Geoscientific
Model Development

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@lsce.ipsl.fr)

S1 Tables

Table S1: Plugin types in pyCIF

⋆ obsoperator:	- purpose:	maps data from the control space to the observation space and conversely
	- inheritance:	- controlvect
		- obsvect
		- model
	- metadata	none
	- data	none
	- methods:	- obsoper:
	- calls:	- controlvect.$\Pi_{\mathcal{X}}^{\mathfrak{F}}$:
		- model.$\Pi_{\mathfrak{F}}^{\mathcal{F}}$:
		- model.run:
		- model.$\Pi_{\mathcal{C}}^{\mathfrak{M}}$:
		- obsvect.$\Pi_{\mathfrak{M}}^{\mathcal{Y}}$:

- methods: - obsoper:
$$\mathbf{x} \rightarrow \mathcal{H}(\mathbf{x})$$
$$\mathcal{H}^*(\mathbf{y}^*) \leftarrow \mathbf{y}^*$$

- calls: - controlvect.$\Pi_{\mathcal{X}}^{\mathfrak{F}}$: $\quad \mathbf{x} \leftrightarrow \mathbf{x}_{\text{model}}$
- model.$\Pi_{\mathfrak{F}}^{\mathcal{F}}$: $\quad \mathbf{x}_{\text{model}} \leftrightarrow$ model inputs
- model.run: $\quad$ model inputs $\leftrightarrow$ model outputs
- model.$\Pi_{\mathcal{C}}^{\mathfrak{M}}$: $\quad$ model outputs $\leftrightarrow \mathbf{y}_{\text{model}}$
- obsvect.$\Pi_{\mathfrak{M}}^{\mathcal{Y}}$: $\quad \mathbf{y}_{\text{model}} \leftrightarrow \mathbf{y}^{\text{o}}$

⋆ controlvect:	- 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:	- domain
		- model
	- metadata:	- components (e.g., fluxes, initial conditions, etc.)
		- dimension
		- correlation patterns
	- data:	- $\mathbf{x}_b$
		- $\mathbf{B}$ if stored, main components otherwise
	- methods:	- $\Pi_{\mathcal{X}}^{\mathfrak{F}}$:
		- $\Pi_{\mathcal{X}}^{\mathfrak{A}}$:
		- $\Pi_{\mathfrak{A}}^{\mathcal{X}}$:
		- init_$\mathbf{B}$:
	- calls:	- domain.resolution

- methods: - $\Pi_{\mathcal{X}}^{\mathfrak{F}}$: $\quad \mathbf{x} \leftrightarrow \mathbf{x}_{\text{model}}$
- $\Pi_{\mathcal{X}}^{\mathfrak{A}}$: $\quad \mathbf{x} \rightarrow \chi \equiv \mathbf{B}_{1/2}\mathbf{x}$
- $\Pi_{\mathfrak{A}}^{\mathcal{X}}$: $\quad \chi \rightarrow \mathbf{x} \equiv \mathbf{B}_{1/2}\chi$
- init_$\mathbf{B}$: $\quad$ some data $\rightarrow \mathbf{B}$

⋆ obsvect:	- 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	- $\mathbf{y}^{\mathrm{o}}$			
		- $\mathbf{R}$ if stored, main components otherwise			
	- methods:	- $\mathbf{\Pi}_{\mathfrak{M}}^{\mathcal{Y}}$:	$\mathbf{y}_{\mathrm{model}}$	$\leftrightarrow$	$\mathbf{y}^{\mathrm{o}}$
		- $\mathbf{\Pi}_{\mathfrak{D}}^{\mathcal{Y}}$:	$\mathbf{y}_{\mathrm{meas}}$	$\leftrightarrow$	$\mathbf{y}^{\mathrm{o}}$
		- $\mathbf{R}^{-1} \cdot ()$:	$\mathbf{y}$	$\rightarrow$	$\mathbf{y} = \mathbf{R}^{-1}\mathbf{y}$
	- calls:	- domain.resolution			

$\star$ 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
		- $\mathbf{\Pi}_{\mathcal{F}}^{\mathcal{C}}$:	$\mathbf{x}_{\mathrm{model}}$	$\leftrightarrow$	model inputs
		- $\mathbf{\Pi}_{\mathcal{C}}^{\mathfrak{M}}$:	model output	$\leftrightarrow$	$\mathbf{y}_{\mathrm{model}}$
	- calls:	none			

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

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

- data   none

- methods: minimize: $(\chi, J) \quad \rightarrow \quad \chi_{\text{opt}}$

- calls: - simulator.simul: $\chi \quad \rightarrow \quad (J(\chi), \nabla_\chi J)$
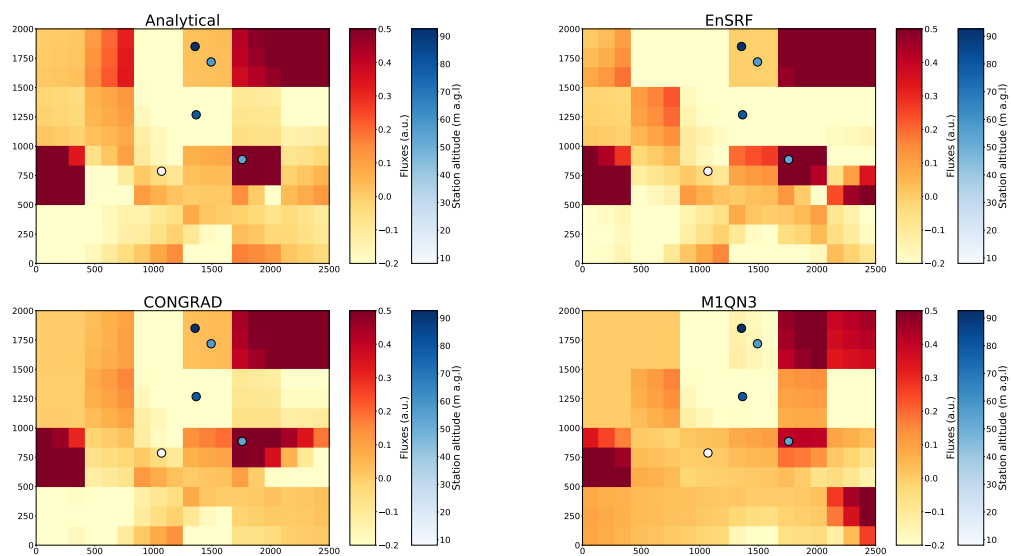
Figure S1. Posterior increments for analytical, EnSRF, variational with CONGRAD and variational with M1QN3 (from top to bottom, left to right) for an inversion set-up with aggregated bands.
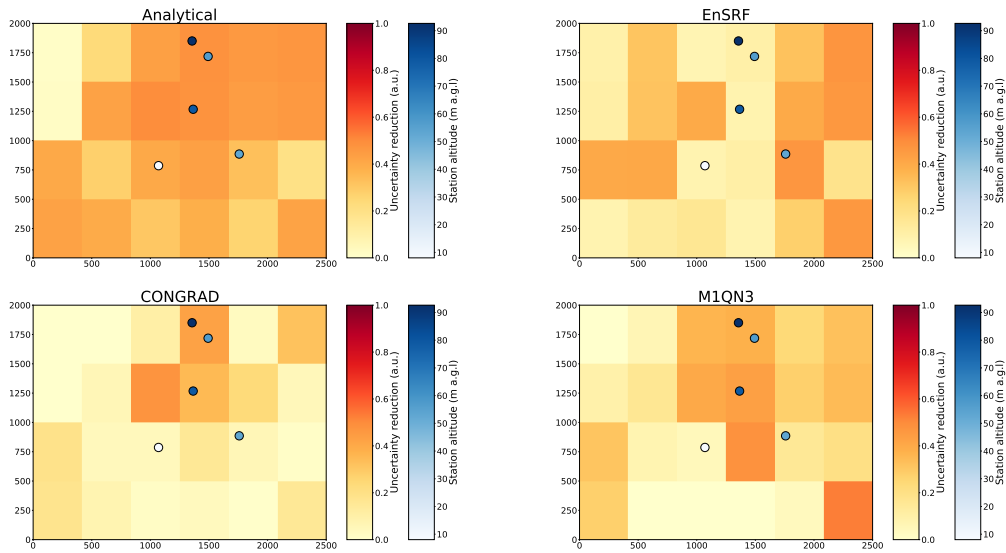
Figure S2. Uncertainty reduction for analytical, EnSRF, variational with CONGRAD and variational with M1QN3 (from top to bottom, left to right) for an inversion set-up with an inversion set-up with aggregated bands.
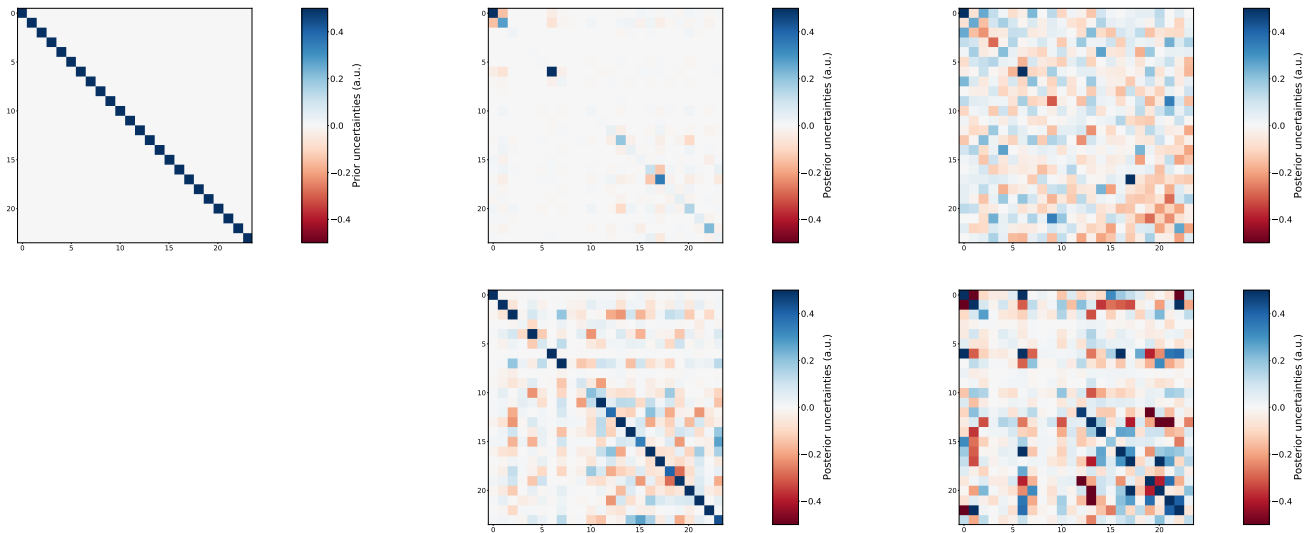

Figure S3. Prior (left) and posterior (middle and right) uncertainty matrices for analytical, EnSRF, variational with CONGRAD and variational with M1QN3 (from top to bottom, middle and right columns) for an inversion set-up with aggregated bands.
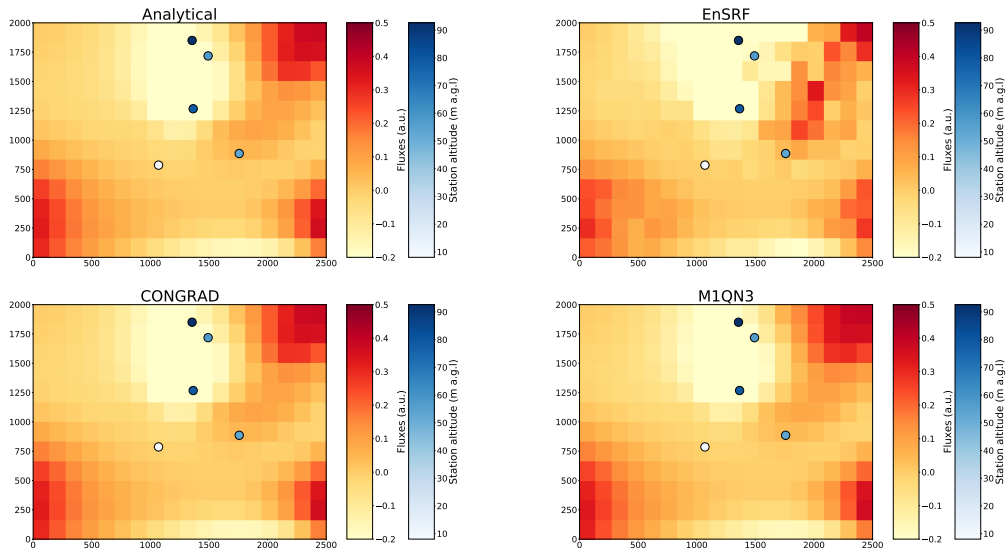
Figure S4. Same as Fig. S1 with an inversion set-up at the pixel resolution with horizontal correlation length of 200000 m.
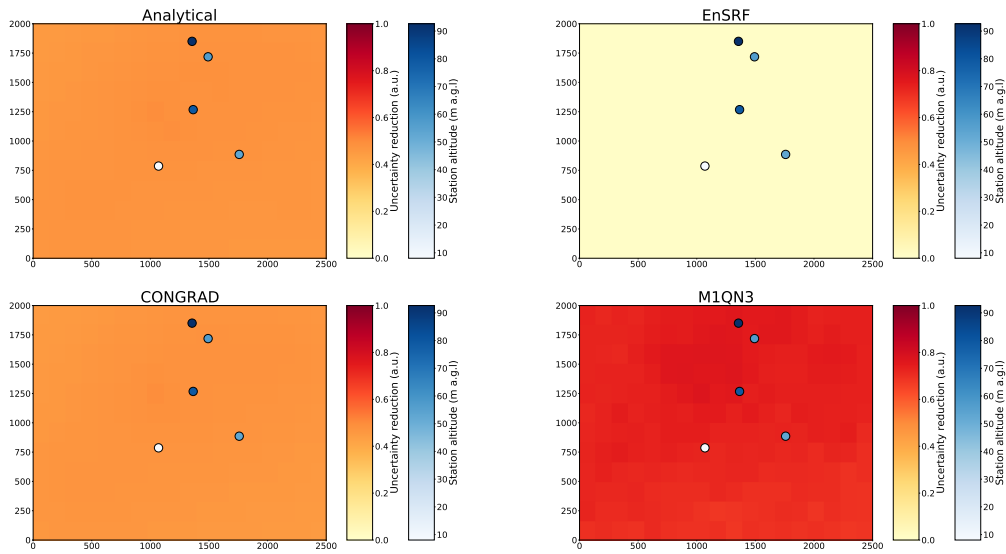

Figure S5. Uncertainty reduction for analytical, EnSRF, variational with CONGRAD and variational with M1QN3 (from top to bottom, left to right) for an inversion set-up at the pixel resolution with horizontal correlation length of 200000 m.
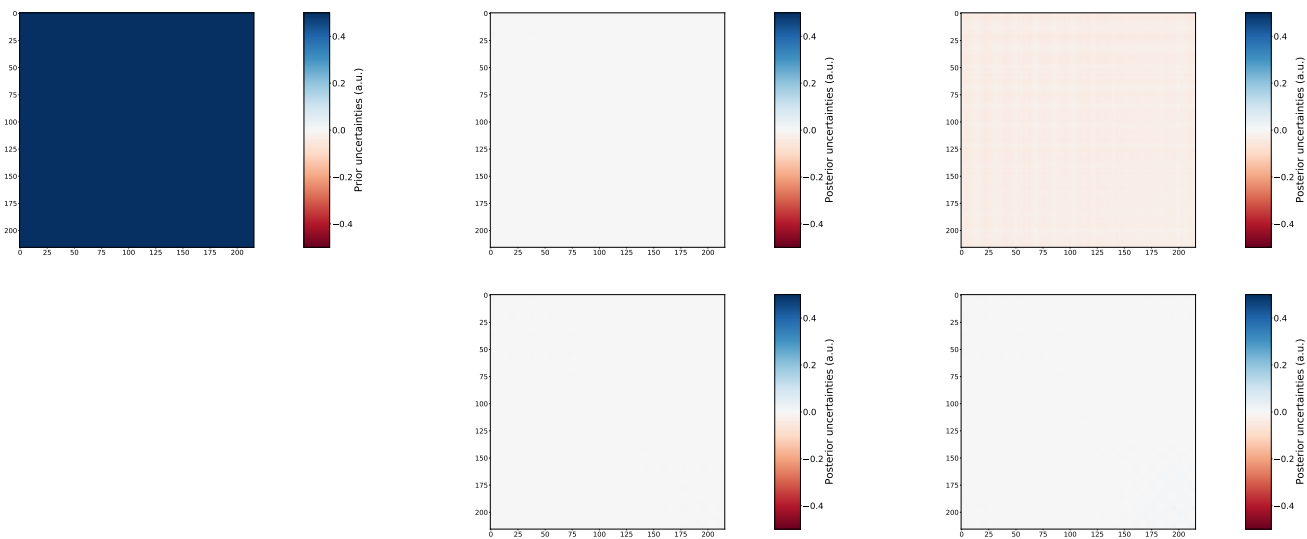
Figure S6. Prior (left) and posterior (middle and right) uncertainty matrices for analytical, EnSRF, variational with CONGRAD and variational with M1QN3 (from top to bottom, middle and right columns) with an inversion set-up at the pixel resolution with horizontal correlation length of 200000 m.

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:
```

```yaml
          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
```

```yaml
      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:
```

```yaml
              - 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:
```

```yaml
                - 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:
    df1: 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:
```

```yaml
              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:
```

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