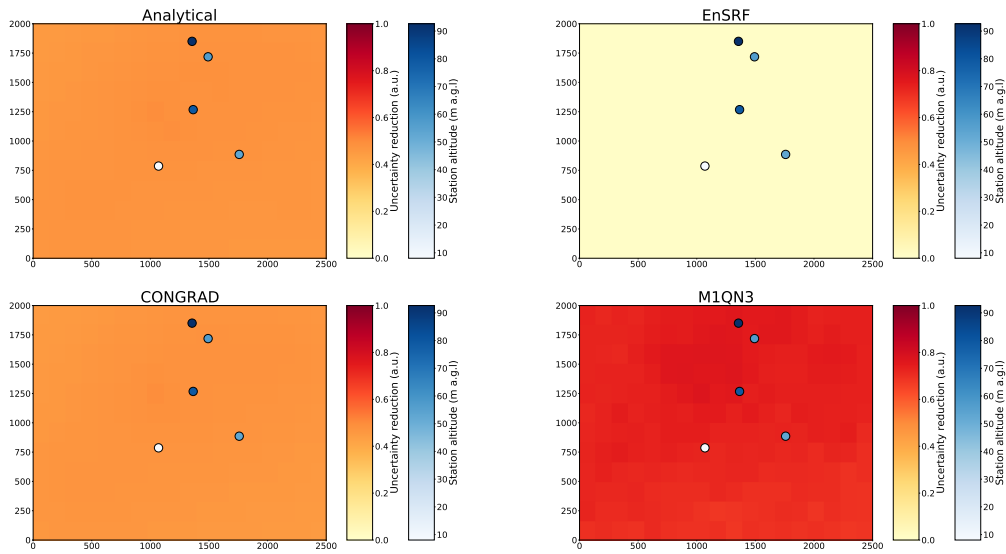


Figure S5. Uncertainty reduction for analytical, EnSRF, variational with CONGRAD and variational with M1QN3 (from top to bottom, left to right) for an inversion set-up at the pixel resolution with horizontal correlation length of 200000 m.

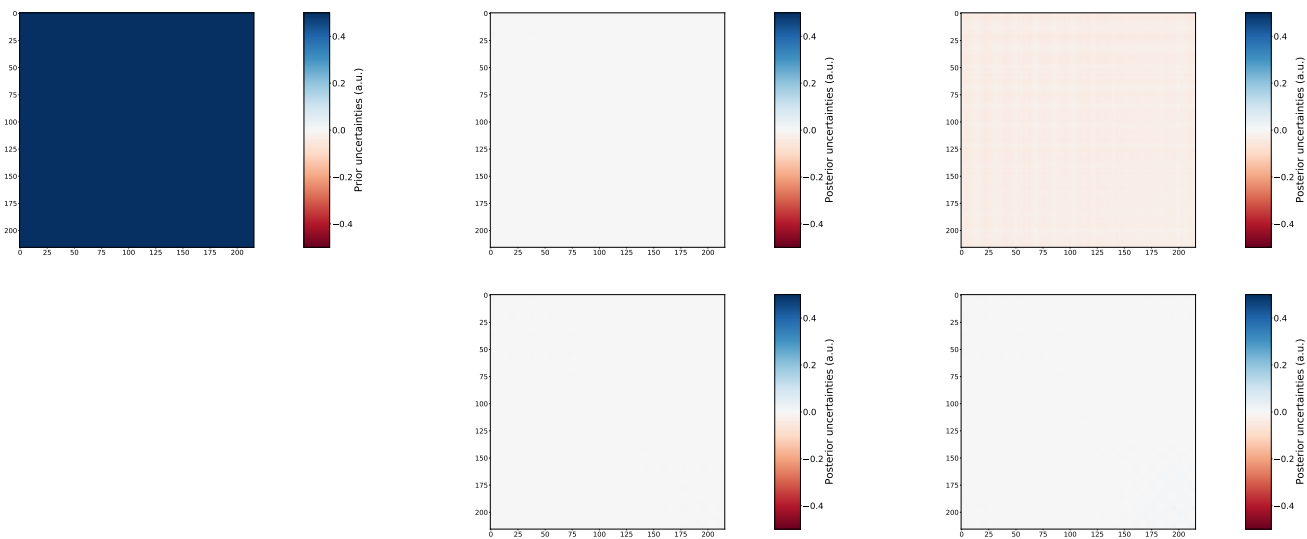


Figure S6. Prior (left) and posterior (middle and right) uncertainty matrices for analytical, EnSRF, variational with CONGRAD and variational with M1QN3 (from top to bottom, middle and right columns) with an inversion set-up at the pixel resolution with horizontal correlation length of 200000 m.

S3 Configuration examples

S3.1 Generate the truth observations

```

rootdir: &rootdir /tmp/CIF/
outdir: &outdir /home/chimereicos/aberchet/debugchimere/pytest
verbose: 2
logfile: pycif.logtest
workdir: !join [*outdir, /fwd_long_families]
datei: 2010-01-01
datef: 2010-01-05 00:00:00
mode:
  obserror: 0.01
  perturb_obsvect: true
  plugin:
    name: forward
    version: std
controlvect:
  plugin:
    name: standard
    version: std
save_out_netcdf: true
transform_pipe:
  trans1:
    component: fluxes
    parameter_out: CH4
    parameters_in:
      - CH4_ref
      - CH4_perturb
  plugin:

```



```

    name: families
    type: transform
    version: std
obsoperator:
  autoflush: true
  plugin:
    name: standard
    version: std
model:
  chemistry:
    acspecies:
      CH4: null
      MCF: null
  file_pg: !join [*rootdir, /model_sources/dummy_gauss/Pasquill-Gifford.txt]
  plugin:
    name: dummy
    version: std
  save_H: true
obsvect:
  dir_obsvect: !join [*outdir, /ref_obsvect]
  dump_type: nc
  plugin:
    name: standard
    version: std
datavect:
  components:
  concs:
  parameters:
    CH4:
      duration: 4H13min
      frequency: 3H22min
      nstations: 5
      plugin:
        name: random
        type: measurements
        version: param
      random_subperiod_shift: true
      seed: true
      seed_id: 5
      zmax: 100
    MCF:
      duration: 5H
      frequency: 4H33min
      nstations: 10
      plugin:
        name: random
        type: measurements
        version: param
      random_subperiod_shift: true
      seed: true
      seed_id: 100
      zmax: 100
  fluxes:
  parameters:
    CH4_perturb:
      err: 1
      errtype: max
      flx_formula:
      - sum:
        - cos: null
          period: 200
          variable: zlat
        - period: 300
          sin: null
          variable: zlon

```

```

hcorrelations:
  dircorrel: !join [*outdir, /test_integration_fwd_dummy_con0/datavect]
  dump_hcorr: true
  evalmin: 0
  landsea: false
  sigma: 500
hresol: hpixels
nlev: 1
plugin:
  name: dummy
  type: fluxes
  version: txt
tcorrelations:
  sigma_t: 12
tresol: 4D
vresol: vpixels
xb_scale: 0.2
CH4_ref:
  err: 1
  errtype: max
  flx_formula:
  - product:
    - sum:
      - cos: null
        period: 500
        variable: zlat
      - period: 1000
        sin: null
        variable: zlon
    - sum:
      - period: 1000
        square: null
        variable: zlat
      - period: 1000
        square: null
        variable: zlon
hcorrelations:
  dircorrel: !join [*outdir, /test_integration_fwd_dummy_con0/datavect]
  dump_hcorr: true
  evalmin: 0
  landsea: false
  sigma: 500
hresol: hpixels
nlev: 1
plugin:
  name: dummy
  type: fluxes
  version: txt
tcorrelations:
  sigma_t: 12
tresol: 4D
vresol: vpixels
MCF:
  err: 1
  errtype: max
  flx_formula:
  - product:
    - sum:
      - cos: null
        period: 100
        variable: zlat
      - period: 200
        sin: null
        variable: zlon
    - sum:

```

```
- period: 500
  square: null
  variable: zlat
- period: 600
  square: null
  variable: zlon
hcorrelations:
  dircorrel: !join [*outdir, /test_integration_fwd_dummy_con0/datavect]
  dump_hcorr: true
  evalmin: 0
  landsea: false
  sigma: 500
hresol: hpixels
nlev: 1
plugin:
  name: dummy
  type: fluxes
  version: txt
tresol: 4D
vresol: vpixels
meteo:
  plugin:
    name: dummy
    type: meteo
    version: csv
  resolution: 1H
  seed: true
plugin:
  name: standard
  version: std
domain:
  nlat: 12
  nlon: 18
plugin:
  name: dummy
  version: std
xmax: 2500
xmin: 0
ymax: 2000
ymin: 0
```

S3.2 EnSRF inversion at pixel resolution

```
rootdir: &rootdir /tmp/CIF/
outdir: &outdir /home/chimereicos/aberchet/debugchimere/pytest
verbose: 2
logfile: pycif.logtest
workdir: !join [*outdir, /inversion_long_full_ensrf_]
datei: 2010-01-01
datef: 2010-01-05 00:00:00
mode:
  nsample: 50
  plugin:
    name: EnSRF
    version: std
controlvect:
  plugin:
    name: standard
    version: std
  save_out_netcdf: true
obsoperator:
  autoflush: true
  plugin:
    name: standard
    version: std
model:
  chemistry:
    acspecies:
      CH4: null
  file_pg: !join [*rootdir, /model_sources/dummy_gauss/Pasquill-Gifford.txt]
  plugin:
    name: dummy
    version: std
  reload_H: !join [*outdir, /H_matrix.pickle]
  save_H: true
obsvect:
  dir_obsvect: !join [*outdir, /ref_obsvect]
  dump_type: nc
  plugin:
    name: standard
    version: std
datavect:
  components:
    concs:
      parameters:
        CH4:
          duration: 4H13min
          frequency: 3H22min
          nstations: 5
          plugin:
            name: random
            type: measurements
            version: param
          random_subperiod_shift: true
          seed: true
          seed_id: 5
          zmax: 100
    fluxes:
      parameters:
        CH4:
          err: 1
          errtype: max
          flx_formula:
            - product:
              - sum:
```

```

- cos: null
  period: 500
  variable: zlat
- period: 1000
  sin: null
  variable: zlon
- sum:
  - period: 1000
    square: null
    variable: zlat
  - period: 1000
    square: null
    variable: zlon
hcorrelations:
  dircorrel: !join [*outdir, /test_integration_inversion_dum2/datavect]
  dump_hcorr: true
  evalmin: 0
  landsea: false
  sigma: 500
hresol: hpixels
nlev: 1
plugin:
  name: dummy
  type: fluxes
  version: txt
tcorrelations:
  sigma_t: 12
tresol: 4D
vresol: vpixels
meteo:
  plugin:
    name: dummy
    type: meteo
    version: csv
  resolution: 1H
  seed: true
plugin:
  name: standard
  version: std
domain:
  nlat: 12
  nlon: 18
  plugin:
    name: dummy
    version: std
  xmax: 2500
  xmin: 0
  ymax: 2000
  ymin: 0
platform:
  plugin:
    name: LSCE
    version: obelix

```

S3.3 Variational inversion with M1QN3 with bands

```
rootdir: &rootdir /tmp/CIF/
outdir: &rootdir !join [*rootdir, /.tox/py38/tmp]
verbose: 2
logfile: pycif.logtest
workdir: &outdir !join [*outdir, /inversion_long_full_4dvar_M1QN3/]
datei: 2010-01-01
datef: 2010-01-05 00:00:00
mode:
  minimizer:
    dfl: 0.5
    epsg: 0.0002
    maxiter: 25
    nsim: 25
  plugin:
    name: M1QN3
    version: std
  simulator:
    plugin:
      name: gausscost
      version: std
    reload_from_previous: true
  montecarlo: 10
  plugin:
    name: 4dvar
    version: std
  save_out_netcdf: true
  controlvect:
    plugin:
      name: standard
      version: std
    save_out_netcdf: true
  obsoperator:
    autoflush: true
    plugin:
      name: standard
      version: std
  model:
    chemistry:
      acspecies:
        CH4: null
    file_pg: &rootdir !join [*rootdir, /model_sources/dummy_gauss/Pasquill-Gifford.txt]
    plugin:
      name: dummy
      version: std
    reload_H: &outdir !join [*outdir, /test_integration_inversion_dum2/./H_matrix.pickle]
    save_H: true
  obsvect:
    dir_obsvect: &outdir !join [*outdir, /test_integration_inversion_dum2/./ref_obsvect/]
    dump_type: nc
    plugin:
      name: standard
      version: std
  datavect:
    components:
      concs:
        parameters:
          CH4:
            duration: 4H13min
            frequency: 3H22min
            nstations: 5
            plugin:
```

```

    name: random
    type: measurements
    version: param
    random_subperiod_shift: true
    seed: true
    seed_id: 5
    zmax: 100
fluxes:
  parameters:
    CH4:
      err: 1
      errtype: max
      flx_formula:
        - product:
            - sum:
                - cos: null
                  period: 500
                  variable: zlat
                - period: 1000
                  sin: null
                  variable: zlon
            - sum:
                - period: 1000
                  square: null
                  variable: zlat
                - period: 1000
                  square: null
                  variable: zlon
      hcorrelations:
        dircorrel: &outdir !join [*outdir, /test_integration_inversion_dum2/datavect/]
        dump_hcorr: true
        evalmin: 0
        landsea: false
        sigma: 2000000.0
      hresol: hpixels
      nlev: 1
      plugin:
        name: dummy
        type: fluxes
        version: txt
      tcorrelations:
        sigma_t: 12
        tresol: 4D
        vresol: vpixels
  meteo:
    plugin:
      name: dummy
      type: meteo
      version: csv
      resolution: 1H
      seed: true
  plugin:
    name: standard
    version: std
domain:
  nlat: 12
  nlon: 18
  plugin:
    name: dummy
    version: std
  xmax: 2500
  xmin: 0
  ymax: 2000
  ymin: 0
platform:

```

```
plugin:  
  name: docker  
  version: cif
```