Weighting factors of satellite data in TM5-4dvar

33rd international TM5 meeting, December 2022

Johann Rasmus Nüß[1]
N. Daskalakis[1] and M. Vrekoussis[1,2,3]

[1]IUP-UB; [2]MARUM; [3]CARE-C

Outline

1. Objective
2. Background: TROPOMI, cost function and prior work
3. New work: Extended variance inflation
   - Intermediate Approach: ”optimized“ inflation
   - New approach: density-depended inflation
4. Open Questions for discussion

Objective

Objective

- Target: Use TROPOMI observations in TM5-4dvar
- Problem: Sheer number of observations ”breaks“ cost function
- Solutions:
  - Variance inflation
  - Observation Thinning
  - Super-Observations

Background: TROPOMI, cost function and prior work

TROPOMI observations

- **TROPO**spheric **M**onitoring **I**nstrument onboard of **S**entinel-**5** **P**recursor
- Local overpass time 13:30
- High resolution (up to $7 \times 7\,\text{km}^2$)

→ After quality filtering around 500.000 observations per day

- For comparison NOAA surface flasks deliver around 12 observations per day

Image: ESA

Cost function

$$J(x) = J_{obs}(x) + J_{prior}(x)$$

- Target of inversion is finding state x (here the global CO emissions) such that J is minimized
- J is quadratic function, i.e. this is a least-squares problem with many dimensions

Cost function

$$J_{obs}(x) = \frac{(F(x) - y)^2}{\sigma^2}$$

- Only looking at observational part
- F is model **and** its sampling to the times and places of the observations
- y are the measurements
- σ is observation error **and** sampling error of model

Cost function

$$J_{obs}(x) = \frac{(F(x) - y)^2}{l^2\sigma^2}$$

- Only looking at observational part
- F is model **and** its sampling to the times and places of the observations
- y are the measurements
- σ is observation error **and** sampling error of model

Variance Inflation

$$J_{obs}(x) = \frac{(F(x) - y)^2}{I^2 \sigma^2}$$

- Increase error by factor I to decrease weight in cost function
- Intended to capture various things, depending on setup e.g.:
  - Error correlation
  - Correlations between observations themselves (redundant information)
  - Model errors
  - Representativeness errors (scales not resolved by model)
- Applied in different ways: flat factor or additive value, possibly depending on other variables like latitude, observation height,..

Previous approach in TM5-4dvar

- Previous TM5-4dvar studies used form of variance inflation: Hooghiemstra et al. 2012, Krol et al. 2013, Nechita-Banda et al. 2018, Naus et al. 2022,..
- All used flat value of $I = \sqrt{50}$ for IASI and/or MOPITT
- Factor mostly arbitrary, was initially intended to represent observation density
- Satellite observations only used in zoom region $\rightarrow$ want global
- Older satellites (MOPITT, IASI,..) have lower observation density $\rightarrow$ cannot use that inflation for TROPOMI

New work: Extended variance inflation

Intermediate Approach

- Use condition $J_{obs,sat} = J_{obs,stations}$
- Leads to $I \approx 200$ (note there is no square root, i.e. cost is reduced 800 times stronger compared to previous studies!)
- Single observation has virtually 0 weight

$\rightarrow$ Problematic over oceans and for small scale signals e.g. from biomass burning

New approach

- Make inflation depend on actual observation density
- In preprocessing step calculate distances d between observations
- "Count" observations that are close to each other
- Calculate individual inflation factors for each observation

New approach: Counting adjacent observations

"Count" observations that are close to each other, different options:

- Box function (initial idea, count observations within certain radius r)

New approach: Counting adjacent observations

"Count" observations that are close to each other, different options:

- Box function (initial idea, count observations within certain radius r)
- $\sum \exp\left(-\frac{d}{r}\right)$ exponential decay, similar to spatial correlations within emissions in TM5-4dvar

New approach: Counting adjacent observations

"Count" observations that are close to each other, different options:

- Box function (initial idea, count observations within certain radius r)
- $\sum \exp\left(-\frac{d}{r}\right)$ exponential decay, similar to spatial correlations within emissions in TM5-4dvar
- $\sum \exp\left(-\frac{d^2}{2r^2}\right)$ Gaussian, suggested for this purpose in e.g. Liu and Rabier 2002, Daley et al 1993

New approach: Counting adjacent observations

"Count" observations that are close to each other, different options:

- Box function (initial idea, count observations within certain radius r)
- $\sum \exp\left(-\frac{d}{r}\right)$ exponential decay, similar to spatial correlations within emissions in TM5-4dvar
- $\sum \exp\left(-\frac{d^2}{2r^2}\right)$ Gaussian, suggested for this purpose in e.g. Liu and Rabier 2002, Daley et al 1993
- ...

New approach: Counting adjacent observations

"Count" observations that are close to each other, different options:

- Box function (initial idea, count observations within certain radius r)
- $\sum \exp\left(-\frac{d}{r}\right)$ exponential decay, similar to spatial correlations within emissions in TM5-4dvar
- $\sum \exp\left(-\frac{d^2}{2r^2}\right)$ Gaussian, suggested for this purpose in e.g. Liu and Rabier 2002, Daley et al 1993
- ...

Currently use $r = 200\,\text{km}$, the spatial correlations length for biomass burning, e.g. smallest distance on which model can change emissions. Is this meaningful?

New approach: Cost function behavior

How should this (weighted) count N affect the (local) cost J_{loc}:

- Without inflation: $J \propto N$

New approach: Cost function behavior

How should this (weighted) count N affect the (local) cost J_{loc}:

- Without inflation: $J \propto N$
- Want something that grows slower, e.g. $J \propto \sqrt{N}$

New approach: Cost function behavior

How should this (weighted) count N affect the (local) cost J_{loc}:

- Without inflation: $J \propto N$
- Want something that grows slower, e.g. $J \propto \sqrt{N}$
- Other possibilities: $J \propto \ln(N)$, ..

Open Questions for discussion

Open questions

- Counting function?
- Correlation length?
- Cost function behavior?
- How to do proper weighting to other datasets?
- Combine with flat factor?

Acknowledgments

- The computations were performed on the HPC cluster Aether at the University of Bremen, financed by DFG in the scope of the Excellence Initiative.
- The PhD position is paid for by the University Bremen.

- ... and of course thank You for your attention