On the parallelisation of CO_2 flux inversion schemes

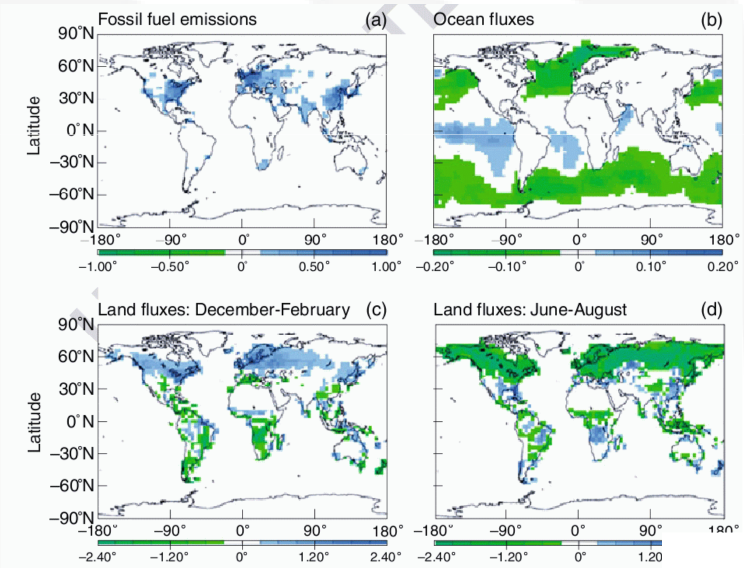
Frédéric Chevallier

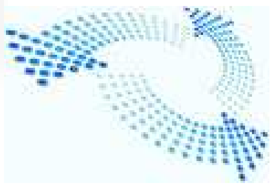
Laboratoire des Sciences du Climat et de l'Environnement
Gif-sur-Yvette
France

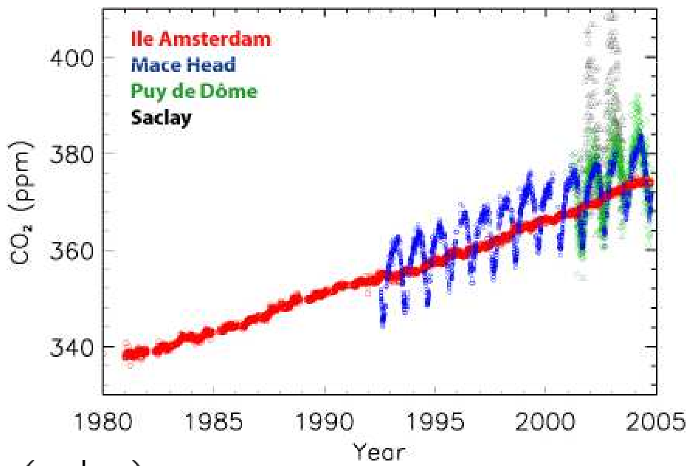
LSCE
Laboratoire des Sciences du Climat
& de l'Environnement

Global CO_2 flux estimation

- Non-reversible atmospheric mixing
  - Need a statistical approach to revert the sign of time







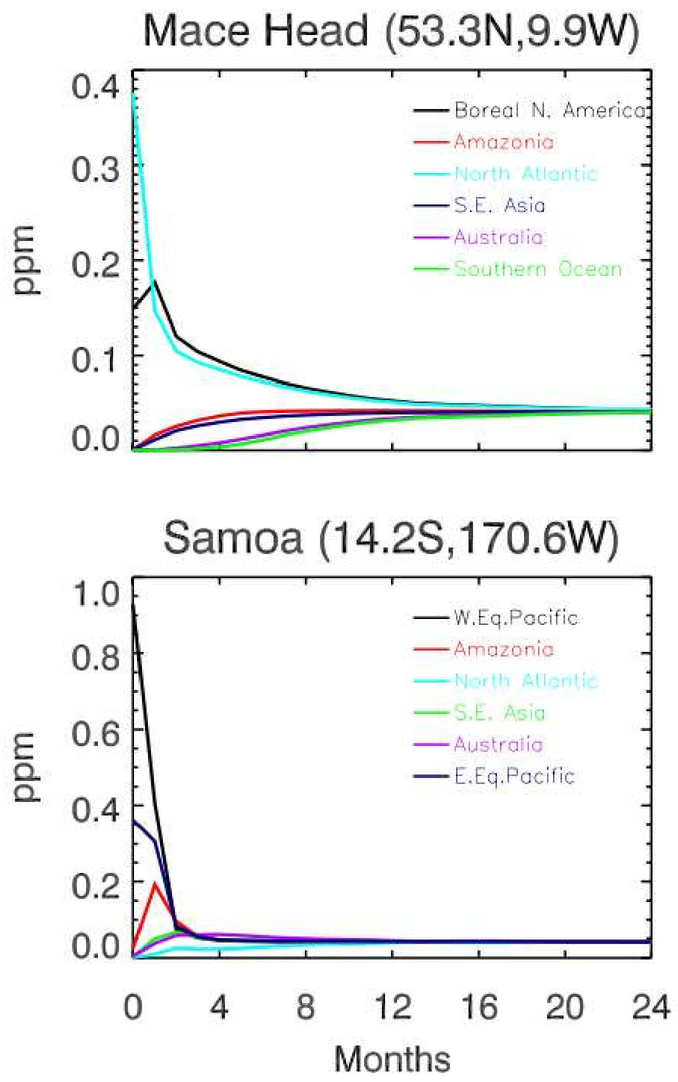
$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x}).p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}$$

LSCE

Time scales

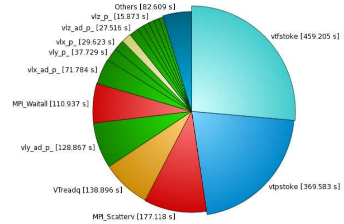
- Long inversion windows needed for long-lived gases
- Interhemispheric exchange time of ~ 1.4 years (Patra et al. 2011)

Monthly average basis functions calculated as one-month flux pulses transported forward from each source region and sampled at a particular observation site. Bruhwiler et al. (2005)

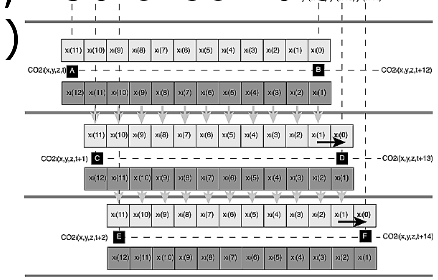


Parallelisation strategies

- Use current supercomputing infrastructures
- Solution 1: variational formulation, parallelise the atmospheric transport model with MPI
  - No simplification
  - Tedious effort
  - Technical hard point: I/O
  - Ex: 32-yr inversion in ~ 2 months on 8 CPUs (Chevallier et al. 2009)



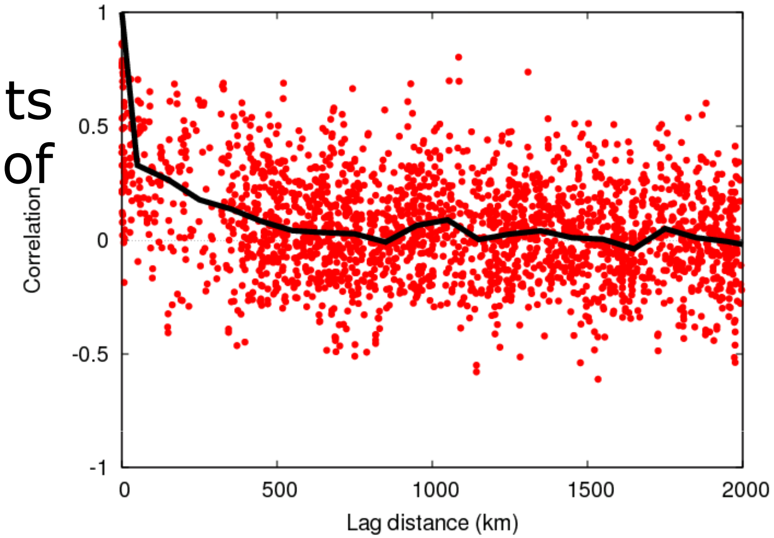
- Solution 2: ensemble formulation, split the state vector **x** into independent pieces
  - Compromise between mixing time and length of **x**
  - Additional localisation simplifications needed
  - Physical consistency is achieved by giving up statistical consistency
  - Ex: NOAA's CarbonTracker (Ensemble Kalman smoother, 150 ensemble members; 5-week inversion window; Peters et al. 1997)



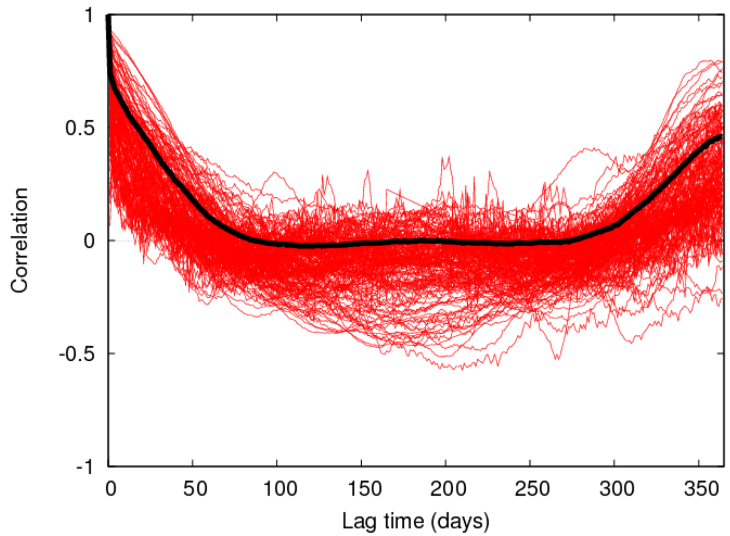
LSCE

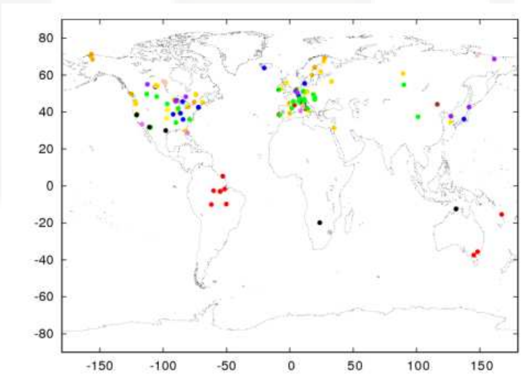
Structure of the prior CO_2 flux errors

- Use daily-mean eddy-covariance flux measurements to assign the error statistics of the prior fluxes



- Results:
  - Small spatial correlations
  - Large temporal correlations



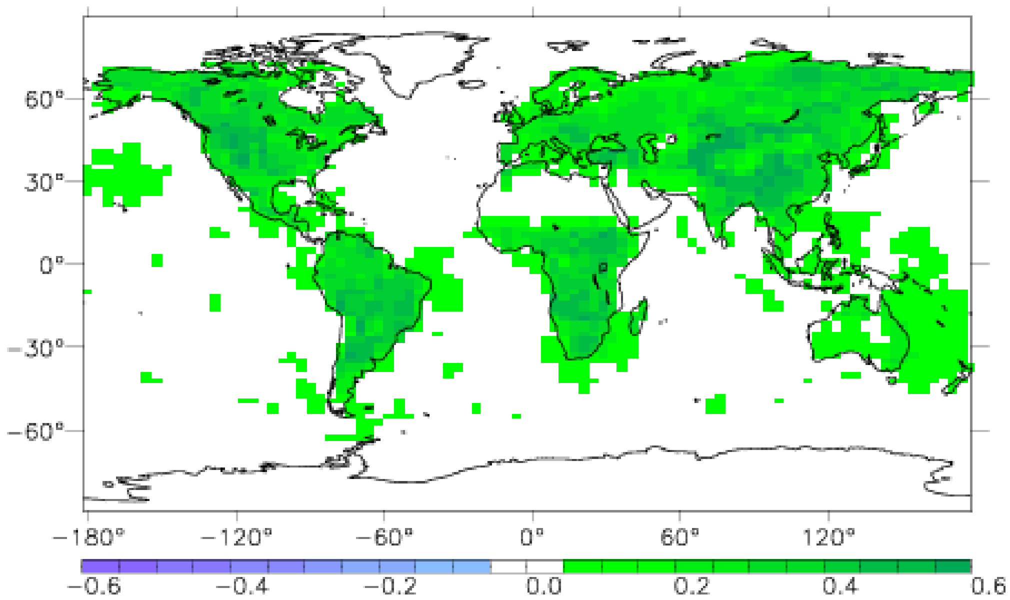


LSCE

Chevallier et al. (2012)

Full-rank results

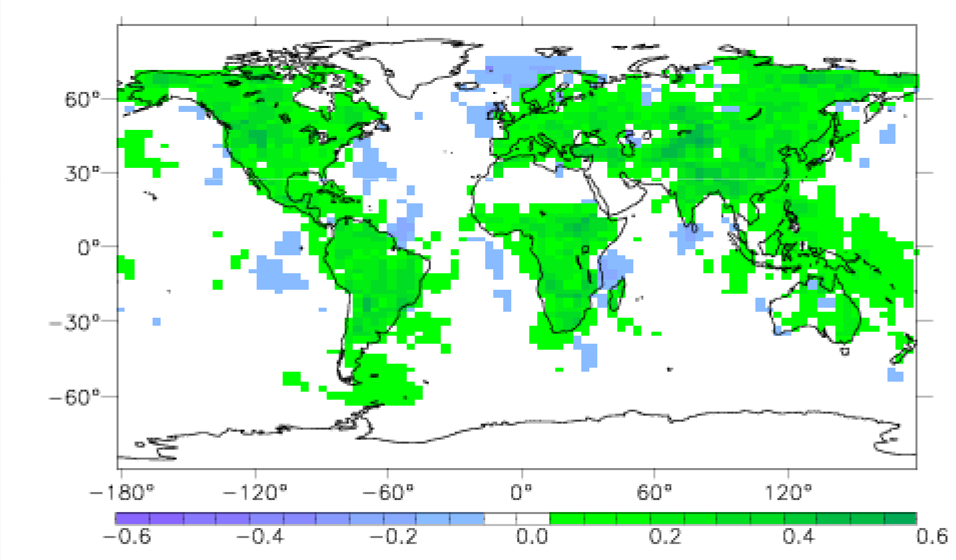
- Theoretical uncertainty reduction ($1 - \sigma_a / \sigma_b$) provided by the GOSAT satellite observations for the estimation of weekly-mean CO_2 surface fluxes.
  - σ_a is the posterior error standard deviation and σ_b is the prior error standard deviation

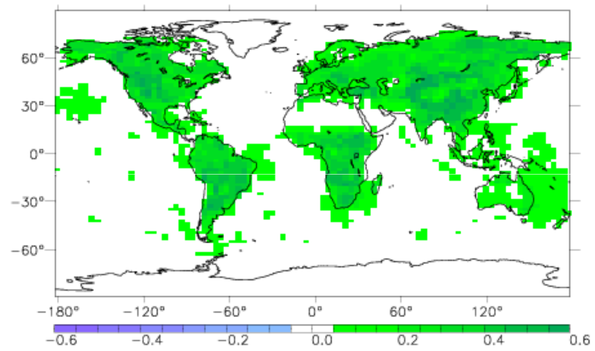


LSCE

Low temporal resolution

- Same as previous, but the inversion is performed with a simplified prior error temporal error covariance matrix.





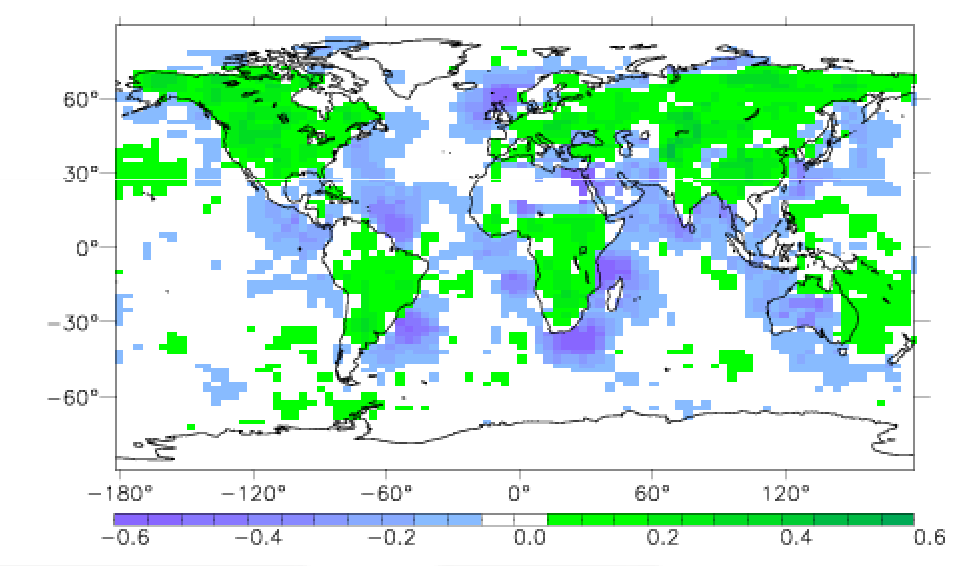
Reference (full-rank)

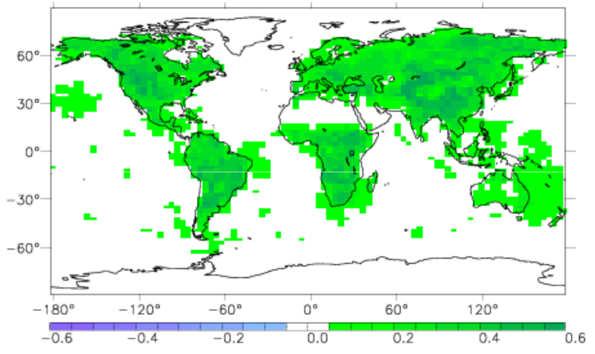
- The positive impact of the GOAT data is reduced compared to the optimal case. The maximum values become less than 50% and some negative values occur for ocean grid points.

LSCE

Low spatial resolution

- Same as reference, but the inversion is performed with a truncated prior error spatial error covariance matrix.
  - 200 leading evs (from 7000) kept for each week





Reference (full-rank)

  - The positive impact of the GOAT data is reduced to values less than 30% over land and becomes significantly negative (i.e. the inverted fluxes are worse than the prior fluxes) over some parts of the ocean basins.

LSCE

Parallelisation strategies: **y**

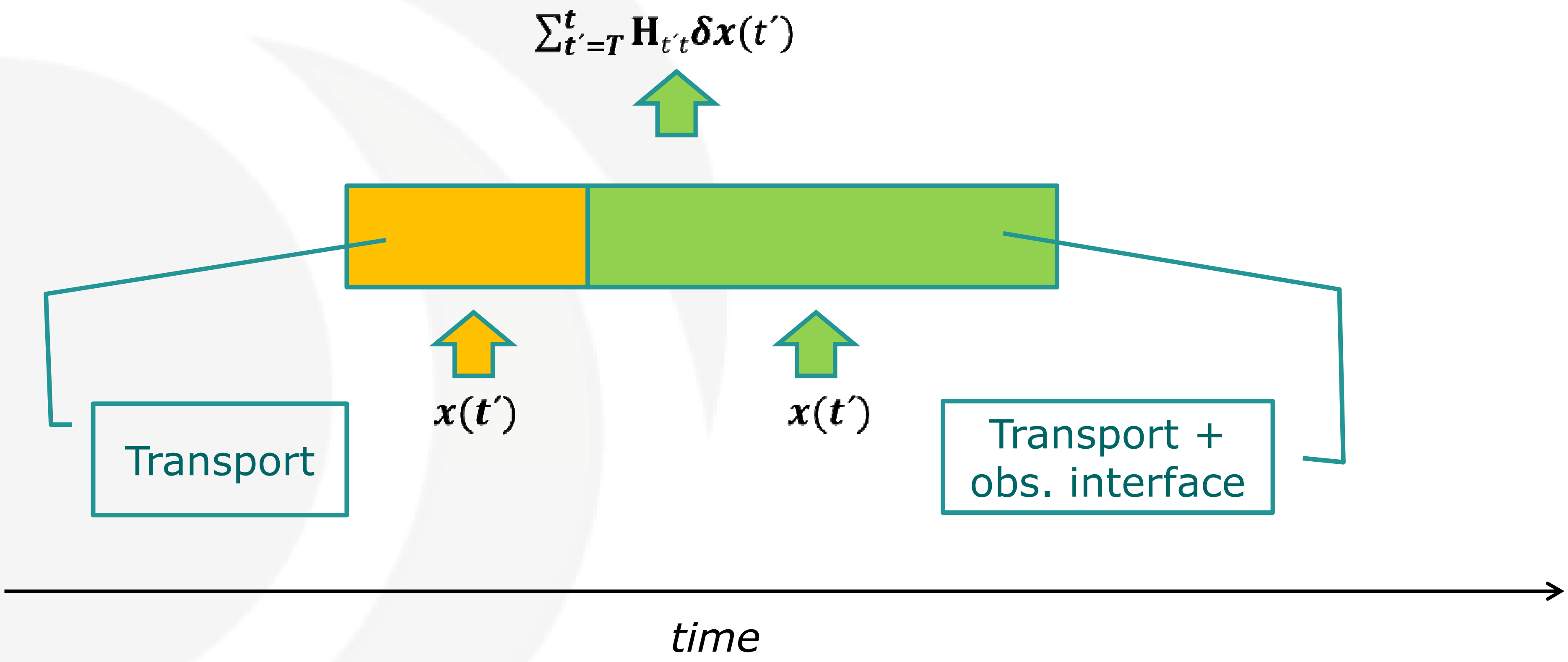
$$J(\mathbf{x}) = (\mathbf{x}-\mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{H}\mathbf{x}-\mathbf{y})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}-\mathbf{y})$$

- Solution 3: variational formulation; process observation sets in parallel
  - All observations are linked through atmospheric transport
  - Split the inversion window into overlapping segments
  - Replace $\boldsymbol{\delta y}(t) = \sum_{t'=-\infty}^{t} \mathbf{H}_{t't}\,\boldsymbol{\delta x}(t') + \boldsymbol{\varepsilon}$

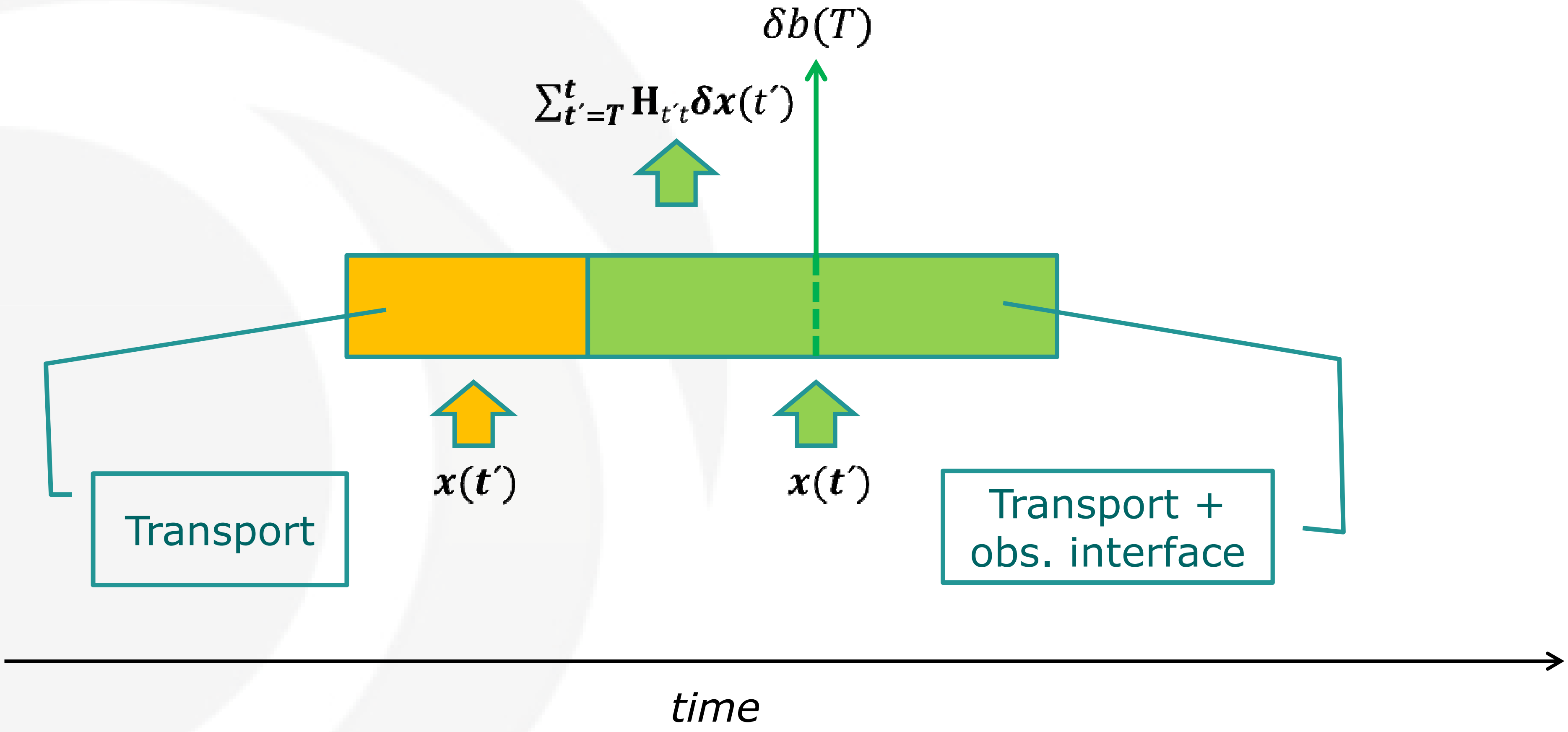
    with $\boldsymbol{\delta y}(t) = \sum_{t'=T}^{t} \mathbf{H}_{t't}\boldsymbol{\delta x}(t') + \delta b(T) + \boldsymbol{\varepsilon_2}$
  - $\delta b(T)$ = global scalar, mass increment of C in the atmosphere
  - The sum term can be computed in parallel
  - $\delta b(T)$ can be computed *afterwards*
  - *Ensemble method*, but without a statistical ensemble

- At the highest level, the inversion manages a single cost

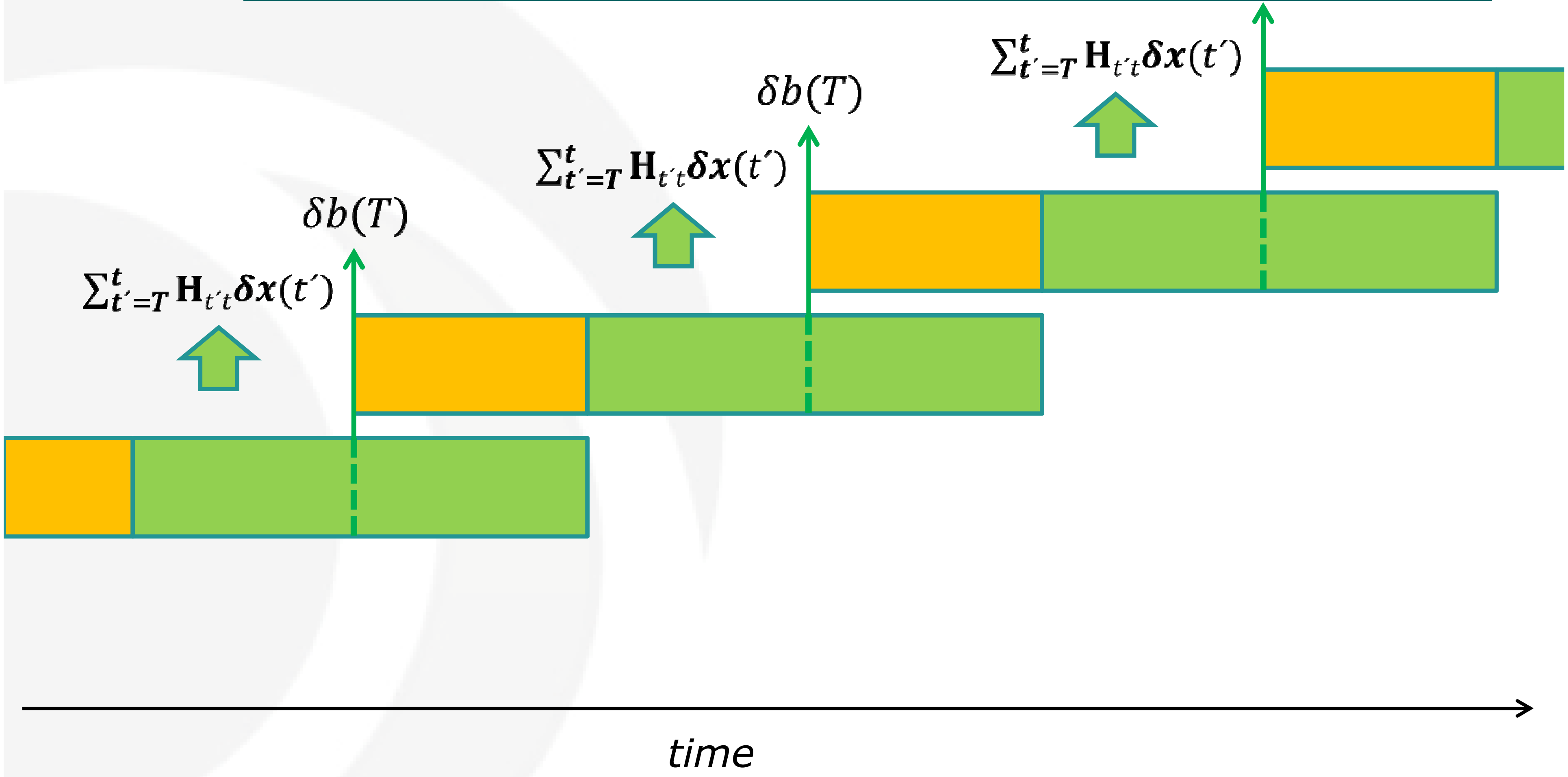
Scheme for the TL computation



Scheme for the TL computation

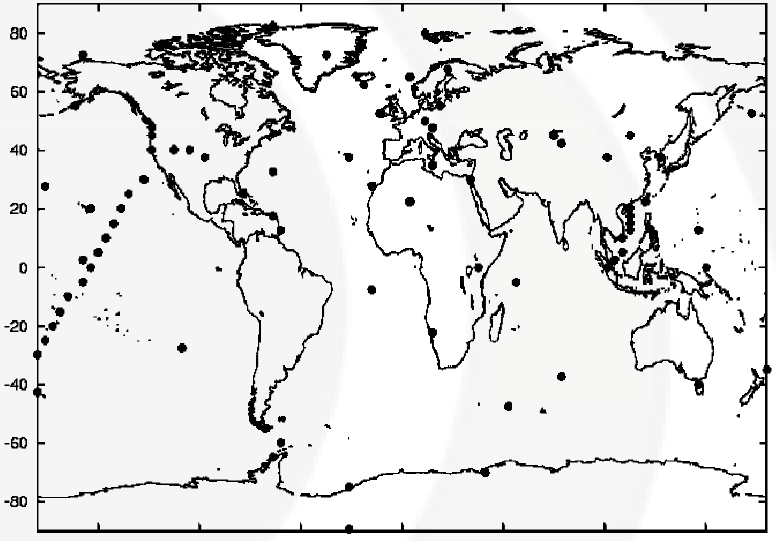


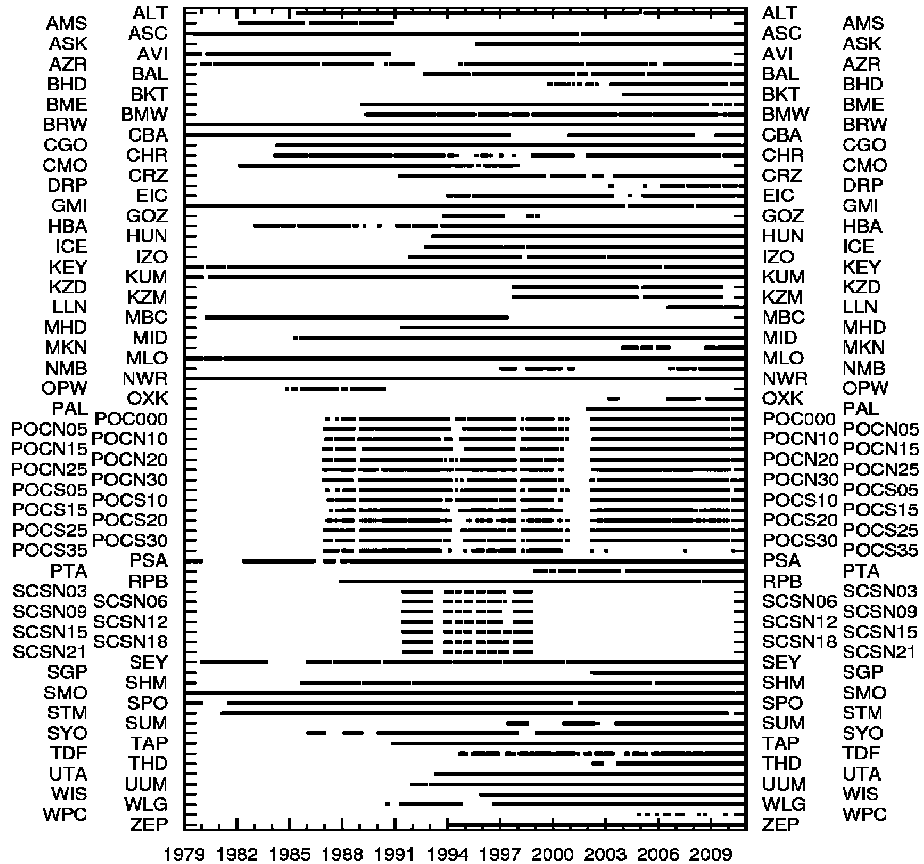
Scheme for the TL computation



Parallelisation strategies: **y**

- $\delta \boldsymbol{y}(t) = \sum_{t'=T}^{t} \mathbf{H}_{t't} \delta \boldsymbol{x}(t') + \delta b(T) + \boldsymbol{\varepsilon}_2$
  - Test on a 32-yr period, with a (3-mnth mixing) and (1-yr obs) configuration. Lanczos-based minimizer, 30 iterations

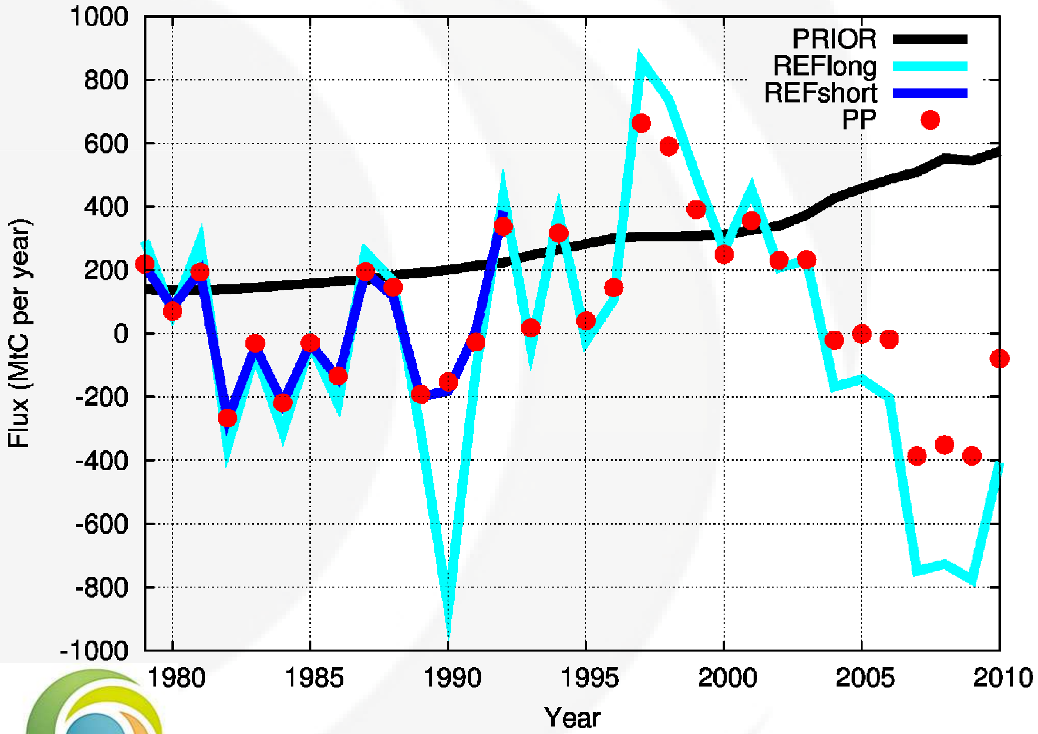




Flask data from

Parallelisation strategies: y

- $\boldsymbol{\delta y}(t) = \sum_{t'=T}^{t} \mathbf{H}_{t't} \boldsymbol{\delta x}(t') + \delta b(T) + \boldsymbol{\varepsilon_2}$
  - Test on a 32-yr period, with a (3-mnth mixing) and (1-yr obs) configuration. Lanczos-based minimizer, 30 iterations

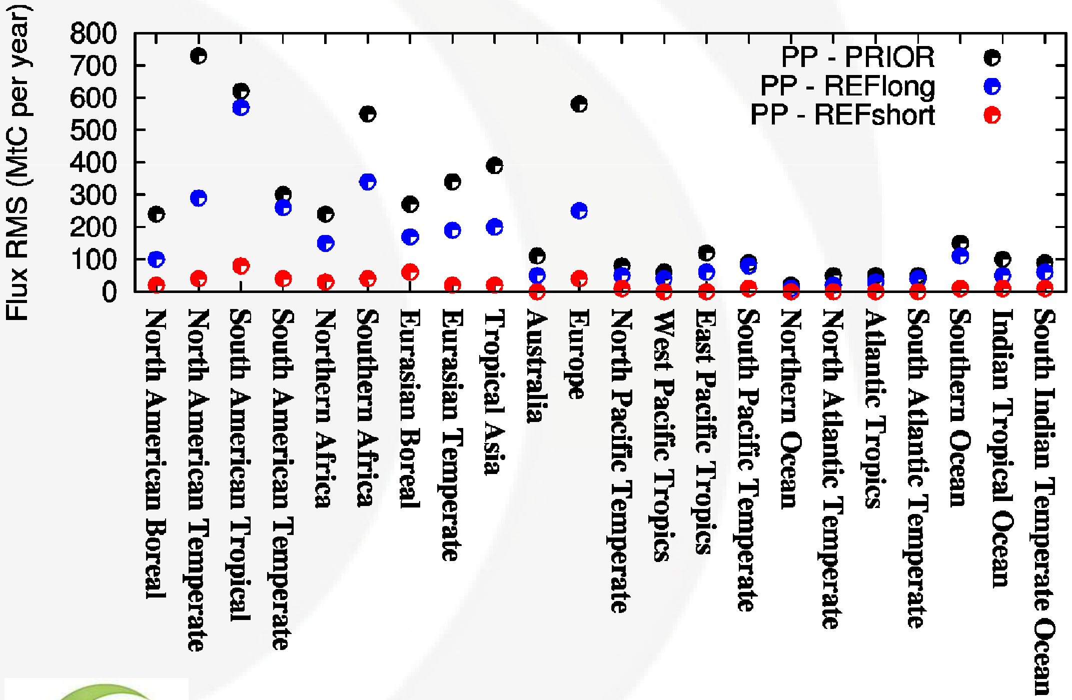


Time series of the annual total CO_2 fluxes (including fossil fuel) in region Tropical Asia.

PRIOR = prior fluxes
REFlong = inverted fluxes for the reference 1979-2010 inversion
REFshort = inverted fluxes for the reference 1979-1992 inversion
PP = 1979-2010 Parallelised inversion

Parallelisation strategies: **y**

- $\boldsymbol{\delta y}(t) = \sum_{t'=T}^{t} \mathbf{H}_{t't}\boldsymbol{\delta x}(t') + \delta b(T) + \boldsymbol{\varepsilon}_2$
  - Test on a 32-yr period, with a (3-mnth mixing) and (1-yr obs) configuration. Lanczos-based minimizer, 30 iterations



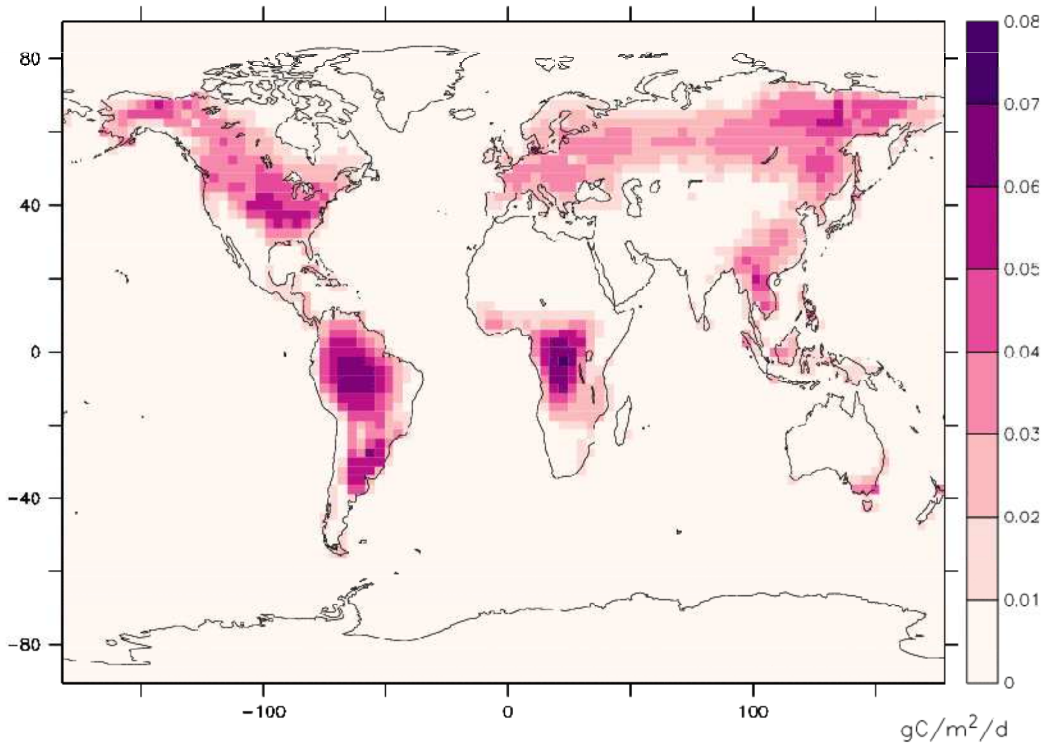
RMS differences for annual regional CO_2 fluxes between

- the PP inversion and its prior for the period 1979-2010 (***PP-PRIOR***),
- the 1979-2010 PP inversion and the 1979-2010 reference inversion for the same period (***PP-REFlong***)
- the 1979-2010 PP inversion and the 1979-1992 reference inversion for 1979-1991 (***PP-REFshort***)

~ 10% of increments

Parallelisation strategies: **y**

- $\boldsymbol{\delta y}(t) = \sum_{t'=T}^{t} \mathbf{H}_{t't}\boldsymbol{\delta x}(t') + \delta b(T) + \boldsymbol{\varepsilon}_2$
  - Test on a 32-yr period, with a (3-mnth mixing) and (1-yr obs) configuration. Lanczos-based minimizer, 30 iterations



RMS differences between the 1979-2010 PP inversion and the reference 1979-1992 inversion when computing the inverted weekly fluxes.

The statistics are computed for the 1979-1991 period.

< 11% of inversion increments

Parallelisation strategies: **y**

- $\boldsymbol{\delta y}(t) = \sum_{t'=T}^{t} \mathbf{H}_{t't} \boldsymbol{\delta x}(t') + \delta b(T) + \boldsymbol{\varepsilon_2}$
  - Test on a 32-yr period, with a (3-mnth mixing) and (1-yr obs) configuration. Lanczos-based minimizer, 30 iterations
  - Cost for 1 iteration (1 TL + 1 AD):

	Reference	// version
Nb of CPUs	8	32
CPU time	352h	192h
Wall clock time	44h	6h
Disk space	155GB	180GB
Swap space per CPU	2.5GB	20.5GB

LSCE

Conclusion

- Parallelisation is less natural for variational DA systems than for ensemble ones, but smart strategies are possible
- For flux inversion, our strategy allows
  - Benefitting from the high-resolution of the variational formulation
  - Keeping the physical and statistical consistency of the inversion over long periods (decades)
  - Improved numerical stability
  - 7x faster execution speed

LSCE