7th Joint ECMWF/ESA/EUMETSAT Training on Atmospheric Composition, 10 October 2025

Jenny Stavrakou, Royal Belgian Institute for Space Aeronomy (BIRA-IASB)

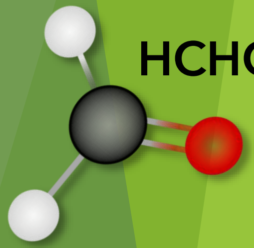
Inversion of sources and application to pyrogenic emissions and organic aerosol precursors

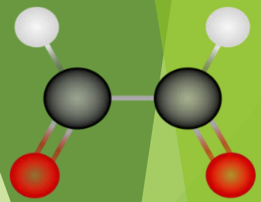
Outline

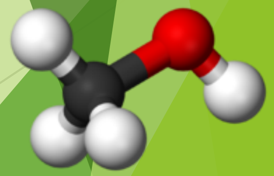
- ✓ What is the source inversion?
- ✓ Cost function
- ✓ Minimization

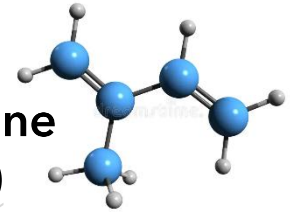
- ✓ How to solve source inversion problems…
  - idealized examples
  - real atmosphere

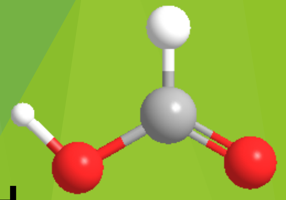
- ✓ Why VOCs matter?
- ✓ From HCHO to VOCs : an example from the real atmosphere
- ✓ Satellite observations of HCHO
- ✓ Top-down VOC emissions from pyrogenic and biogenic sources
- ✓ Satellite-based trends


HCHO

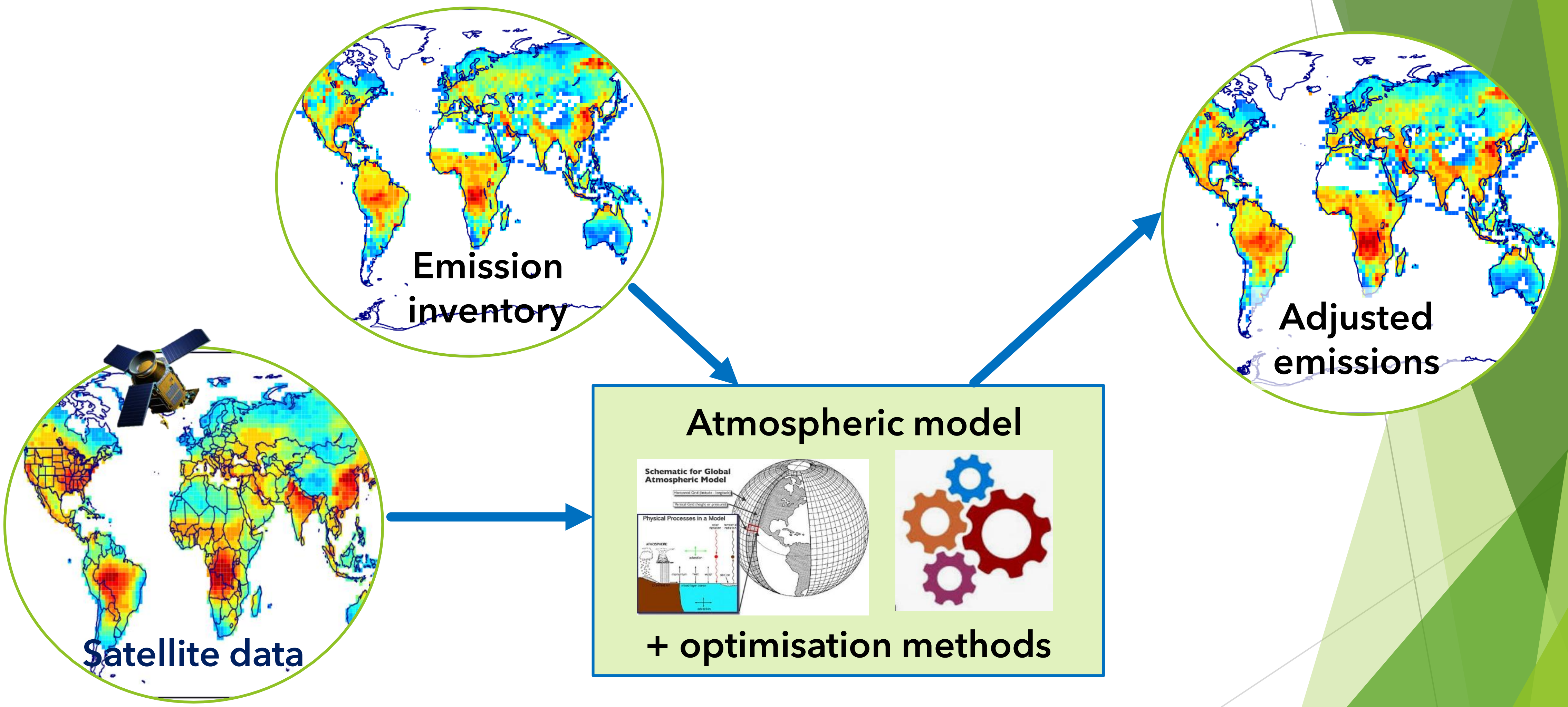

CHOCHO


CH_3OH


Isoprene (C_5H_8)


HCOOH

In the nutshell: source inversion uses observations to determine emissions



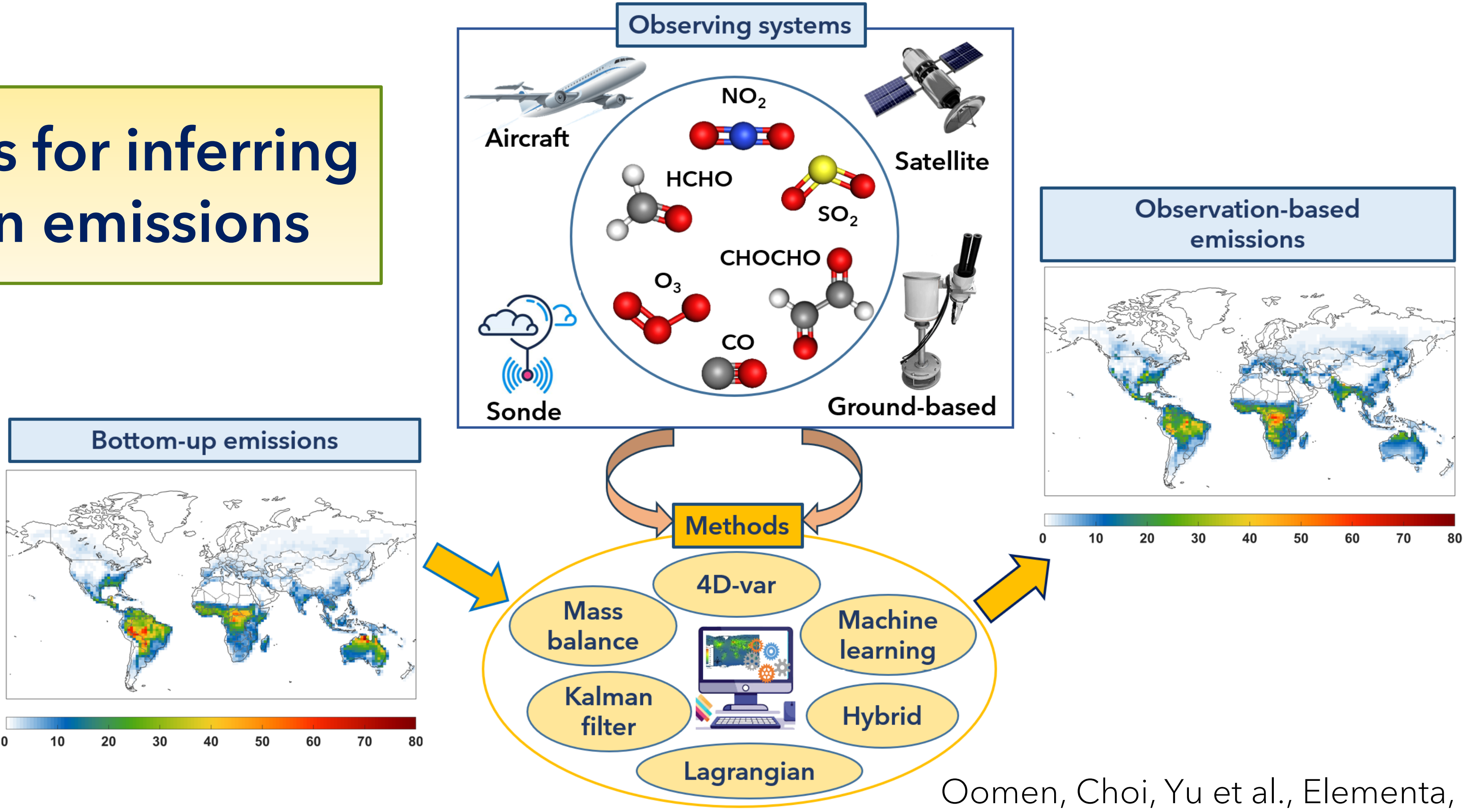
Ground-based: Offers continuous time series of species (in situ or column) at fixed locations

Sonde: Supplies in situ vertical profiles through the atmosphere

Aircraft: Captures atmospheric composition across altitude ranges during campaigns.

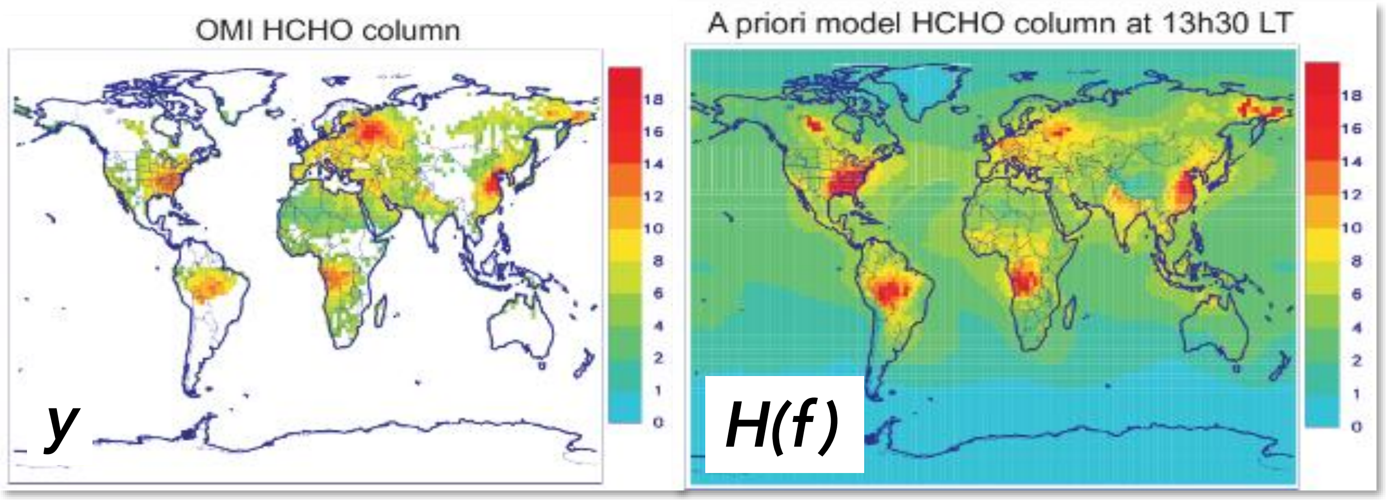
Satellite: Provides columns of trace gases over wide areas and extended periods. e.g. OMI, TROPOMI, CrIS, TES, GEMS, TEMPO, Sentinel-4

Ingredients for inferring top-down emissions



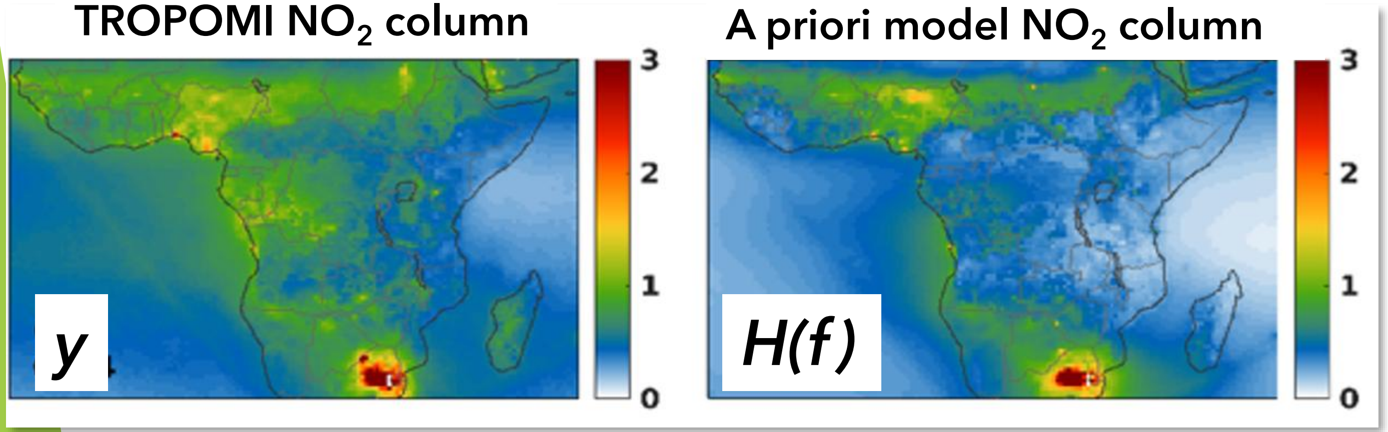
Oomen, Choi, Yu et al., Elementa, in review

Cost function *J* : measure of the mismatch between model and data



Stavrakou et al. 2015

10^{15} molec.cm^{-2}



Opacka et al. 2025

A priori emission distributions

$$G_0(x,t) = \sum_{j=1}^{m} \Phi_j(x,t)$$

Optimised emission distributions

$$G(x,t) = \sum_{j=1}^{m} \exp(f_j)\cdot\Phi_j(x,t)$$

$$J(f) = J_{OBS}(f) + J_B(f)$$

$$= \frac{1}{2}\left[(H(f)-y)^T \mathbf{E}^{-1}(H(f)-y)^T + \mathbf{f}^T \mathbf{B}^{-1} f\right]$$

f = *parameters to be optimized*

f_B = *first guess value for control parameters*

y = *atmospheric observations*

E = *matrix of errors on observations*

B = *matrix of errors on a priori fluxes*

Adjust emission distributions to best reproduce the observations

Minimum of the cost function?

The cost function *J* is an example of complex numerical algorithm consisting in a composition of differentiable mappings

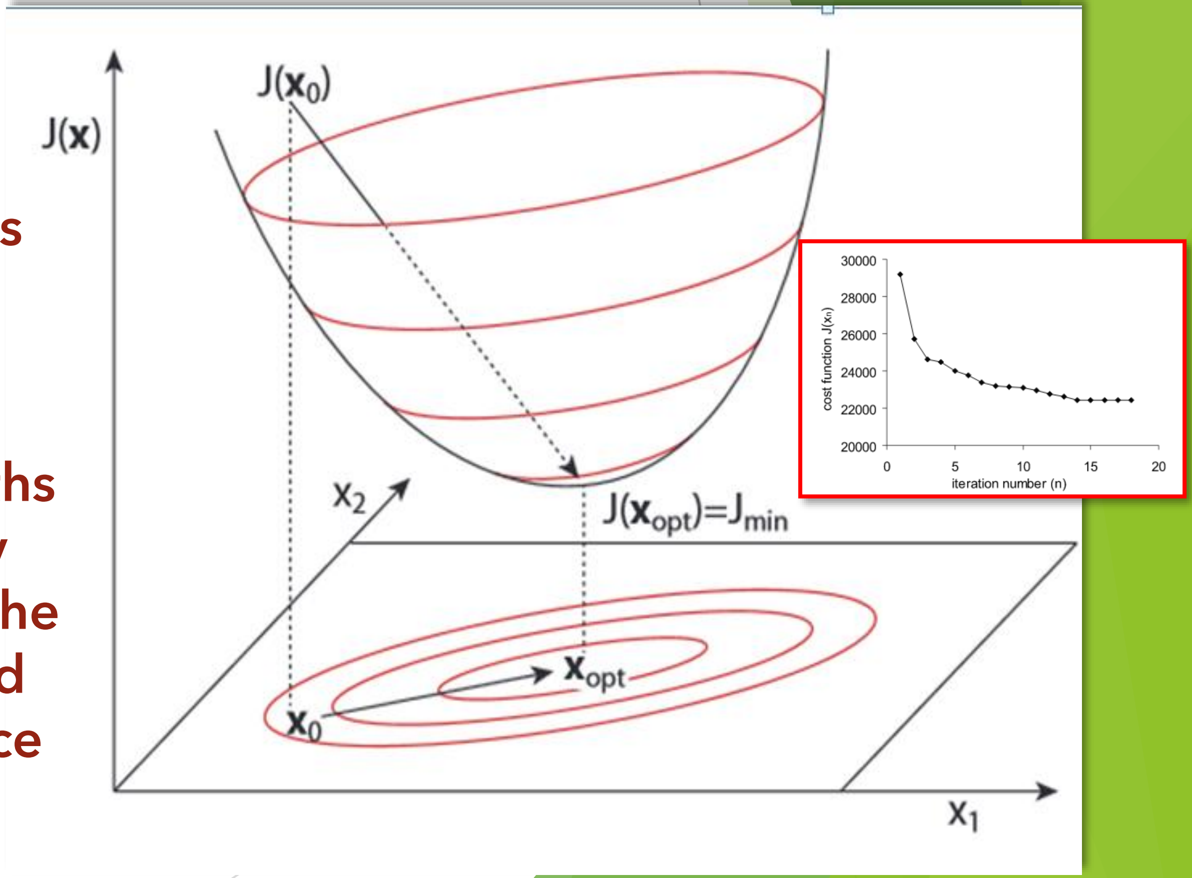
$$
\begin{aligned}
J(f) &= J_{obs}(f) + J_B(f) \\
&= \frac{1}{2}\sum_{i=1}^{p}(\boldsymbol{H}_i(f) - \boldsymbol{y}_i)^T \mathbf{E}^{-1}(\boldsymbol{H}_i(f) - \boldsymbol{y}_i) \\
&\quad + \frac{1}{2}(f - f_B)^T \mathbf{B}^{-1}(f - f_B),
\end{aligned}
$$

$$
\begin{aligned}
(\nabla J)_f = &\sum_{i=1}^{p}(D\boldsymbol{H}_i)_f^T \mathbf{E}^{-1}(\boldsymbol{H}_i(f) - \boldsymbol{y}_i) \\
&+ \mathbf{B}^{-1}(f - f_B)
\end{aligned}
$$

Minimum ? Gradient of *J*=0

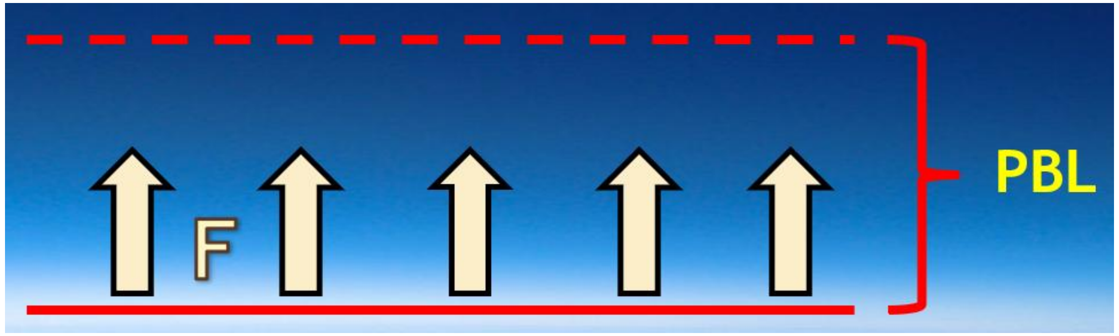
Example of cost function defined in a 2d-parameter space

J(x) projections on parameter plane: elliptic iso-cost lines with axis lengths determined by grad*J* and by the uncertainty and error covariance of x_{opt}

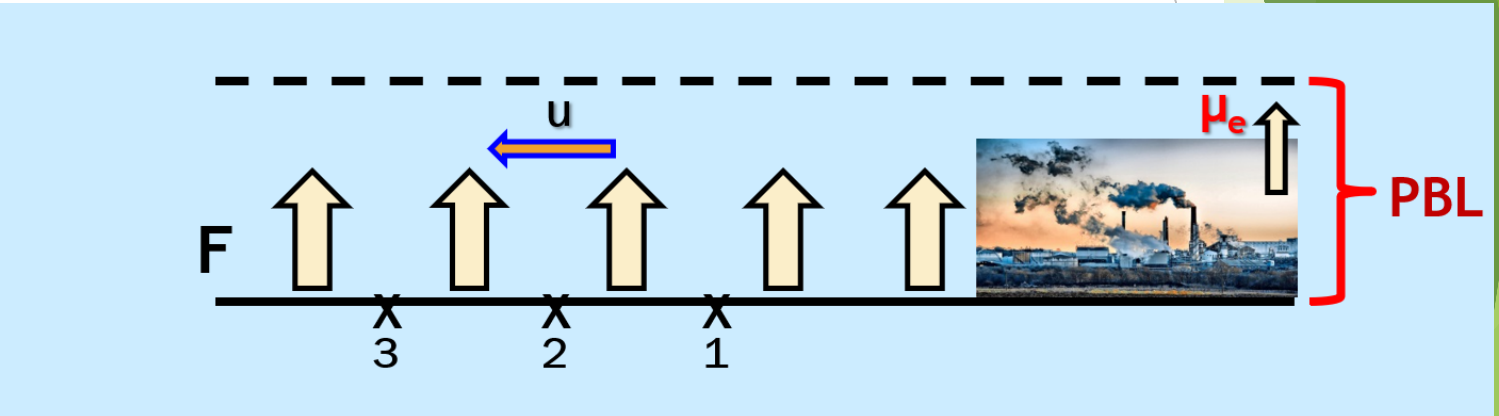


Learning how to solve an inversion problem!

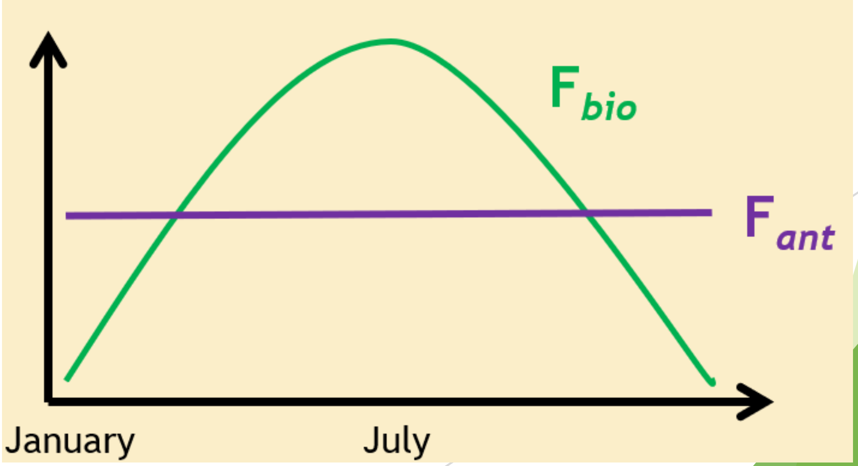
- ✓ No transport, constant sink, 1 source



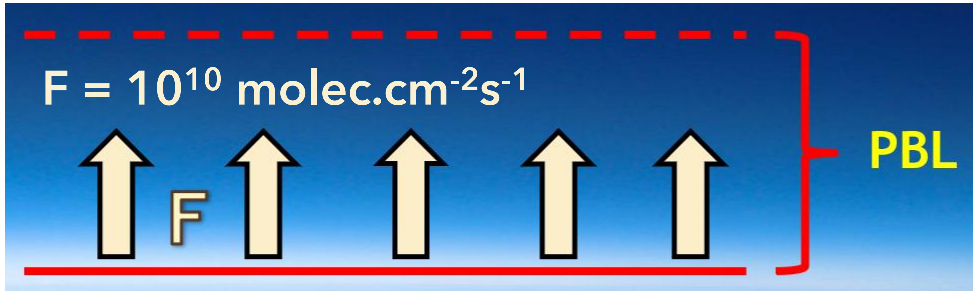
- ✓ Transport, constant sink, 2 sources



- ✓ Seasonal variation of emission sources, 2 sources



The simplest source inversion problem: *No transport, constant sink, 1 source*



- Uniform flux F of compound A
- A + OH → rate $k=10^{-11}$ $molec^{-1}cm^3\ s^{-1}$, $[OH]=10^6$ $molec.cm^{-3}$
- PBL : assumed well-mixed at all times ($z=10^5$ cm)
- Air number density (n) is constant ($n=2.5\times10^{19}$ $molec.cm^{-3}$)

Forward problem :
- What is the resulting vertical column V of A ?
- What is the average mixing ratio (μ)?

→ $V = F / (k \cdot [OH])$ → $V = 10^{15}$ $molec.cm^{-2}$
$\mu = V / n \cdot z$ → μ = 400 pptv

- 3 measurements μ^1_o=450 pptv, μ^2_o=390 pptv, μ^3_o=330 pptv, $\Delta\mu_o$= 20 pptv
- A priori flux estimate : $F_o = 5\times10^9$ $molec.cm^{-2}s^{-1}$, uncertainty=100%
- Uncertainties in n, z, k and [OH] lead to additional uncertainty on modelled mixing ratio of $\Delta\mu_m$=50 pptv

→ **Inverse problem :**
What is the top-down flux F?

- A priori mixing ratio : $\mu_o = F_o / n \cdot z \cdot k \cdot [OH]$ = 200 pptv
- Set: $F = F_o . f$, f= dimensionless adjustable parameter

$$J = \frac{1}{2}\sum_1^3\left(\frac{\mu_{\mathrm{mod}}(f) - \mu_o^i}{\Delta\mu}\right)^2 + \frac{1}{2}\left(\frac{f - f_0}{\Delta f}\right)^2$$

f_o=1

Δf = 1 (=100%)

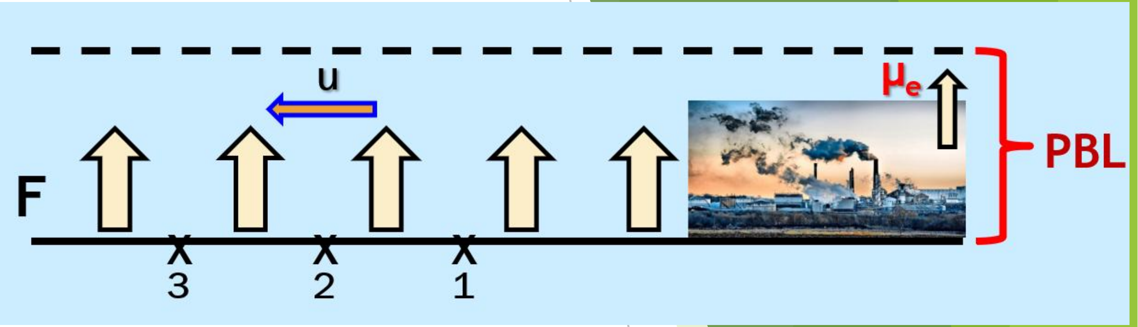
$(\Delta\mu)^2 = (\Delta\mu_o)^2 + (\Delta\mu_m)^2 = (20^2 + 50^2)$ $pptv^2$

$$\frac{\partial J}{\partial f} = \frac{200}{\Delta\mu^2}\sum_1^3 200\cdot(200\cdot f - \mu_o^i) + (f-1) = 0 \Rightarrow f = 1.927$$

→ $F = F_o . f$ =
= *9.64 x10^9* $molec.cm^{-2}s^{-1}$
Very close to the true flux!

A bit more complicated: *Constant transport, constant sink, 2 sources to retrieve*

- Background source $F=10^{10}$ molec.cm^{-2}s^{-1} + local source generating an enhancement $\mu_e=1$ ppbv of compound A
- Constant horizontal wind : $u = 5$ m/s
- Measurements at sites 1,2,3 downwind, at $d_1=200$ km, $d_2=500$ km and $d_3=1000$ km



Forward problem: What is the mixing ratio (μ) at each site?

- $\mu = \mu_b + \mu_e$, with $\mu_b = F / (n \cdot z \cdot k \cdot [OH]) = 400$ pptv (as before)
- $\mu_e = 1$ ppbv $\cdot e^{-(d/u)\cdot k \cdot [OH]} \rightarrow \mu_e(1) = 670$ pptv, $\mu_e(2) = 368$ pptv, $\mu_e(3) = 135$ pptv

$\mu(1) = 1070$ pptv
$\mu(2) = 768$ pptv
$\mu(3) = 535$ pptv

- ✓ Assume : $\mu^1_o=1100$ pptv, $\mu^2_o=750$ pptv, $\mu^3_o=500$ ppt
- ✓ Combined measurement/model uncertainty : $\Delta\mu=100$ pptv
- ✓ $F_o = 5\times10^9$ molec.cm^{-2}s^{-1}, uncertainty=100%
- ✓ A priori mixing ratio enhancement : $\mu_{eo} = 2$ ppbv, 100% error

Inverse problem :
What are the flux F and mixing ratio μ_e?

- $F = F_o . f_1$, $\mu_e = \mu_{eo} . f_2$ *with a priori* $f_1 = f_2 = 1$, $\Delta f_1 = \Delta f_2 = 1$

$$J = \frac{1}{2}\sum_1^3\left(\frac{\mu_{mod}(f)-\mu_o^i}{\Delta\mu}\right)^2 + \frac{1}{2}\left(\frac{f_1-f_0}{\Delta f_1}\right)^2 + \frac{1}{2}\left(\frac{f_2-f_0}{\Delta f_2}\right)^2$$

$$\mu^i_{mod}(f) = (F_o . f_1 / (n . z . k . [OH])) + (\mu_{eo} . f_2 \cdot e^{-(d_i/u)\cdot k\cdot[OH]}) \Rightarrow$$

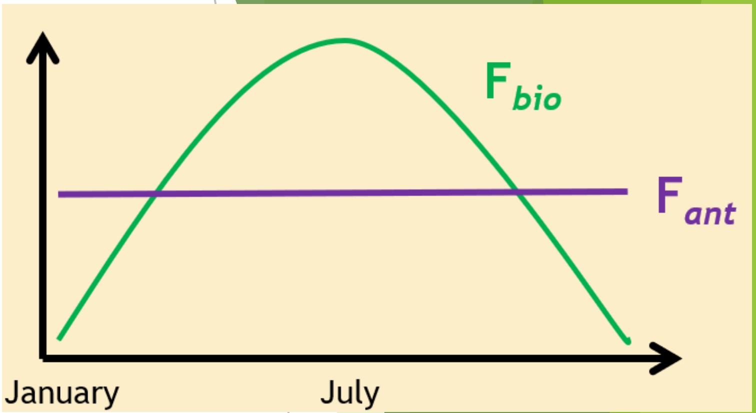
$$\mu^1_{mod} = 200 \cdot f_1 + 1340 . f_2$$
$$\mu^2_{mod} = 200 \cdot f_1 + 736 . f_2$$
$$\mu^3_{mod} = 200 \cdot f_1 + 270 . f_2$$

$$\frac{\partial J}{\partial f_1} = \frac{\partial J}{\partial f_2} = 0 \Rightarrow f_1 = 1.515, f_2 = 0.603$$

$F = 7.575 \times 10^9$ molec.cm^{-2} s^{-1} & $\mu_e = 1.206$ ppbv

One step further : *Retrieve 2 sources with different seasonality*

- Compound A has a constant anthropogenic source (F_{ant}), and a biogenic source (F_{bio}) with a seasonality peaking in summer
  - $F_{ant} = 10^{10}$ molec.cm^{-2}s^{-1}
  - F_{bio}(spring) = 10^{10} molec.cm^{-2}s^{-1}
  - F_{bio}(summer) = $3\text{x}10^{10}$ molec.cm^{-2}s^{-1}
- μ (*spring*) = (F_{ant} + $F_{bio\ (spring)}$) / (n · z · k · [OH]) = 0.8 ppbv
- μ (*summer*) = (F_{ant} + $F_{bio\ (summer)}$) / (n · z · k · [OH]) = 1.6 ppbv



- Assume μ^1_o=0.8 ppbv in spring, μ^2_o =1.6 ppbv in summer, Δμ =0.2 ppbv
- No correlation btw anthropogenic and biogenic source
- Correlation btw errors on biogenic source in spring and summer

<u>Inverse problem :</u>
What is the top-down flux F_{ant} and F_{bio} ?

- $F_{ant} = 2\text{x}10^{10} \cdot f_1$, F_{bio}(spring) = $10^{10} \cdot f_2$, F_{bio}(summer) = $2\text{x}10^{10} \cdot f_3$ *with* $f_1 = f_2 = f_3 = 1$, $\Delta f_1 = 1$, $\Delta f_2 = \Delta f_3 = 2/\sqrt{3}$
- Correlation between f_2 *and* f_3 is *c*=0.5

$$J = \frac{1}{2}\sum_i \left(\frac{\mu_{\text{mod}}(f) - \mu^i_o}{\Delta\mu}\right)^2 + \frac{1}{2}\sum_{j=1}^{3}\sum_{k=1}^{3}(f_j - 1)B^{-1}_{jk}(f_k - 1)$$

$$B = \begin{pmatrix} \Delta f_1^2 & 0 & 0 \\ 0 & \Delta f_2^2 & c\Delta f_2 \Delta f_3 \\ 0 & c\Delta f_2 \Delta f_3 & \Delta f_3^2 \end{pmatrix}$$

$$\frac{\partial J}{\partial f_1} = \frac{\partial J}{\partial f_2} = \frac{\partial J}{\partial f_3} = 0 \Rightarrow f_1 = 0.66, f_2 = 0.76, f_3 = 1.31$$

$F_{ant} = 1.33\text{x}10^{10}$, $F_{bio\ (spring)} = 0.76\text{x}10^{10}$, $F_{bio\ (summer)} = 2.6\text{x}\ 10^{10}$

μ^1_{mod} = 0.839 ppbv, μ^2_{mod} = 1.579 ppbv

Very close to the real values!

And in the real atmosphere?

Same formula!

$$J = \frac{1}{2}(H(f) - y)^T E^{-1}(H(f) - y) + \frac{1}{2}(f - f_B)^T B^{-1}(f - f_B)$$

- *y* : chemical observations from satellite, ground-based, airborne…
- *H(f)* : global or regional CTM
- E : errors of the retrievals (systematic/random)
- B : based on spatiotemporal correlations

In numbers…

- Global CTM (2°x2.5°, 144 longitudes x 90 latitudes)
- Optimize monthly fluxes of compound A
- 3 emission sources (e.g. biogenic, pyrogenic, anthropogenic)
- # *f*'s : 12 x 144 x 90 x 3 x 0.3 ~140,000 parameters to optimize

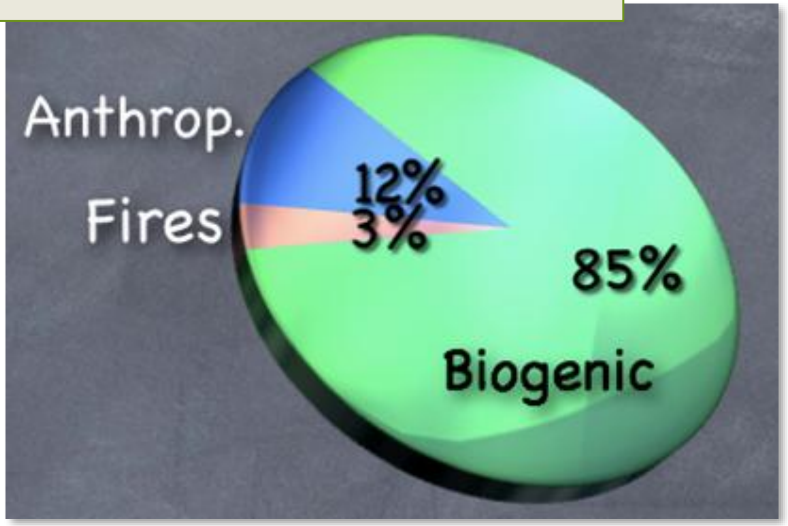
Even when we use satellite measurements to constrain the fluxes, we have fewer observations (12 x 144 x 90 x 0.3 ~ 50,000) than parameters to optimize → *Underdetermined problem*

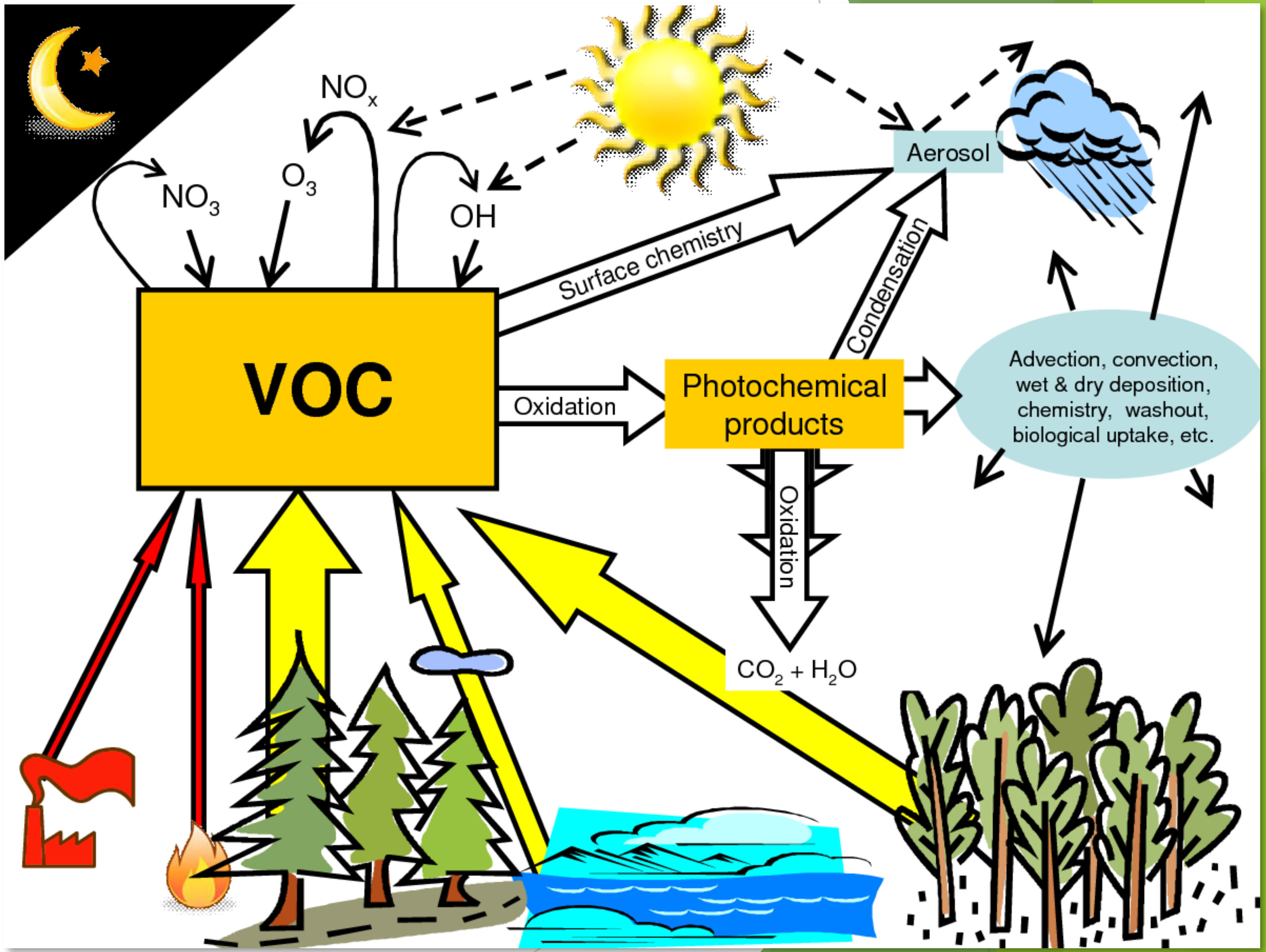
Reduce the number of effective variables by omitting cells with very low a priori flux, and by using a correlation setup in matrix B

NMVOCs : why do they matter?

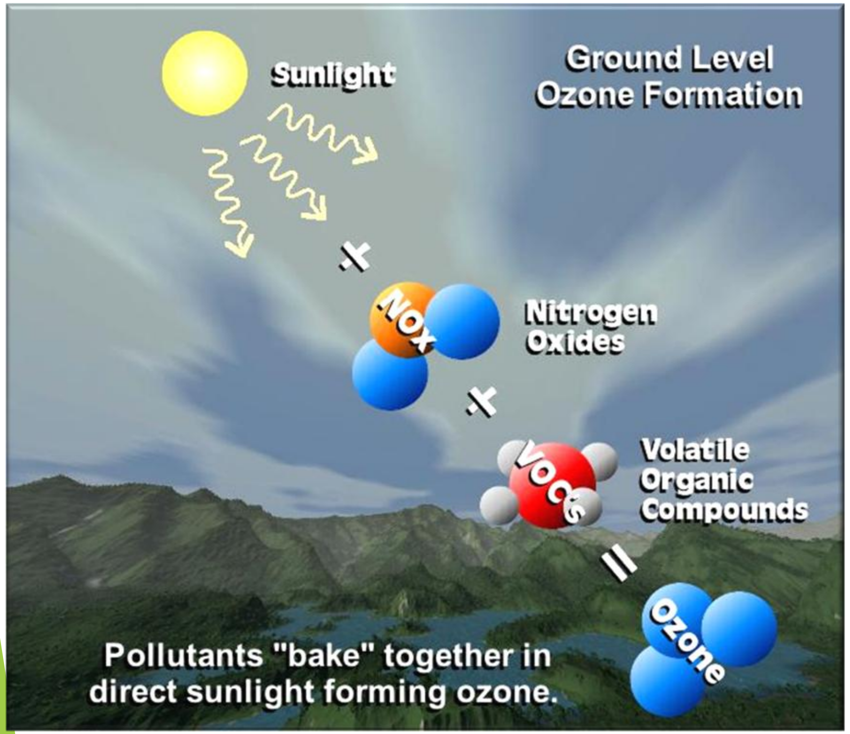
- Broad variety of fastly reacting species
- Influence atmospheric composition : contribute to O_3 & PM formation
- Influence radiation, clouds, air quality
- Influence human health & climate : precursors of SOA, affect GHG

Annual emission : ~1000 TgC (isoprene~500 TgC), but large uncertainties!

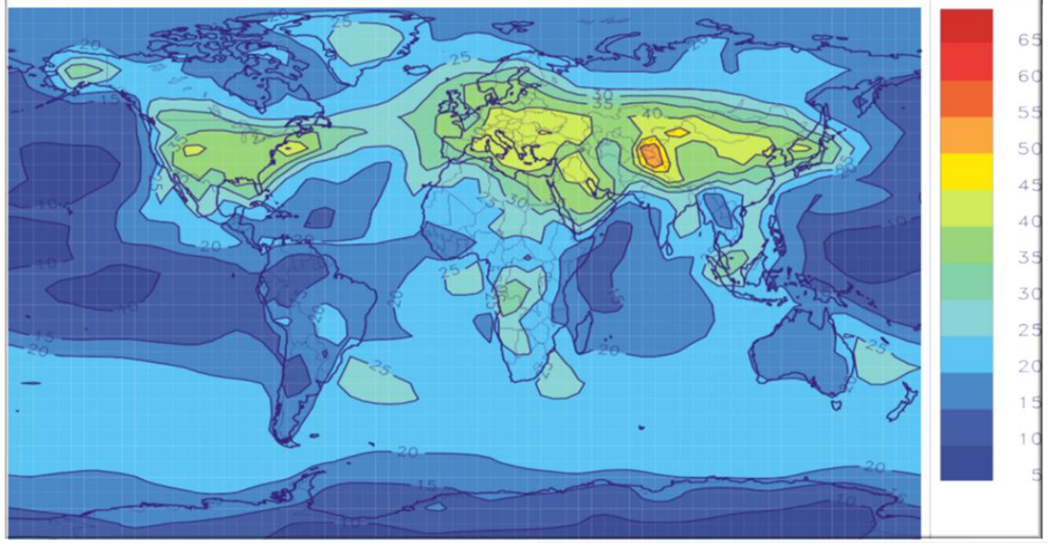




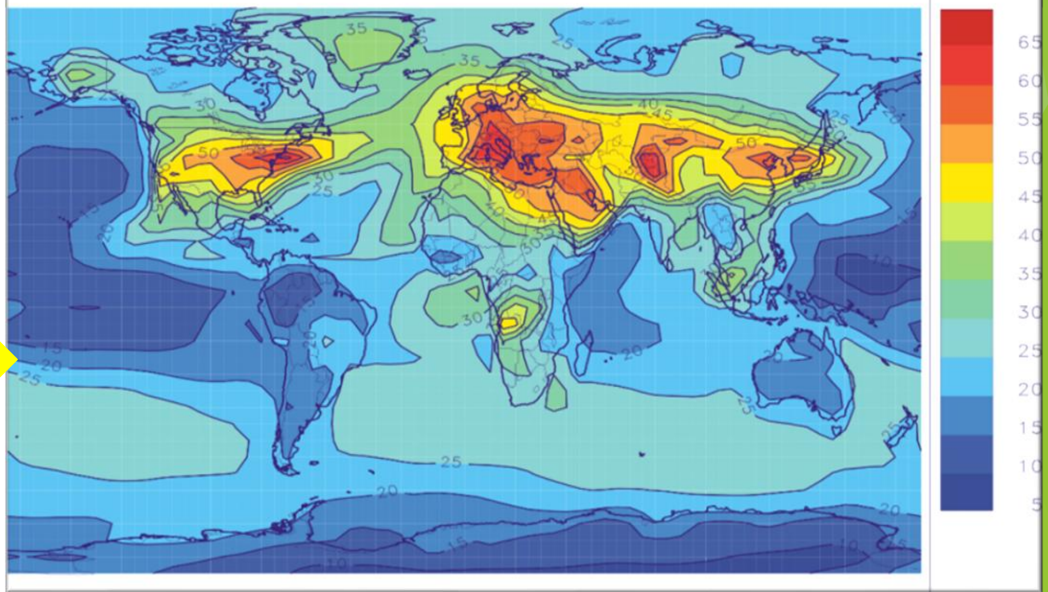
Impact of NMVOC precursors on surface ozone levels



Model simulation neglecting NMVOCs



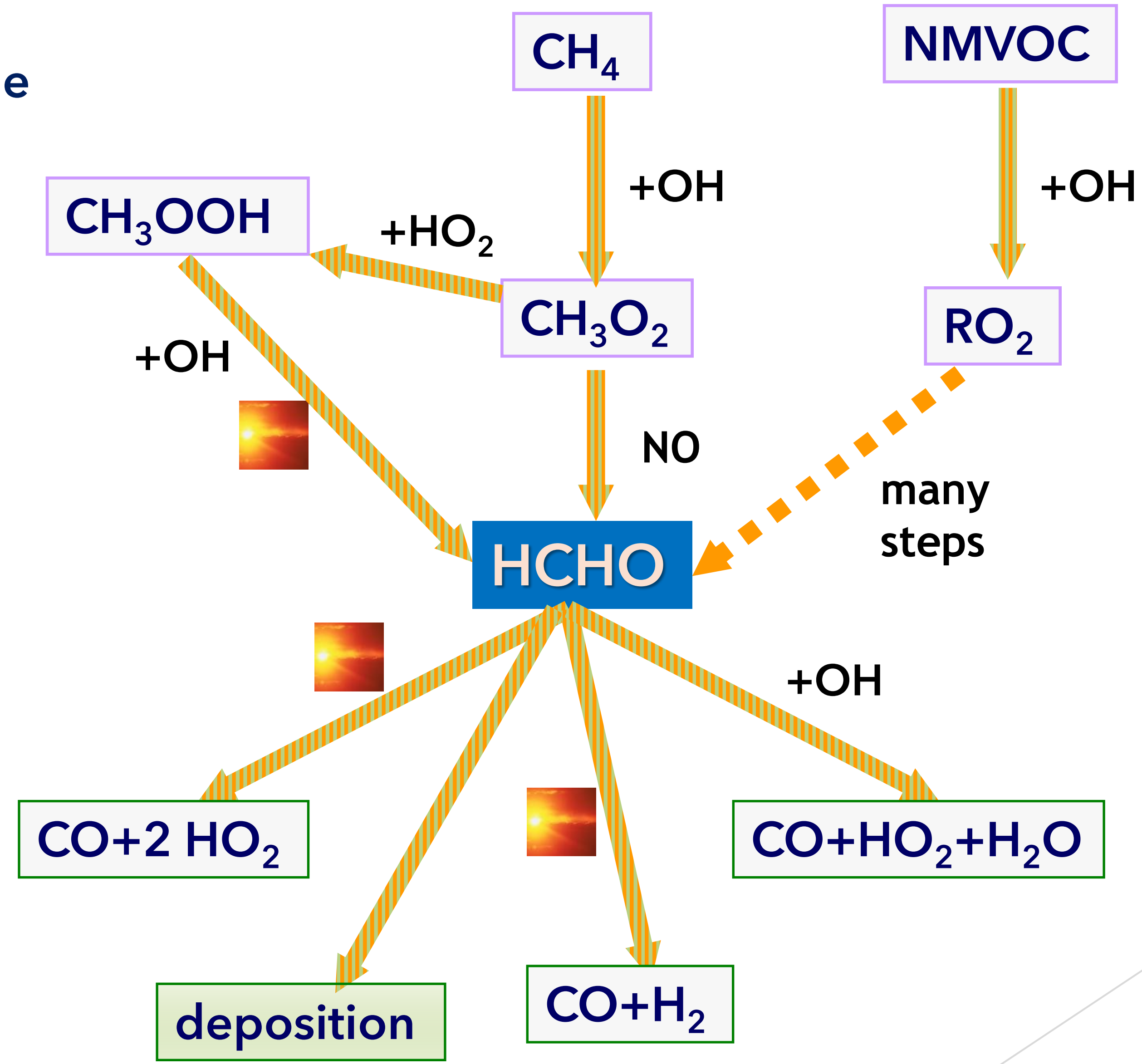
Model simulation with NMVOCs

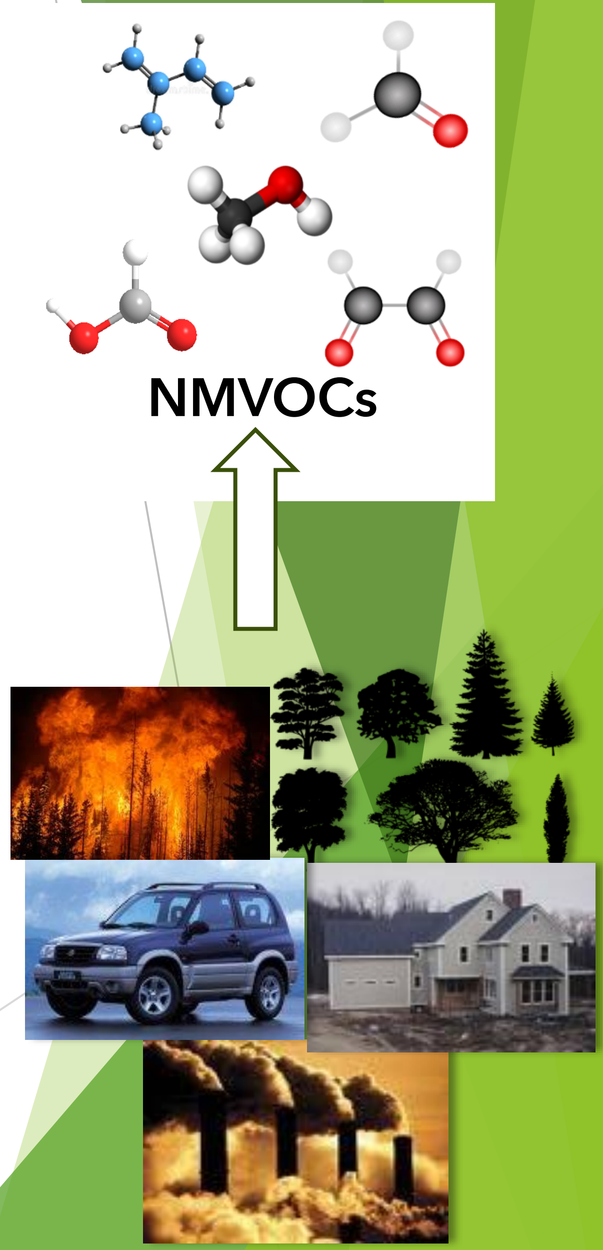


In the presence of VOCs → strong increases of surface ozone levels in the Northern Hemisphere, up to factor of 2 increase in eastern US, Europe and China

HCHO : the most abundant carbonyl in the atmosphere

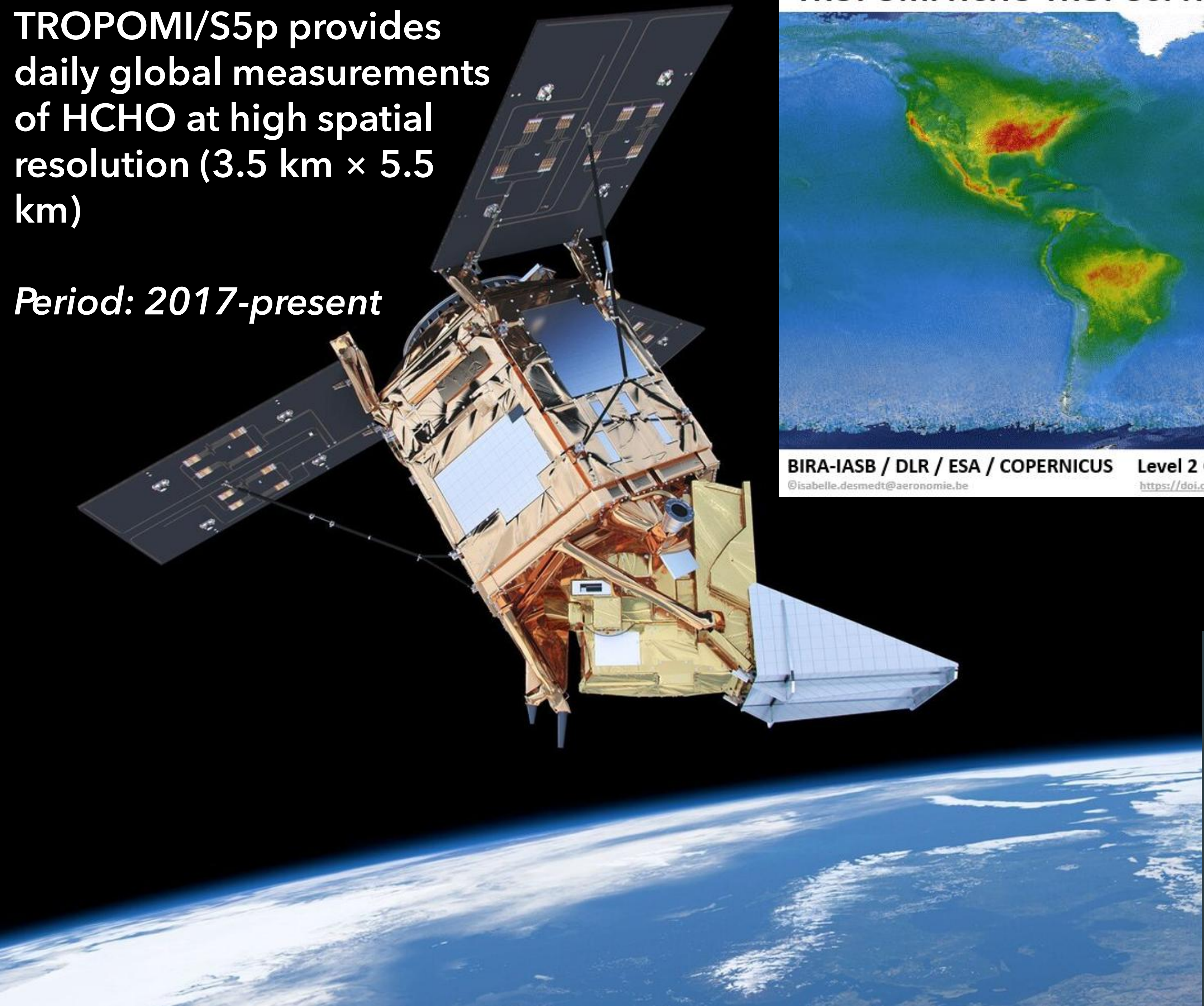
- ✓ Short-lived - lifetime on the order of a few hours
- ✓ Directly emitted from fossil fuel combustion and biomass burning
- ✓ Also formed as a high-yield secondary product in the CH_4, and NMVOC oxidation

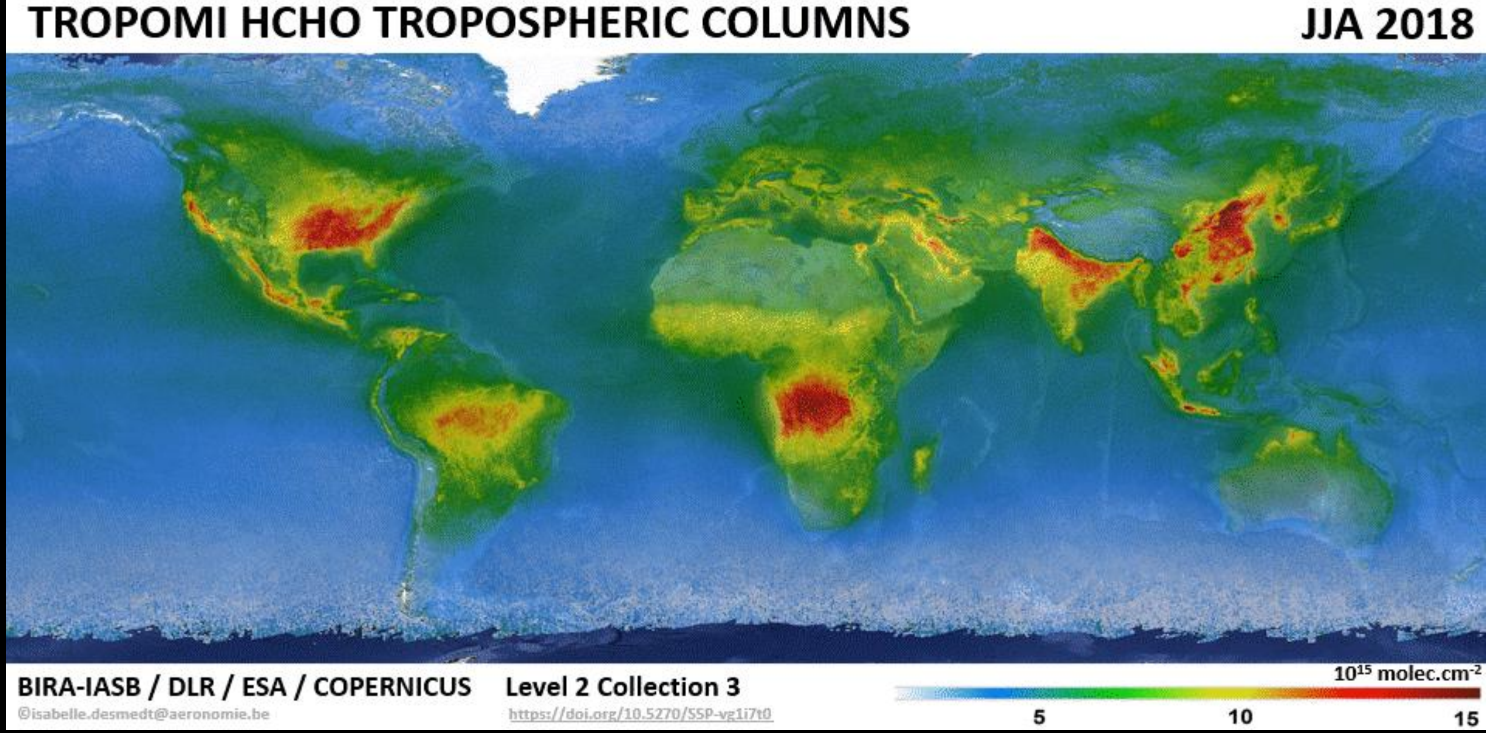




TROPOMI/S5p provides daily global measurements of HCHO at high spatial resolution (3.5 km × 5.5 km)

Period: 2017-present





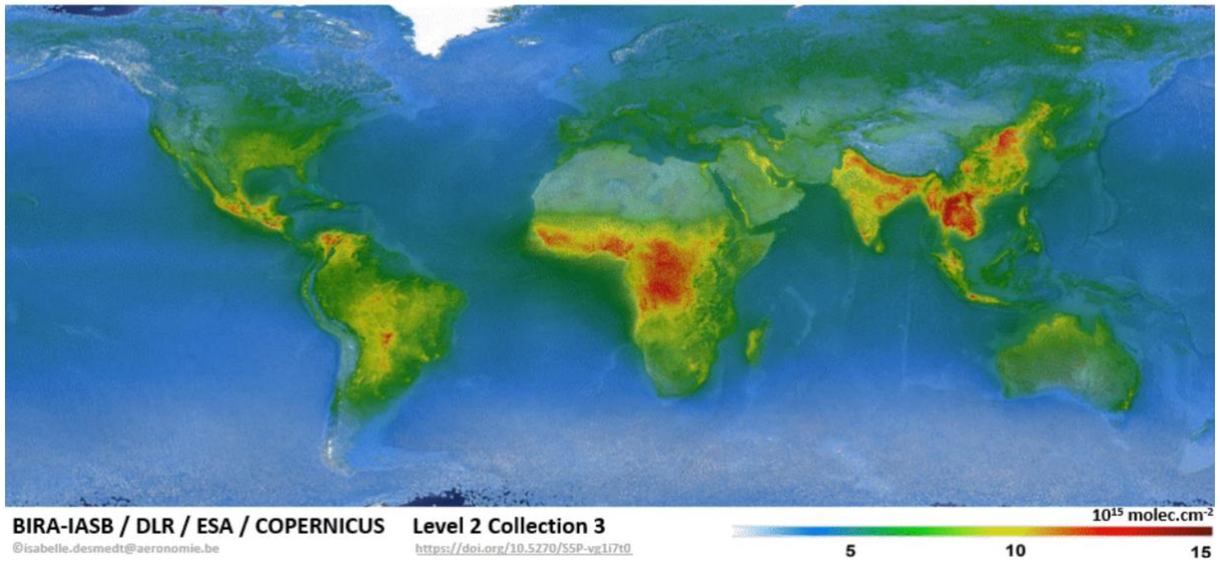
Infer top-down

- pyrogenic VOC emissions
- biogenic VOC emissions

based on MAGRITTE global CTM and its adjoint (4D-Var)

Focus on pyrogenic and biogenic sources

TROPOMI HCHO TROPOSPHERIC COLUMNS JJA 2020

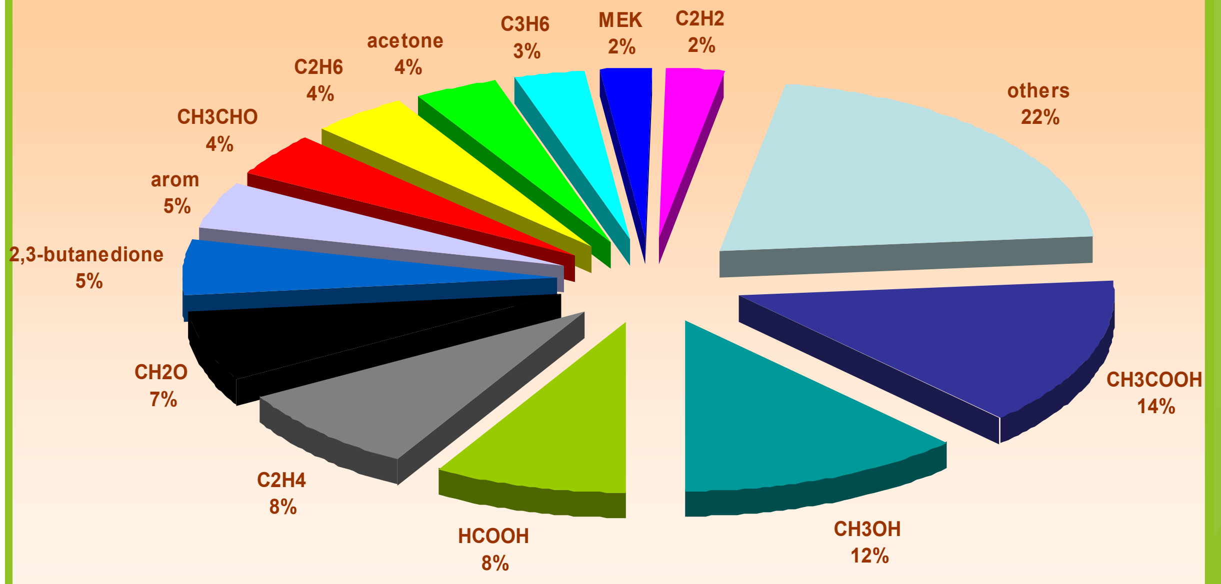


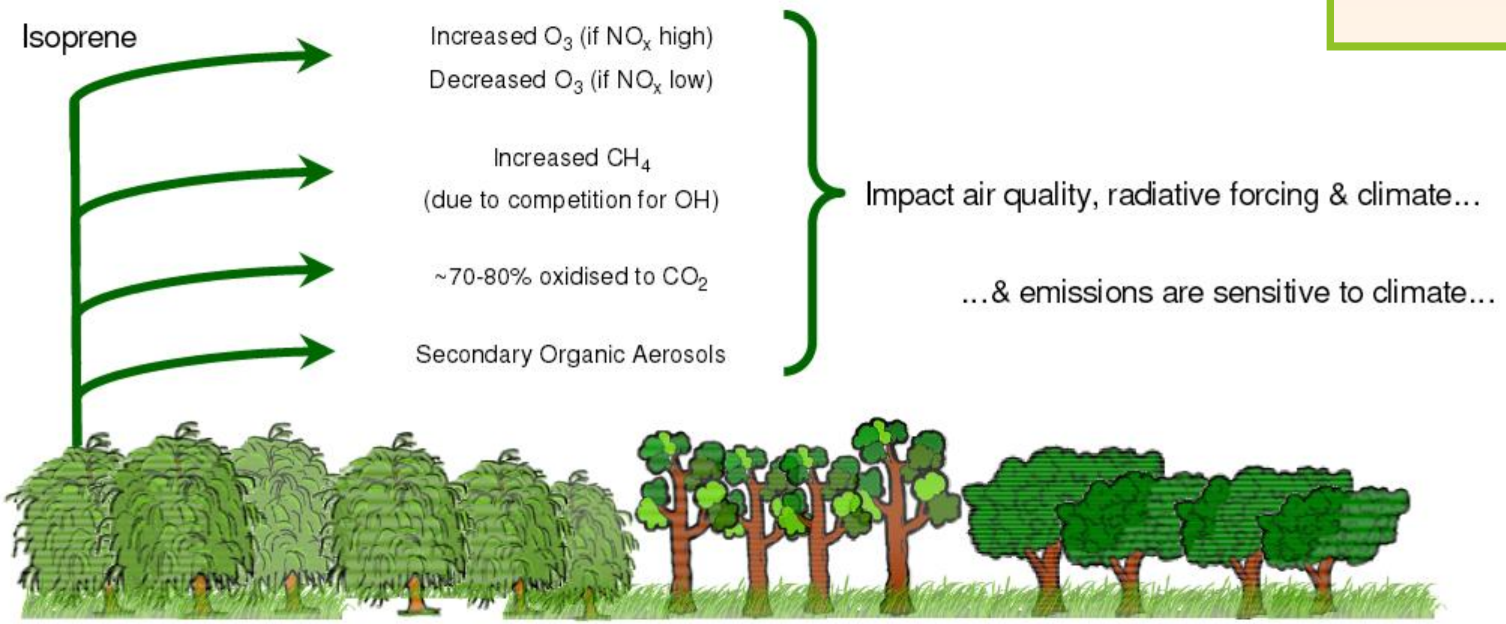
BIRA-IASB / DLR / ESA / COPERNICUS Level 2 Collection 3

©isabelle.desmedt@aeronomie.be https://doi.org/10.5270/S5P-vg1i7t0

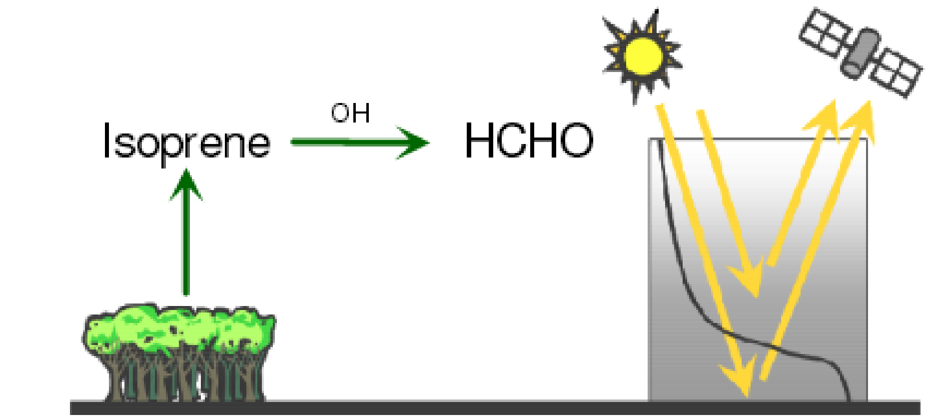
NMVOC emission from biomass burning

Total : 65 Tg / year



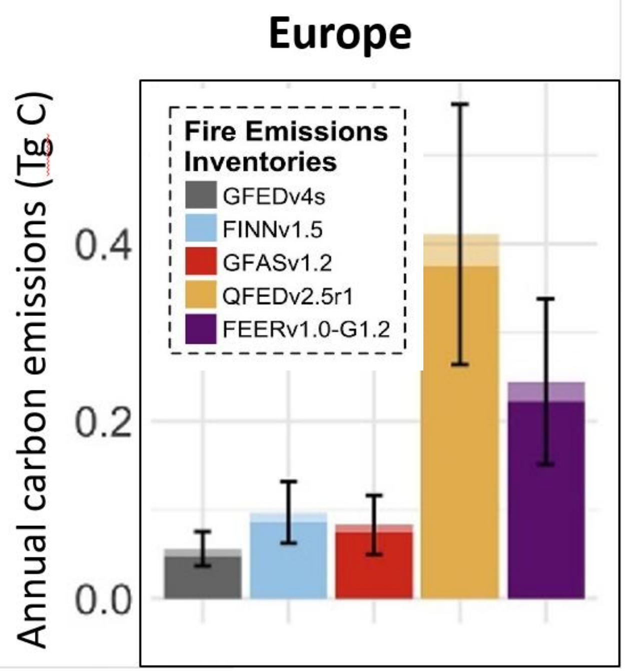


HCHO : high-yield product in isoprene oxidation



Large differences between bottom-up inventories!

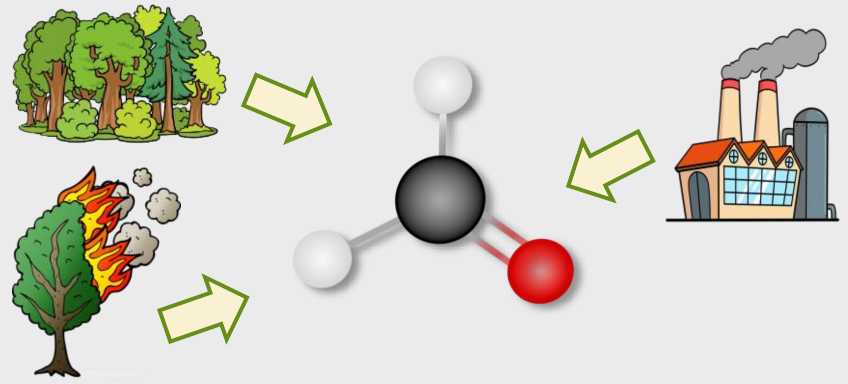
- Uncertainties due to detection of burnt area, FRP, emission factors, biome types, fuel consumption + difficulty to account for understory fires, peatland fires → hampers our understanding of fire impacts



- Factor of ~4 between global bottom-up estimates, larger differences at regional scale

BB datasets	Relies on
GFED4s	MODIS burnt area + active fires
FINN	MODIS active fire counts + active fires
GFAS	Assimilated MODIS FRP
FEER	As in GFAS, constrained by MODIS AOD
QFED	MODIS FRP + AOD
SEEDS	Top-down, based on HCHO data

- Inventories perform differently depending on species, season, location

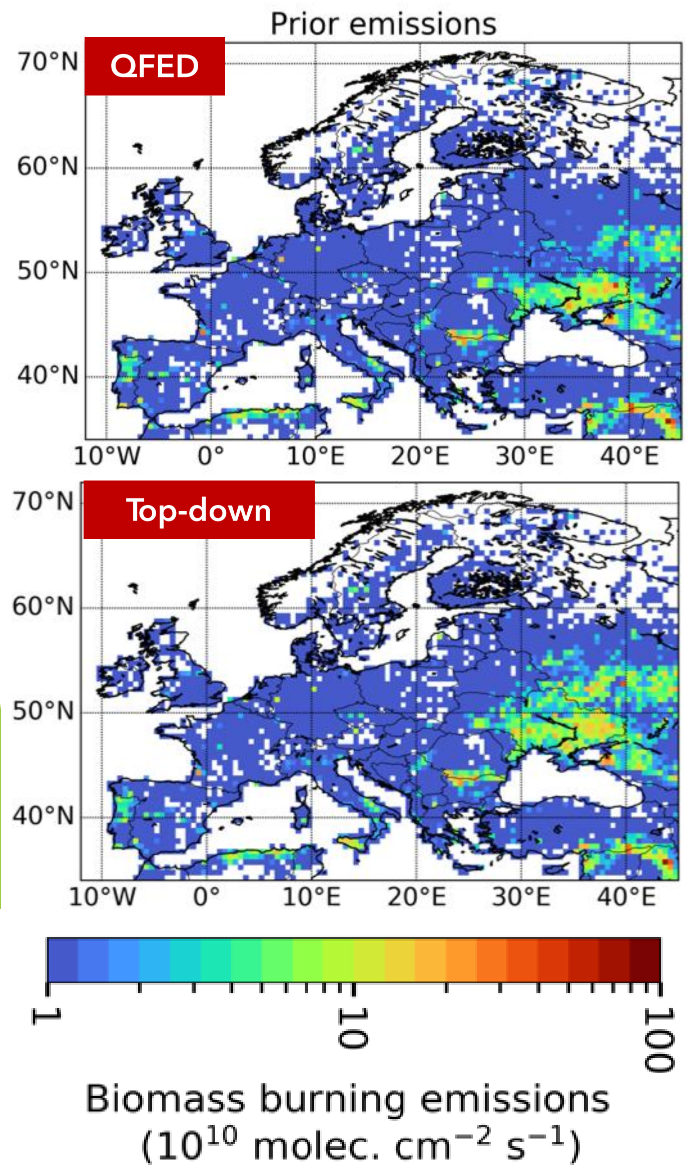


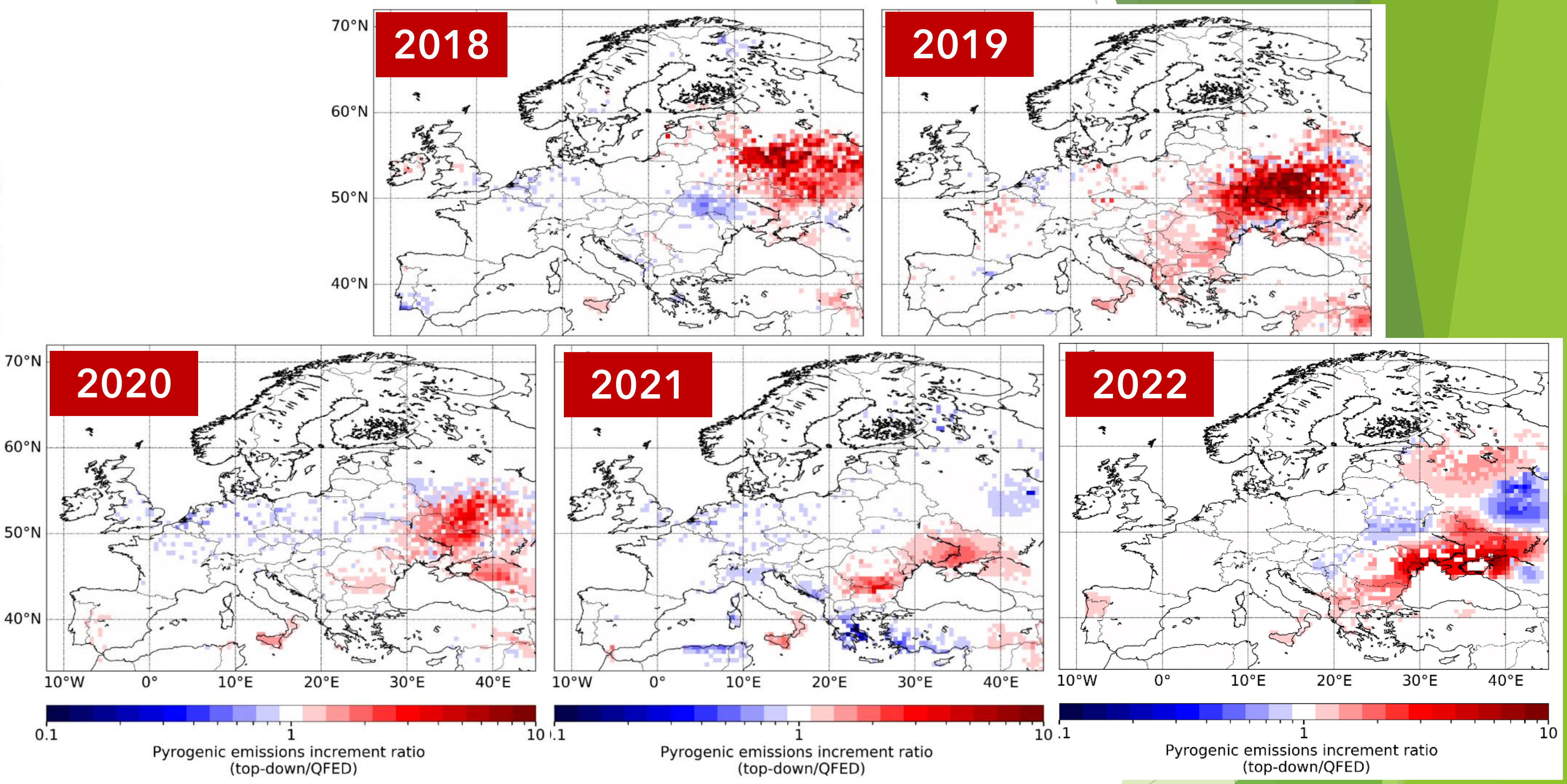
Is satellite HCHO an alternative way to constrain biomass burning emissions?

Funded by the European Union

SEEDS Sentinel EO-based Emission and Deposition Service

Emission enhancements: average May-September



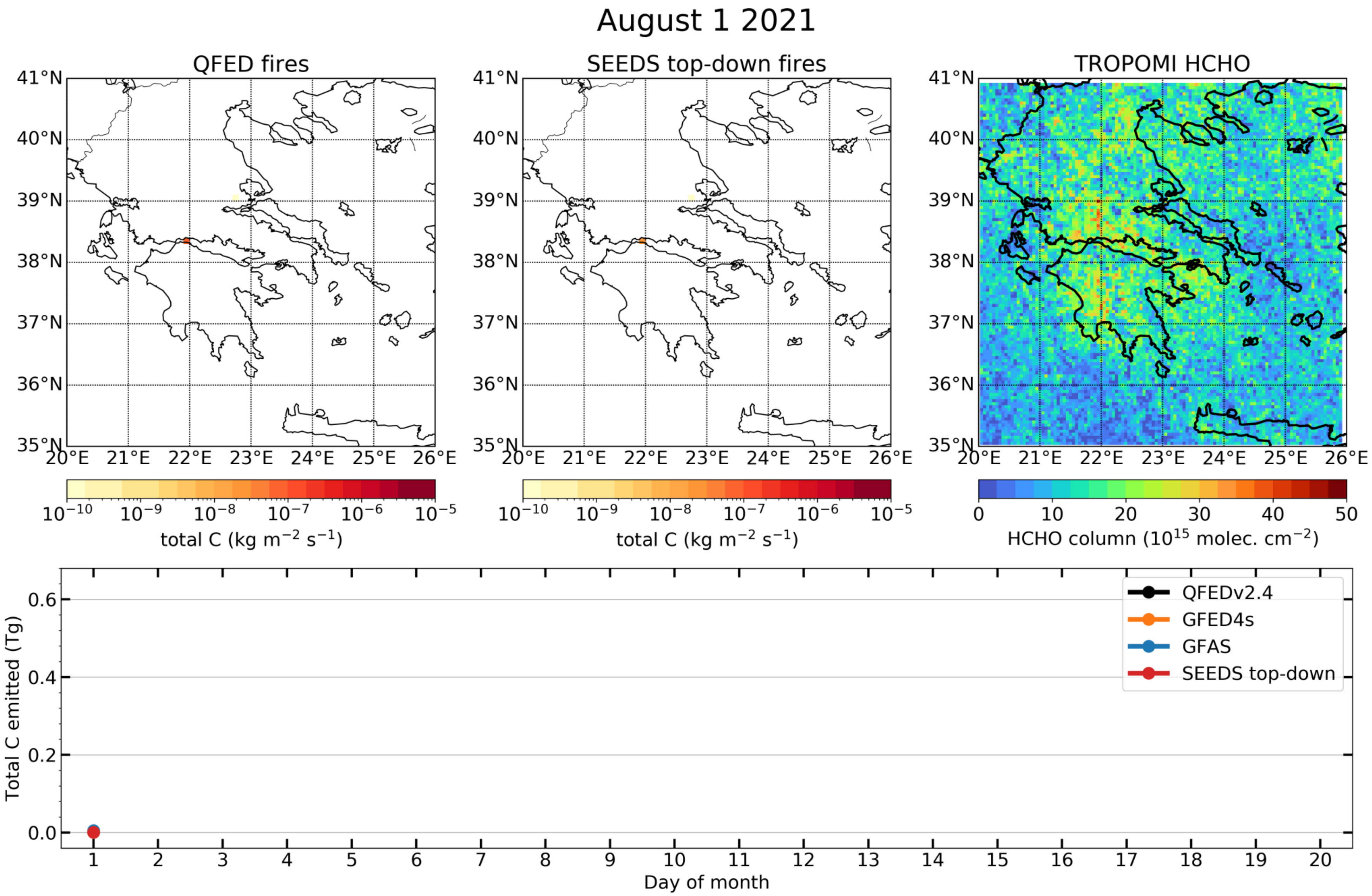


Fire emissions (Tg/yr)	2018	2019	2020	2021	2022
A priori	1.6	2.5	2.6	1.7	2.0
Top-down	2.1	3.5	2.9	1.7	1.8

Oomen et al. 2024

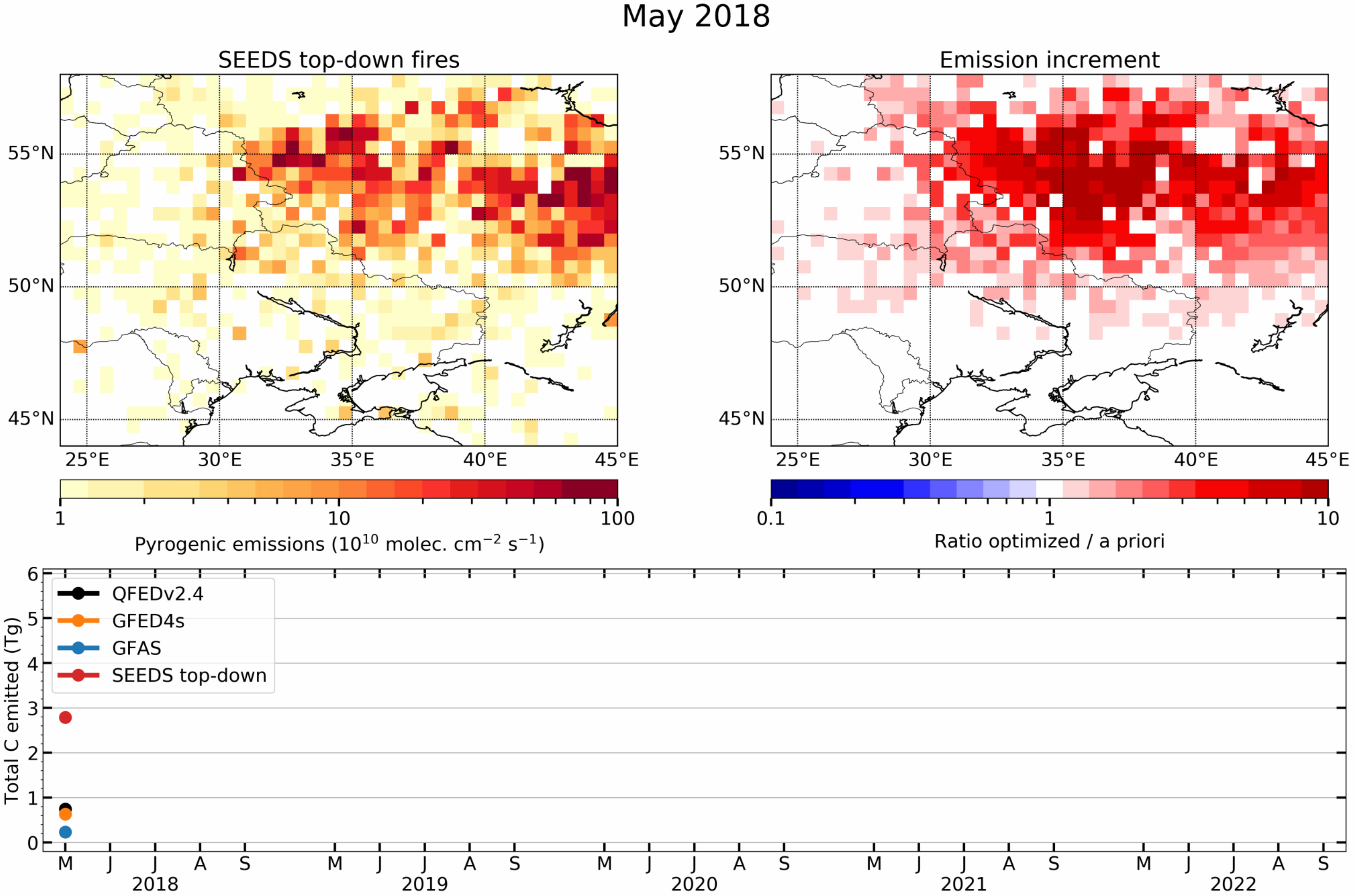
Top-down emissions during an extreme event

- Top-down emissions are lower than all inventories
- The peak on 6 Aug is well captured in all datasets, except for GFED
- The top-down peak is x2-3 lower than QFED/GFAS, could be due to the export of pollution due to strong winds



Underestimated cropland burning in Ukraine/Russia

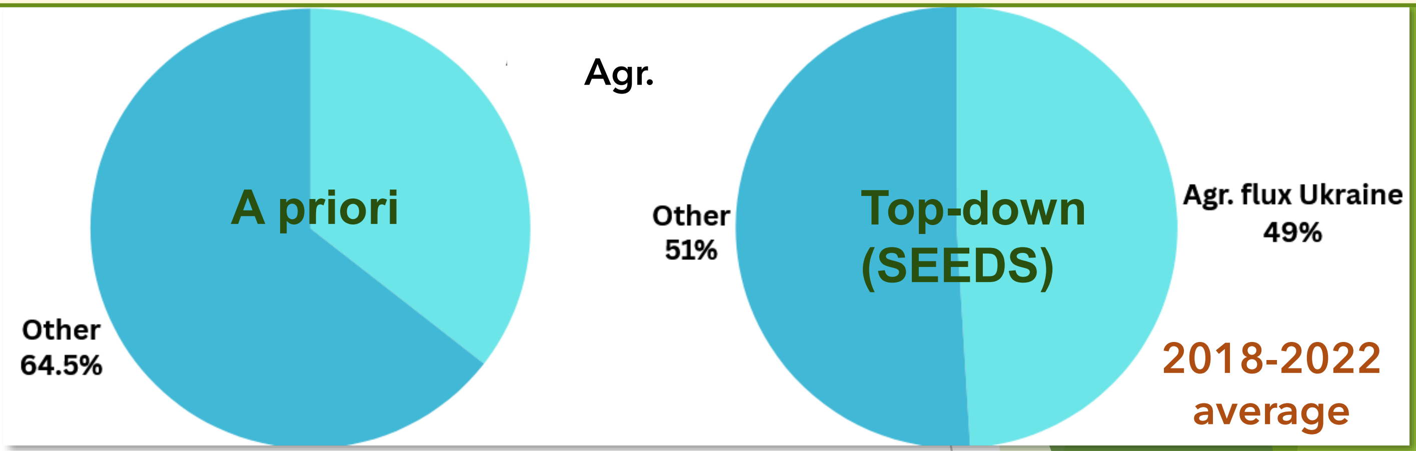
- ✓ ~Half of Ukraine is cultivated area, 70% of land is dedicated to agricultural use
- ✓ Due to the small size of cropland fires, satellite burnt area is often underestimated
- ✓ SEEDS estimates are factor of 1.5-2 higher on average than QFED, GFAS estimates are the lowest



Oomen et al. 2024

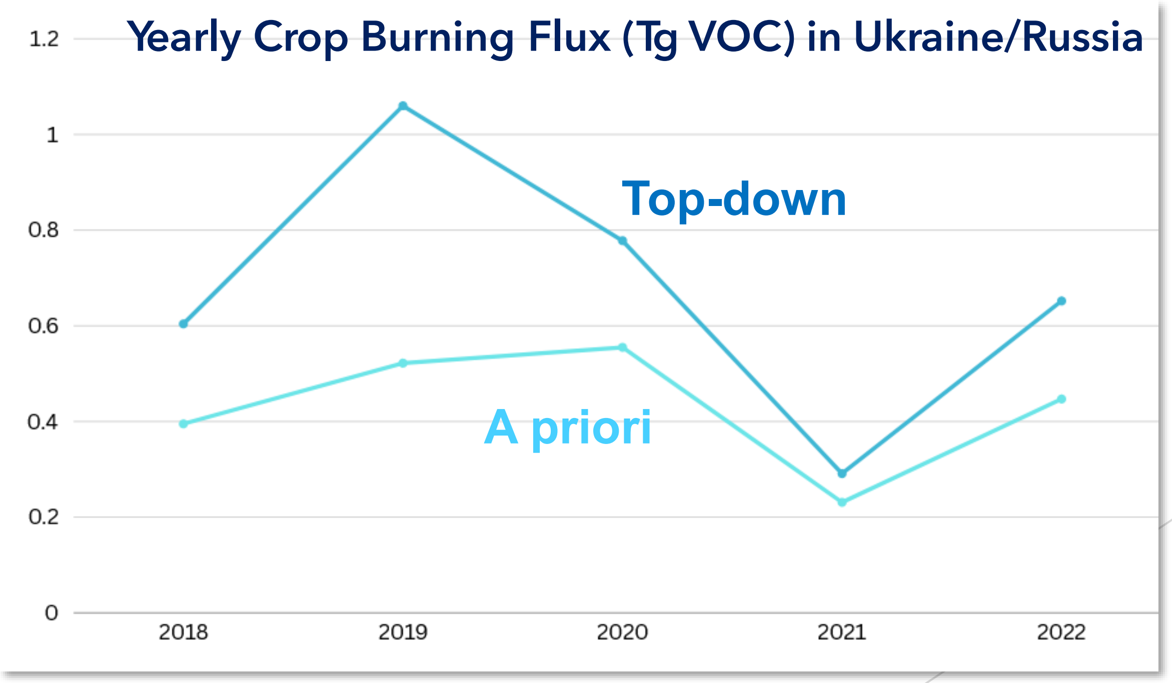
Crop residue burning in Ukraine

The share of top-down crop residue burning in Ukraine accounts for half of the total flux estimate in the European domain; Increased wrt to the a priori



Ukraine/Russia fires	TgC (5-yr mean)
A priori (QFED)	5.9
GFED	*3.7*
GFAS	*2.5*
Top-down (SEEDS)	11.2

- ✓ Small fires are underrepresented in inventories, due to difficulties to map burnt area from satellites



- ✓ Top-down products offer an alternative estimate consistent with chemical observations, independent of fire proxies

Biogenic BVOCs from the bottom-up perspective

- ✓ Natural emissions from vegetation are currently poorly constrained
- ✓ Large source of uncertainty in models

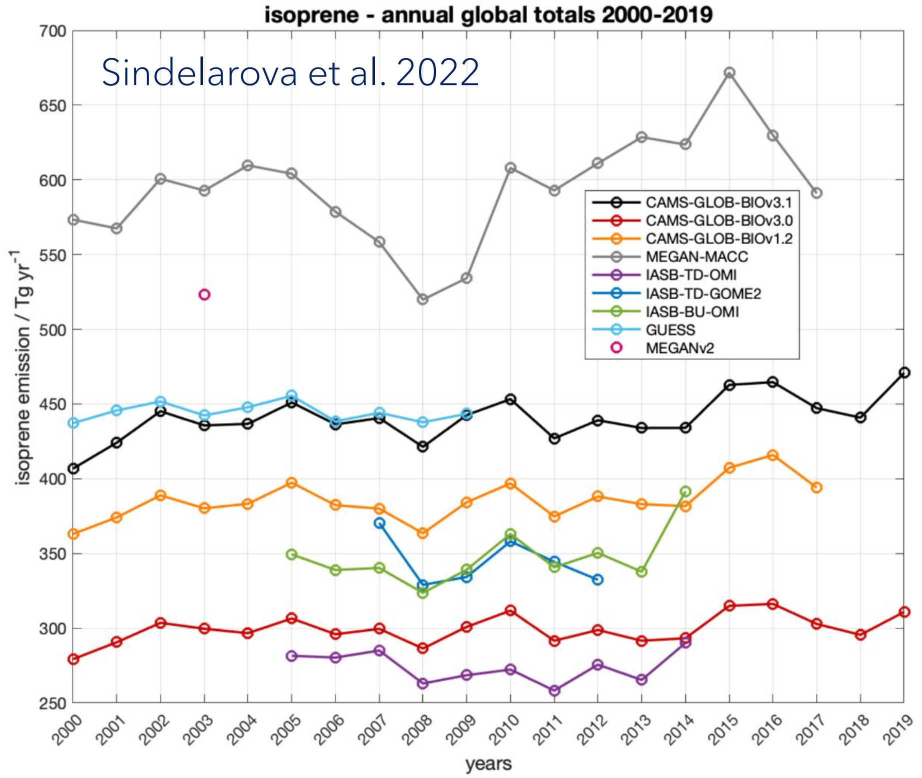
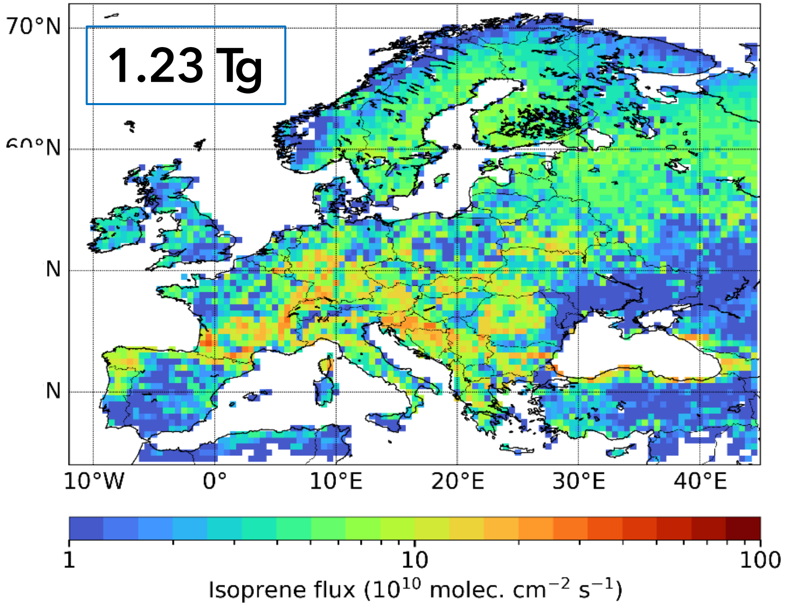
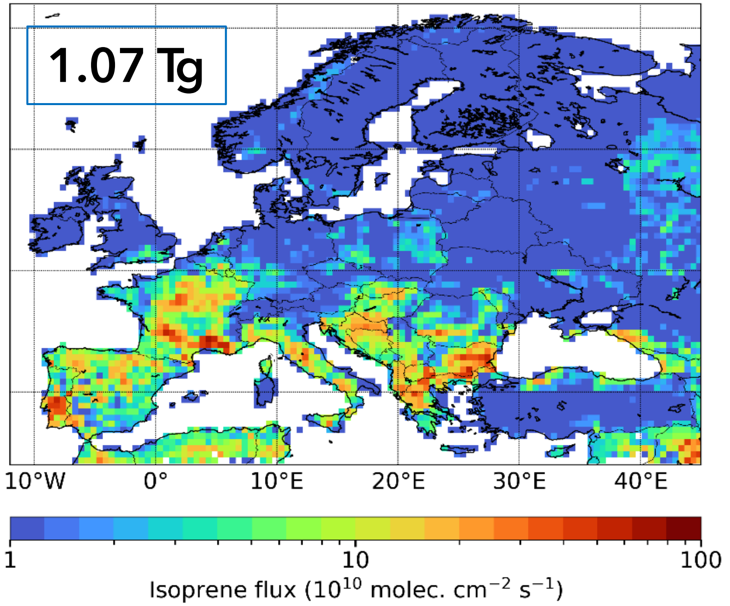


Figure 7. Comparison of isoprene global annual totals from CAMS-GLOB-BIOv3.1 (black), CAMS-GLOB-BIOv3.0 (red), CAMS-GLOB-BIOv1.2 (orange) and other available inventories within the 2000–2019 period.

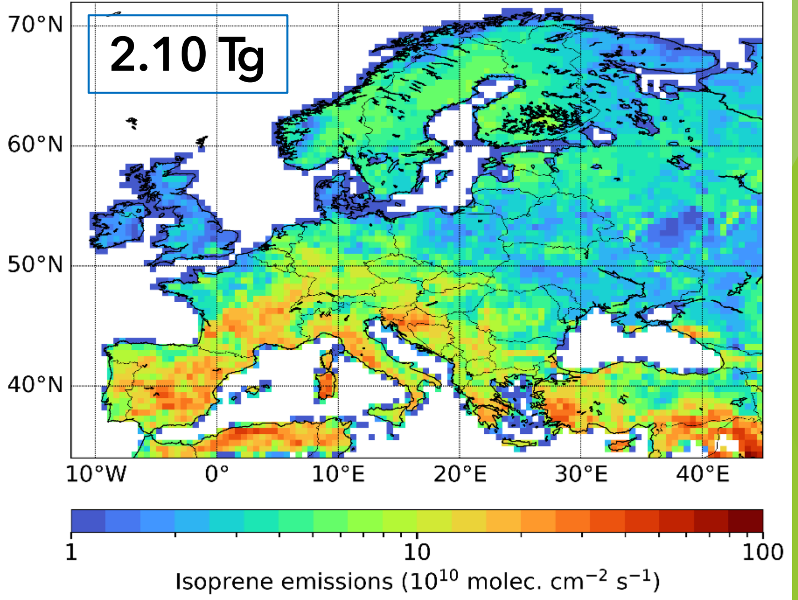
MEGAN-SURFEX



CAMS-GLOB-BIOv3.1 (Sindelarova et al. 2022)

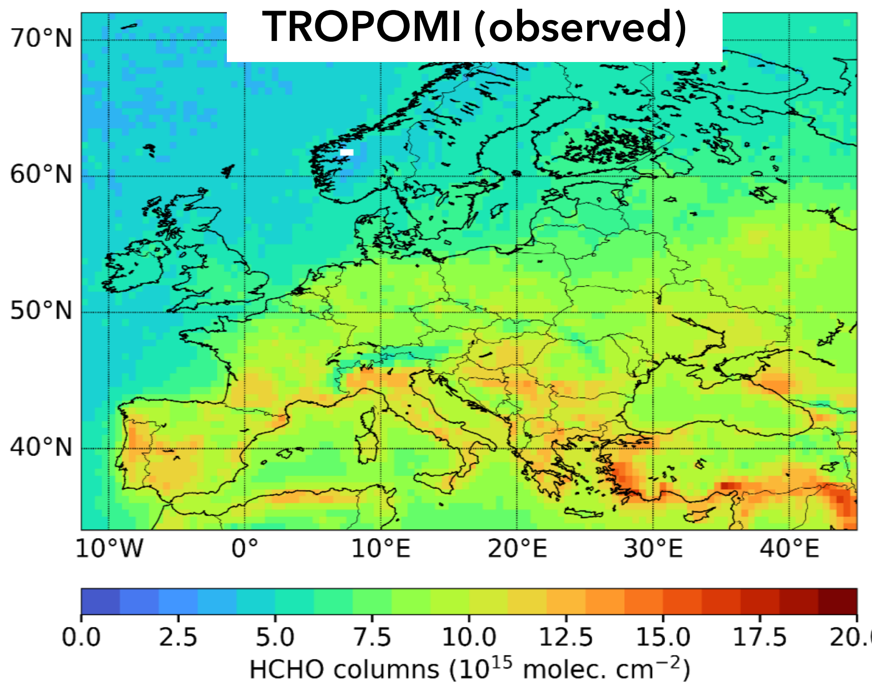


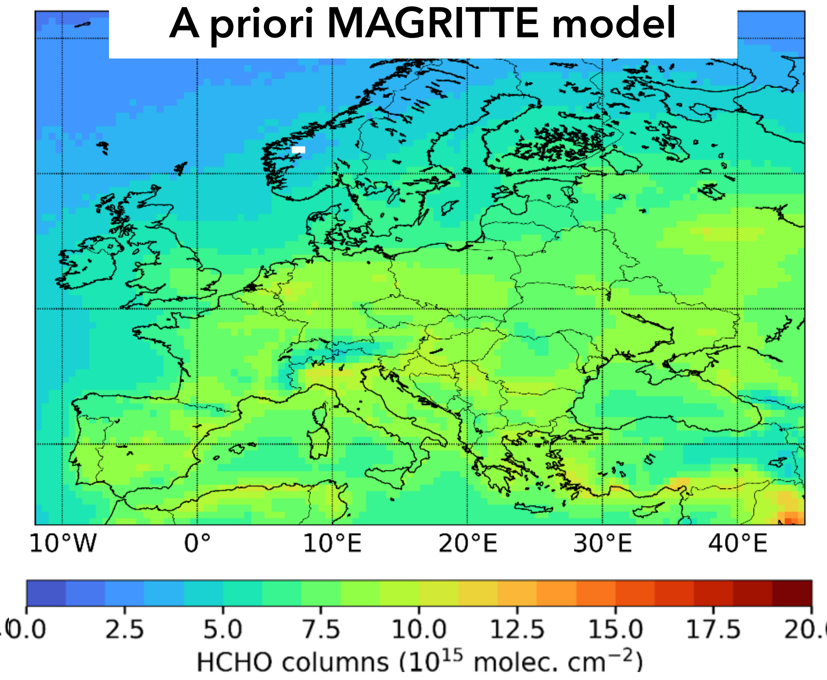
MEGAN-MOHYCAN (Stavrakou et al. 2018)

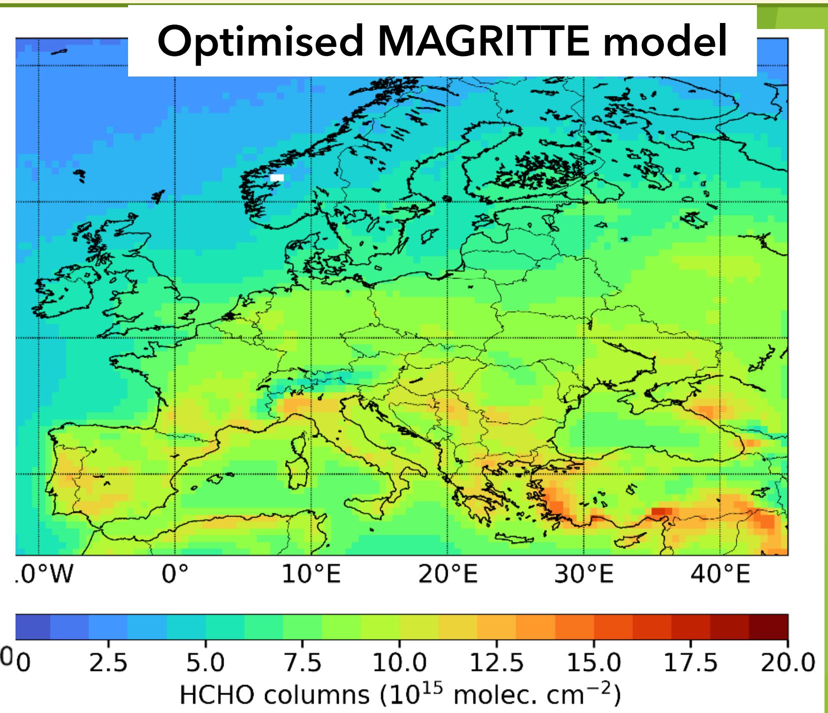


- ✓ In Europe, large differences found even among inventories built on the same emission model (MEGAN)

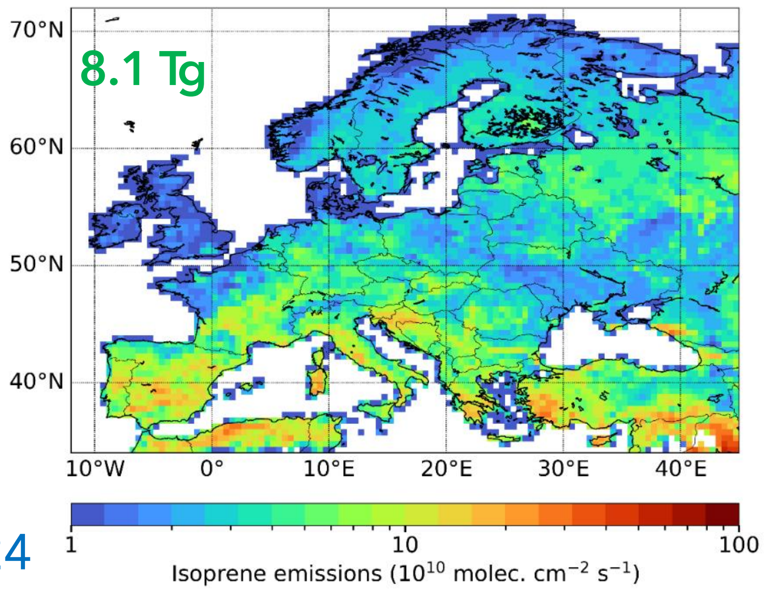
…and from the satellite perspective



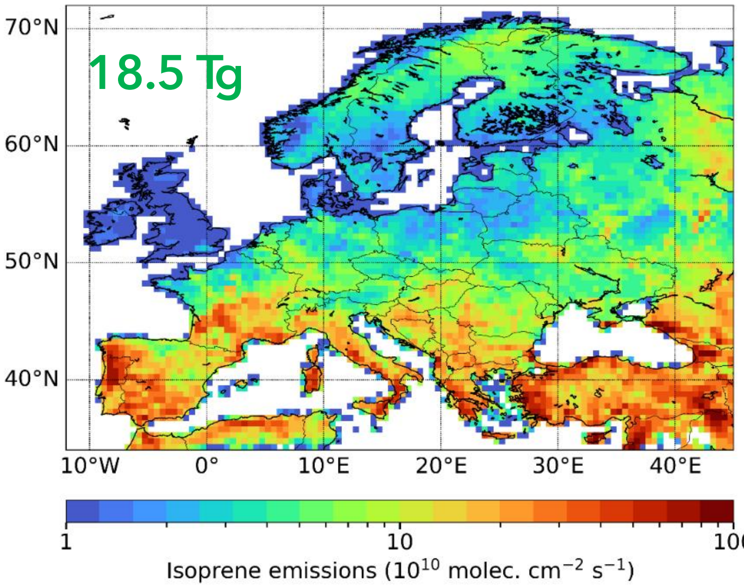




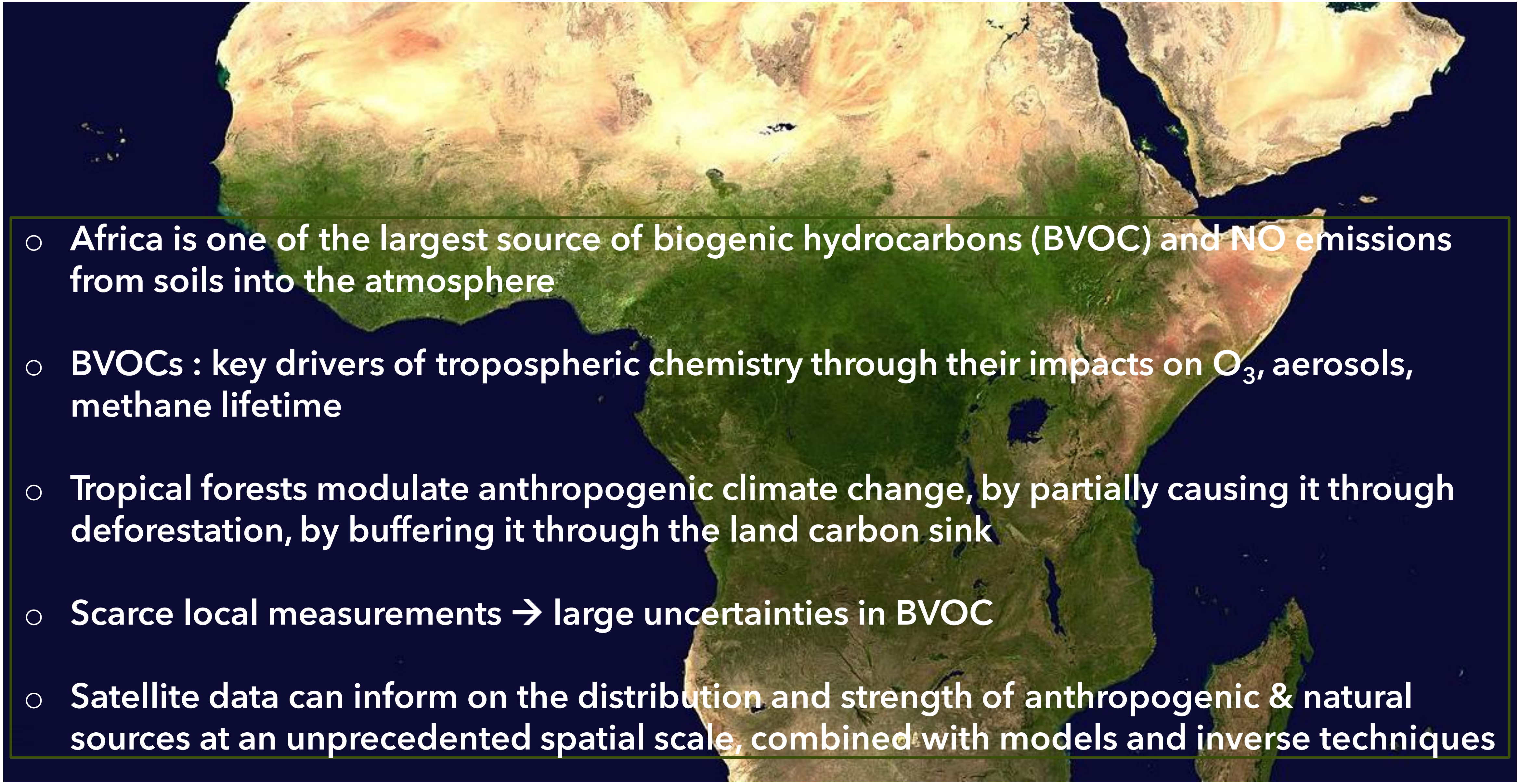
A priori emissions



Optimized (top-down) emissions

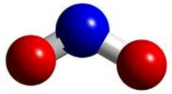


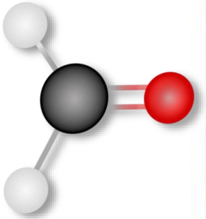
Oomen et al. 2024

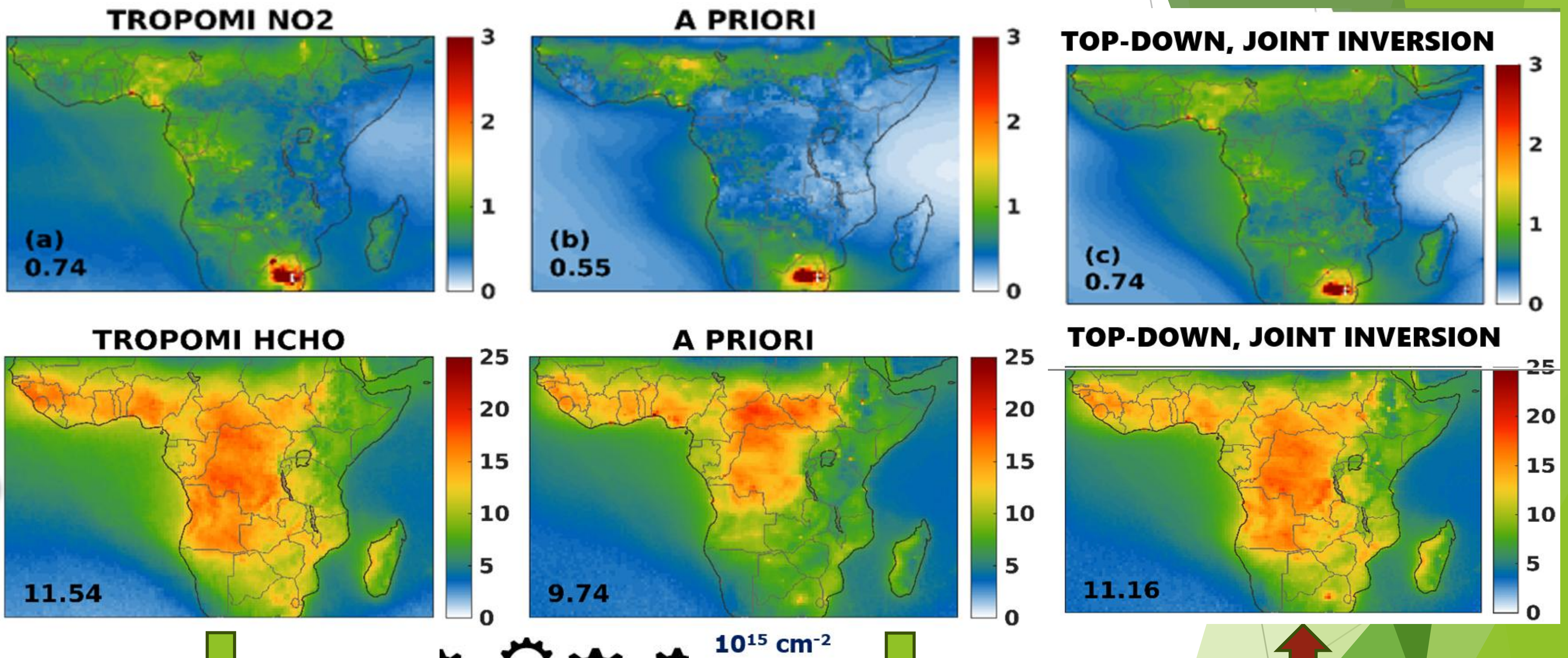


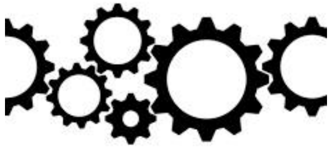
- Africa is one of the largest source of biogenic hydrocarbons (BVOC) and NO emissions from soils into the atmosphere
- BVOCs : key drivers of tropospheric chemistry through their impacts on O_3, aerosols, methane lifetime
- Tropical forests modulate anthropogenic climate change, by partially causing it through deforestation, by buffering it through the land carbon sink
- Scarce local measurements → large uncertainties in BVOC
- Satellite data can inform on the distribution and strength of anthropogenic & natural sources at an unprecedented spatial scale, combined with models and inverse techniques

Satellite NO_2 and HCHO data can inform about emission sources!







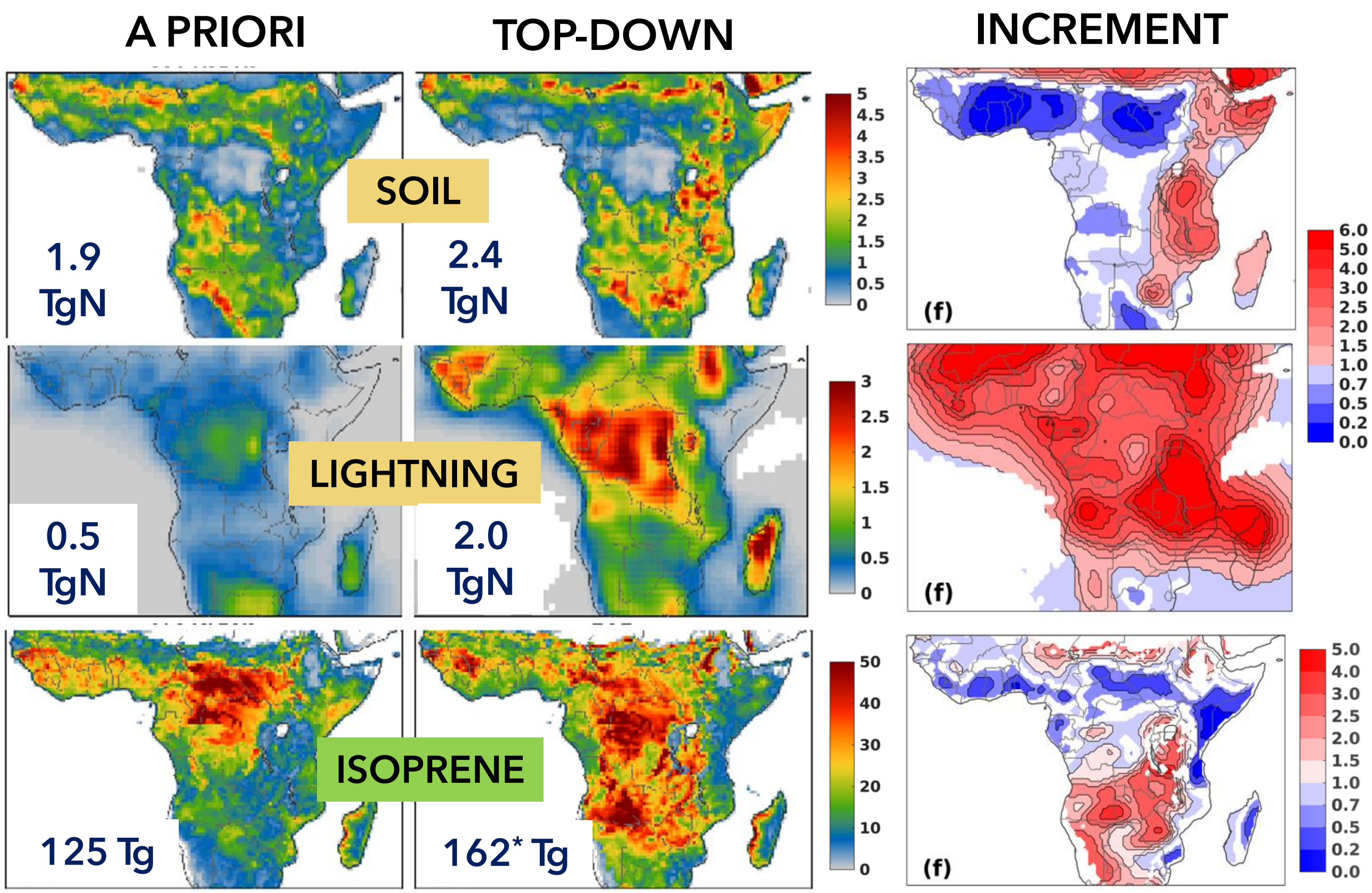


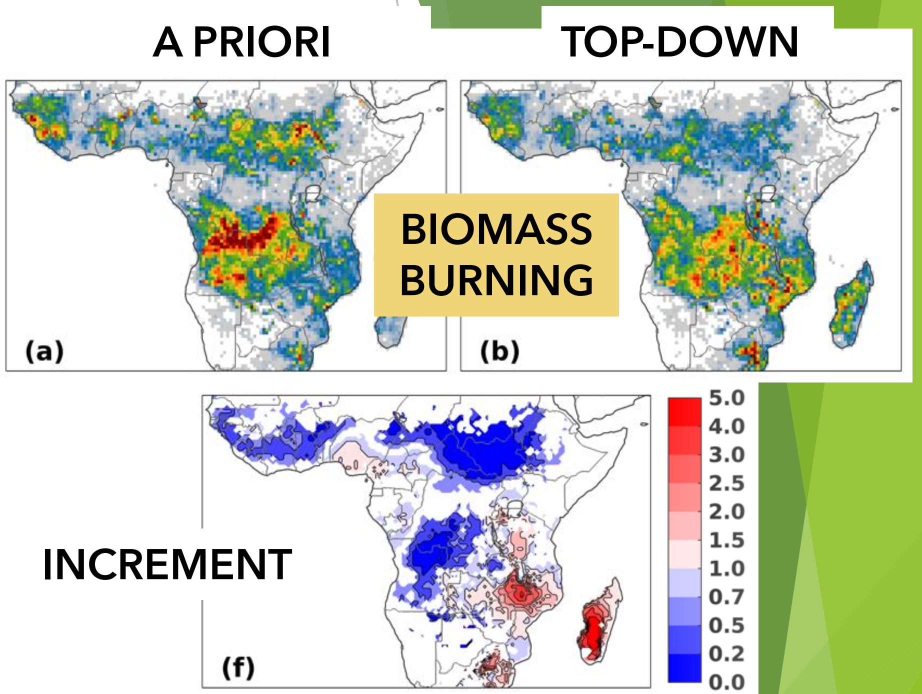
Inverse methods

Very close agreement with observations achieved by significant emission changes

Opacka et al. 2025

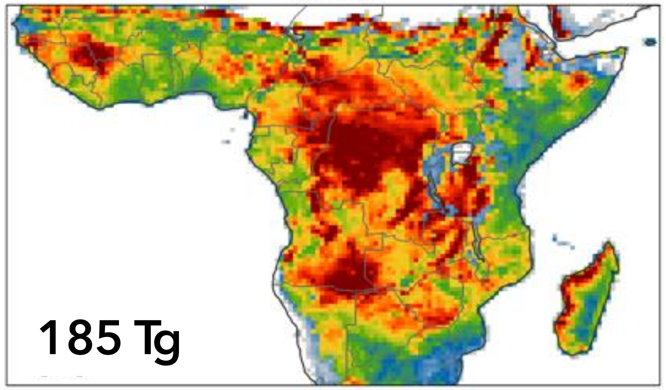
Significant emission changes!



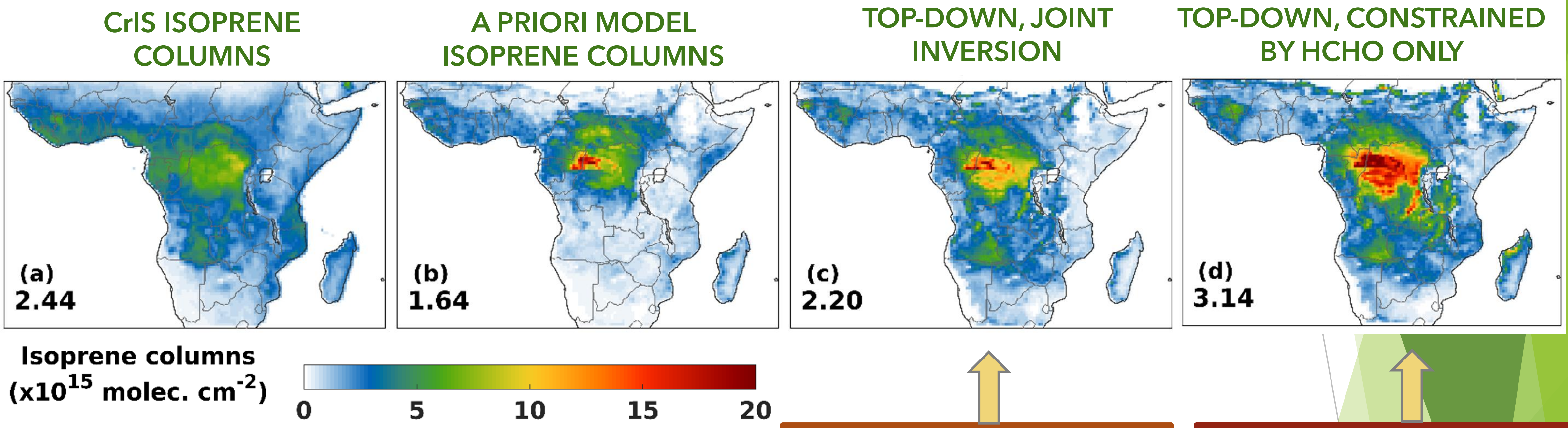


185 Tg when NO_2 is not used to constrain the inversion

TOP-DOWN ISOPRENE, CONSTRAINED BY HCHO ONLY



Evaluation against CrIS isoprene columns



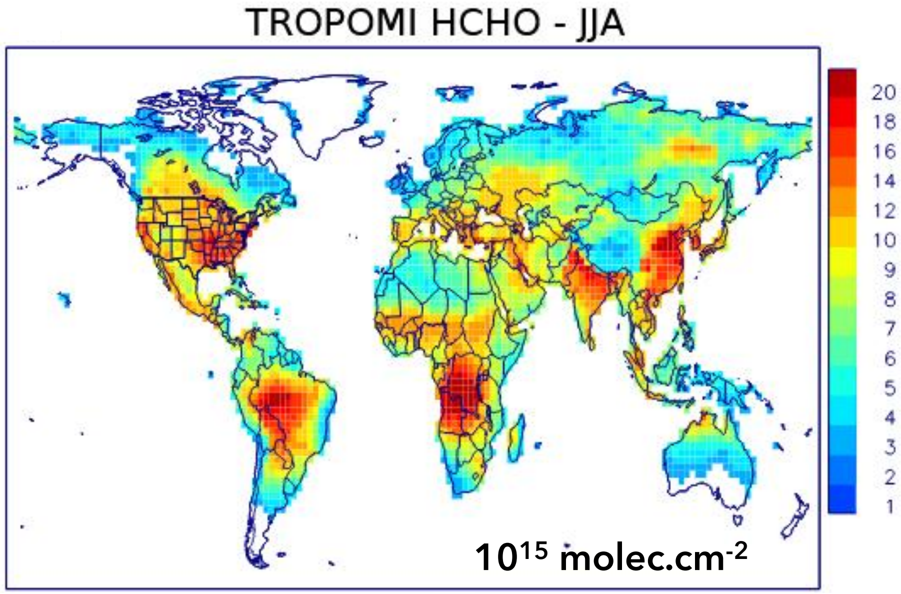
A priori spatial patterns are more contrasted than in the observations

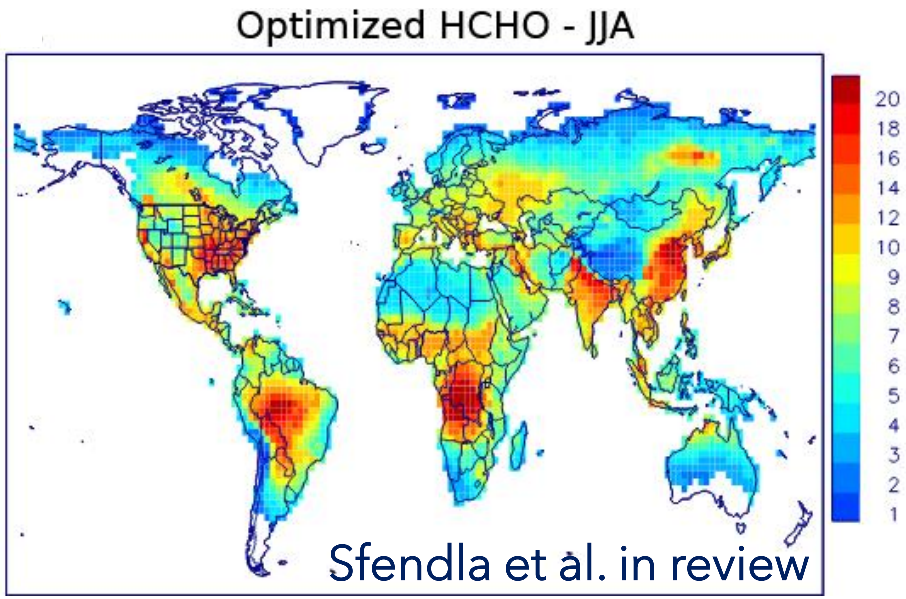
Improved agreement with CrIS in the joint inversion (-33% → -10%), improved spatial distribution

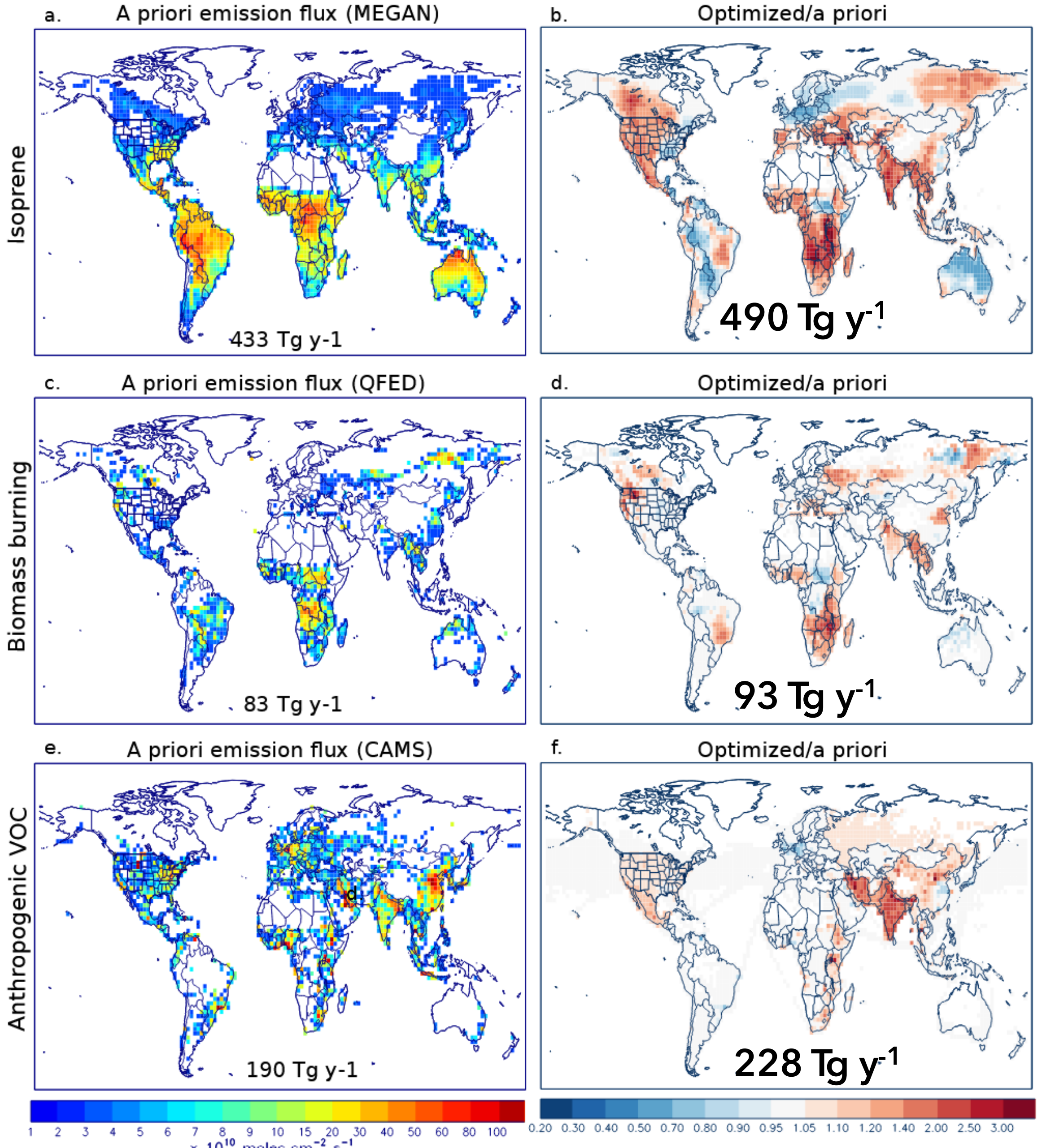
Compared to the joint inversion, increased columns by 40%. This is due to lower NOx fluxes, lower OH levels and longer isoprene lifetimes in HCHO-only inversion

Opacka et al. 2025

Inferred VOC emission fluxes on global scale using TROPOMI

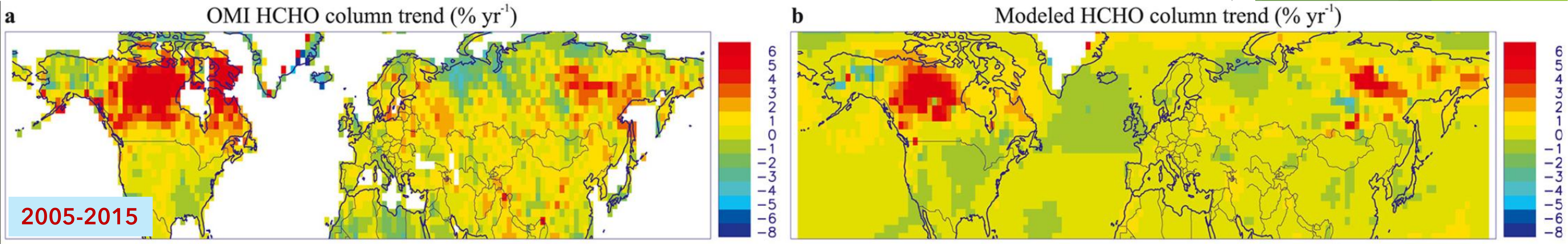




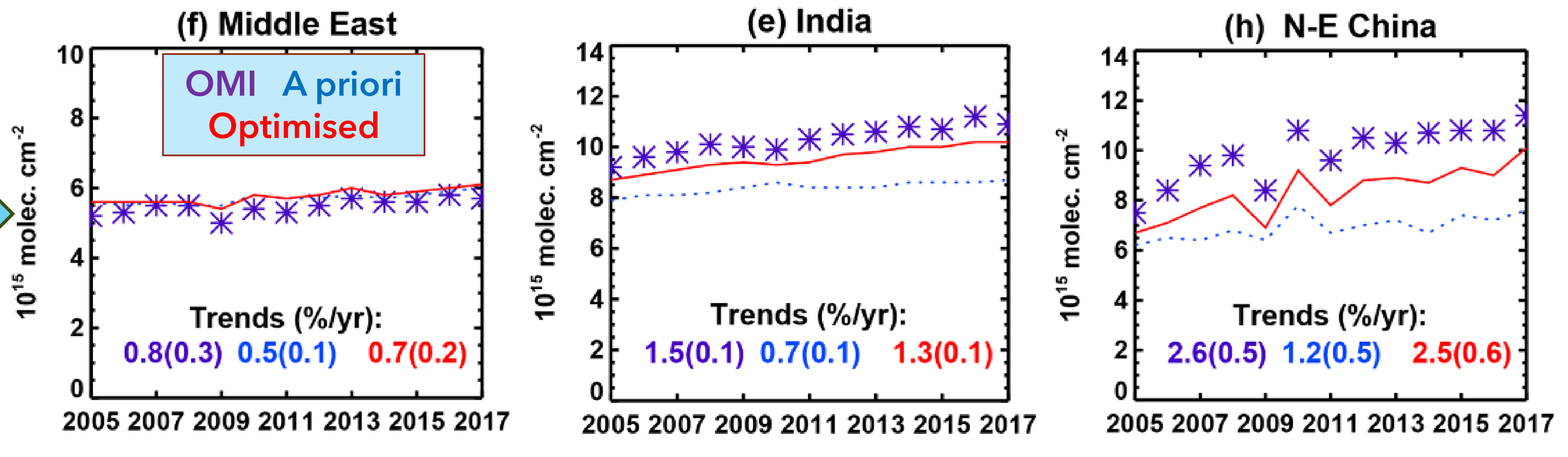


Satellite-based trends

Clear evidence of a significant impact of climate variability over BVOC-dominated regions



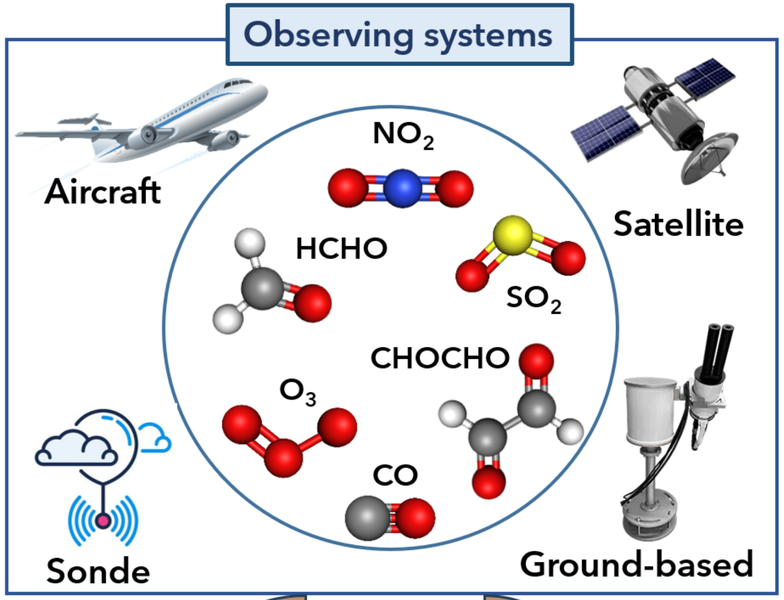
Inversion based on OMI HCHO over 2005-2017

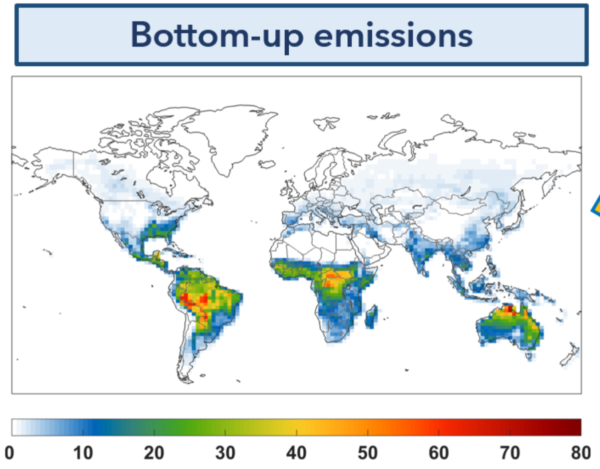


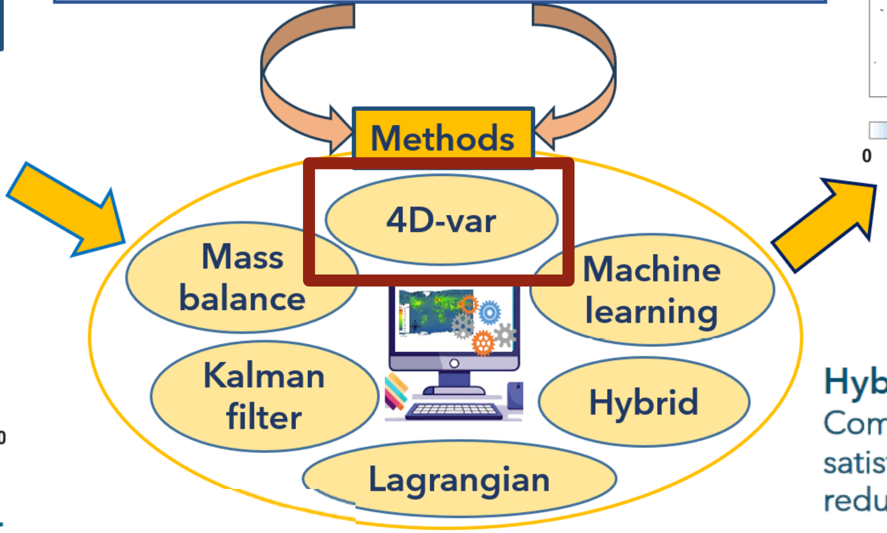
Müller et al. 2024
Stavrakou et al. 2018

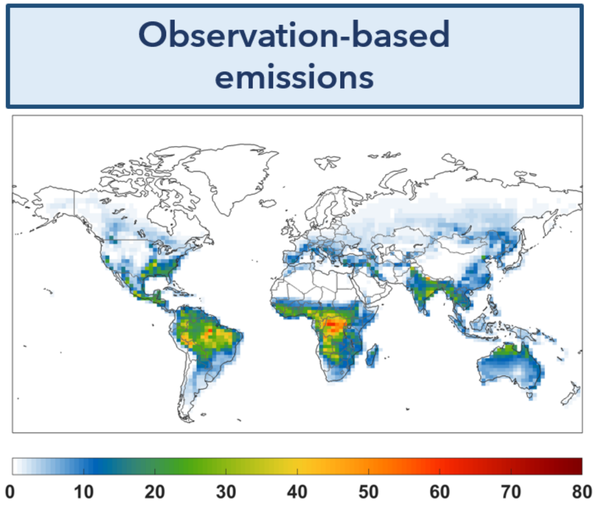
European Space Agency
PIRAMID

Other top-down approaches









Machine learning
Data-driven inference of emissions and processes from large datasets

Hybrid
Combines techniques to satisfy accuracy and reduce compute cost

Mass-balance
Top-down approach inferring emissions from observed concentration changes assuming known transport and chemical changes

Kalman filter
Sequential update scheme combining simulations with new observations; suited for real-time applications

Lagrangian
Particle or trajectory modeling tracing plumes backward (sources) or forward (dispersion).

Take-home messages

- Use of satellite data to learn about emission sources and their evolution, and to interpret the observed long-term satellite trends
- TROPOMI HCHO suggest increased fire fluxes from crop residue burning wrt bottom-up inventories and changes in spatial distribution of biogenic sources → geostationary observations offer promise to better determine these emissions
- Co-occurrence of sources (fires and enhanced vegetation emissions) during summertime makes it challenging to separate the sources
- Need for improved representation of pyrogenic and biogenic VOCs in models
- Inverse methods allow for a large array of applications & improved assessments of air quality (compounded by independent ground/satellite observations)