Universität Bremen

LAMOS

Model evaluation with satellite retrievals

Behind the scenes of validating the new MOGUNTIA chemistry

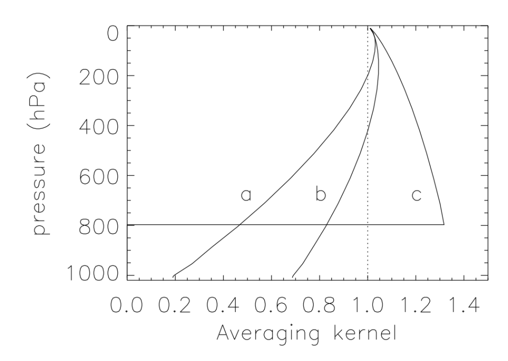
Andreas Hilboll, N. Daskalakis, M. Vrekoussis, S. Myriokefalitakis, *et al.*

29th International TM Meeting, 21–22 November 2019

What’s so special about satellites?

Vertical sensitivity / averaging kernels

- *sensitivity* of the satellite measurement depends on light path
    - altitude
    - wavelength (i.e., trace gas)
    - cloudiness
    - aerosols
    - surface albedo



Column AKs for (a) clear-sky, albedo 0.02; (b) clear-sky, albedo 0.15; (c) cloudy, cloud top @800 hPa [Eskes and Boersma, 2003]

What *is* an averaging kernel?

- for *profile* retrievals from satellite, the AK is a *matrix* $\mathbf{A}_{ij}$, indicating the sensitivity of the retrieved concentration in layer j to trace gas present in atmospheric layer i
- for *column* retrievals from satellite, the AK is a *vector* $\mathbf{A}_i$, indicating the sensitivity of the retrieved column to trace gas present in atmospheric layer i

How do I use an averaging kernel?

- using the modelled concentrations, calculate the column which the sensor *would have retrieved* $\hat{y}_m$ from the modelled profile x_m

$$\hat{y}_m = \mathbf{A} \cdot x_m$$

- in case of *tropospheric* columns (i.e., NO_2), need to apply conversion factor:

$$\hat{y}_{m,\text{trop}} = \mathbf{A} \cdot \frac{AMF}{AMF_{\text{trop}}} \cdot x_{m,\text{trop}}$$

- this $\hat{y}_m$ can then be compared to the satellite retrievals

Data aggregation workflow

From single satellite measurements to gridded data

1. read satellite data (lv 2: retrieval output of single measurements, i.e., column densities and averaging kernels)
2. (strictly) filter satellite data (according to product specification)
3. (linearly) interpolate hourly model concentration profiles to time and horizontal location of each (remaining) measurement
4. for each satellite measurement, apply the averaging kernel to the model concentration profile to yield $\hat{y}_m$
5. average all $\hat{y}_m$ from one day within a model grid cell

From single satellite measurements to gridded data

1. read satellite data (lv 2: retrieval output of single measurements, i.e., column densities and averaging kernels)
2. (strictly) filter satellite data (according to product specification)
3. (linearly) interpolate hourly model concentration profiles to time and horizontal location of each (remaining) measurement
4. for each satellite measurement, apply the averaging kernel to the model concentration profile to yield $\hat{y}_m$
5. average all $\hat{y}_m$ from one day within a model grid cell

What does this leave us with?

For each model dataset (different chemistry schemes / solvers)

- vertical column densities *as the satellite would have seen them*
- at model resolution (here: 1×1°)
- averaged daily

Model validation

Maps

- monthly / annual mean maps (global or regional)
- one map per dataset
- additional map(s) for relative and/or absolute differences

Why (not) to use maps

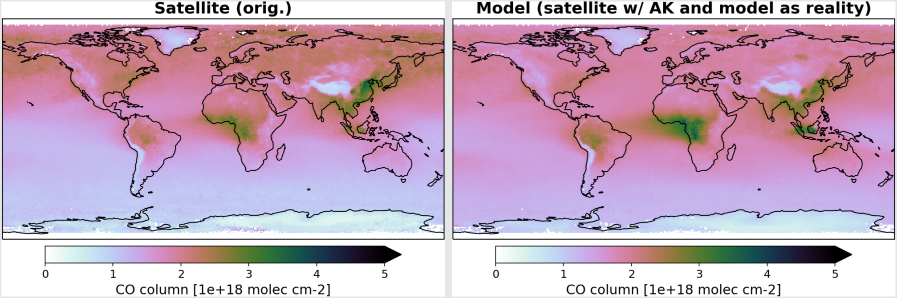
- make sense only for long temporal aggregates (noisy measurement data!)
- are hard to interpret quantitatively
- give indication for areas to look into at more detail

Maps

- monthly / annual mean maps (global or regional)
- one map per dataset
- additional map(s) for relative and/or absolute differences

Example

2006 mean CO total VCDs from MOPITT (l) and TM5mp-MOGUNTIA (r)

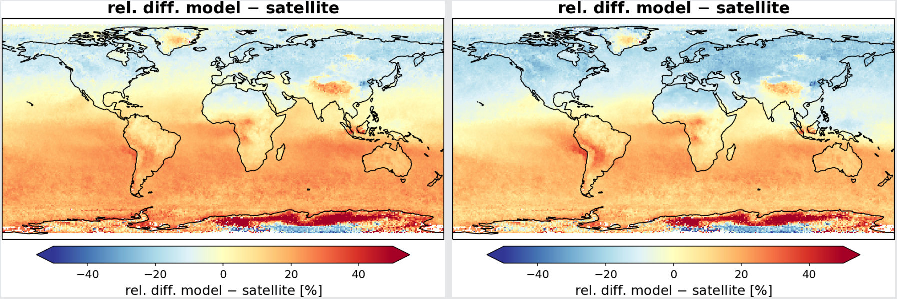


Maps

- monthly / annual mean maps (global or regional)
- one map per dataset
- additional map(s) for relative and/or absolute differences

Example

Relative difference TM5—MOPITT for MOGUNTIA (l) and CB05/EBI (r)



Time-series

- monthly mean time-series over individual regions
- error bars to show variability within one month
- several datasets in one plot

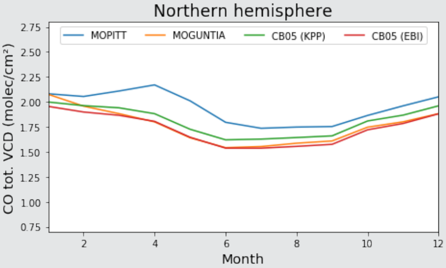
Why (not) to use time-series

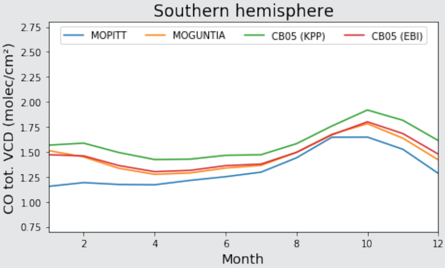
- spatial (regional) averages
- longer temporal averages
- can give ideas about seasonal changes in the differences

Time-series

- monthly mean time-series over individual regions
- error bars to show variability within one month
- several datasets in one plot

Example

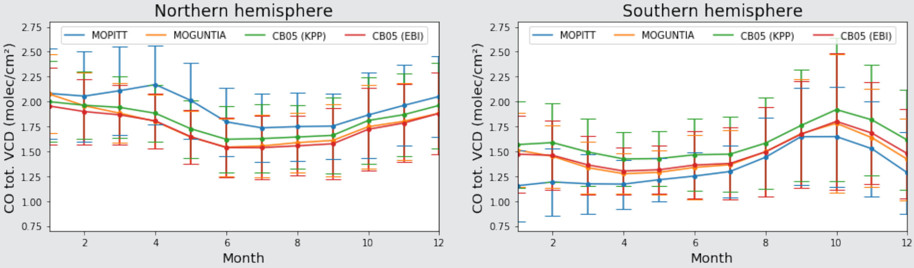




Time-series

- monthly mean time-series over individual regions
- error bars to show variability within one month
- several datasets in one plot

Example



Scatter plots / heatmaps

- scatter plots/heatmaps of daily/monthly/annual data
- comparison against reference data (one plot per model dataset)
  - draw 1:1 and linear regression lines
  - show distribution / histogram for individual datasets on outer axes

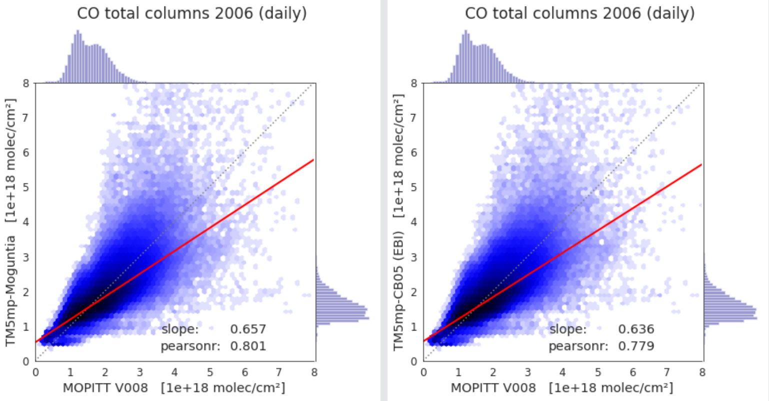
Why (not) to use scatter plots

- make the model look very bad when used on daily data, especially for HCHO and NO_2
- make the model look very good when used on annual averages
- nicely shows correlation and slope

Scatter plots / heatmaps

- scatter plots/heatmaps of daily/monthly/annual data
- comparison against reference data (one plot per model dataset)
    - draw 1:1 and linear regression lines
    - show distribution / histogram for individual datasets on outer axes

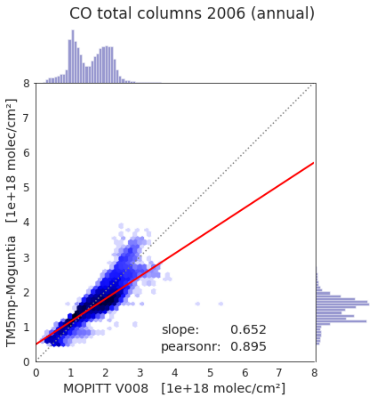
Example

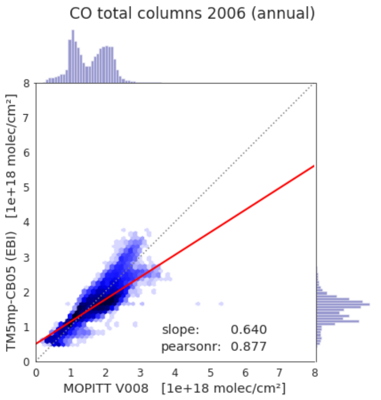


Scatter plots / heatmaps

- scatter plots/heatmaps of daily/monthly/annual data
- comparison against reference data (one plot per model dataset)
    - draw 1:1 and linear regression lines
    - show distribution / histogram for individual datasets on outer axes

Example



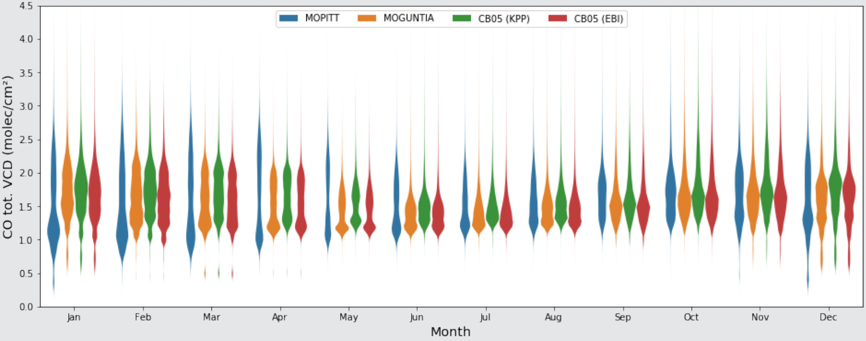


Other distribution plots (violin plots, ...)

- violin plots basically show many histograms simultaneously

Example

Monthly distribution of MOPITT and TM5 (violin plot)



Aggregated statistics

- correlation coefficient
- slope/intercept of regression line
- (normalized) mean bias, root mean square error, ...
- unbiased *symmetric* metrics [Yu et al., 2006]:
  - normalized mean bias factor (NMBF)
  - normalized mean absolute error factor (NMAEF)

How to visualize aggregated statistics?

- Tables (not really visual ...)
- Bar plots of, e.g., correlation
- Taylor plots
- ???

Aggregated statistics

- correlation coefficient
- slope/intercept of regression line
- (normalized) mean bias, root mean square error, ...
- unbiased *symmetric* metrics [Yu et al., 2006]:
  - normalized mean bias factor (NMBF)
  - normalized mean absolute error factor (NMAEF)

Example

Correlation and slope of daily mean gridded model CO VCDs against MOPITT V008 CO total VCDs:

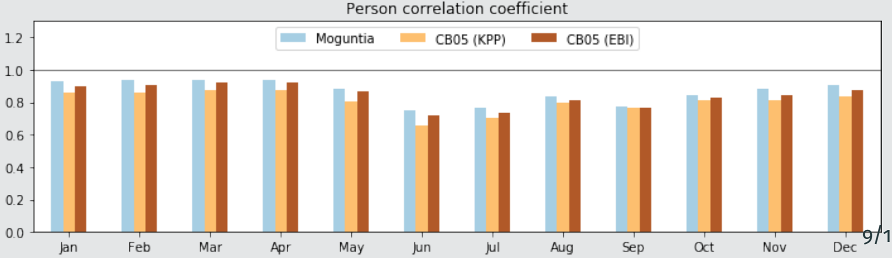
Dataset	Pearson's *r*	Slope
Moguntia	0.810	0.657
CB05 (KPP)	0.740	0.631
CB05 (EBI)	0.779	0.636

Aggregated statistics

- correlation coefficient
- slope/intercept of regression line
- (normalized) mean bias, root mean square error, ...
- unbiased *symmetric* metrics [Yu et al., 2006]:
  - normalized mean bias factor (NMBF)
  - normalized mean absolute error factor (NMAEF)

Example

Correlation of daily mean gridded model CO VCDs vs. MOPITT V008:



Beyond column densities: what else can be done?

- surface concentration measurements (from satellite or in-situ)
- isobar concentration measurement (from satellite profiles)
- profile measurements from sonde flights
- concentration measurements from aircraft (non-regular locations and timings)

Conclusions

Summary

- modelled concentration profiles have to be adapted for the measuremts' vertical sensitivity before comparison
- there are many different ways to evaluate these global datasets
    - maps look nice and give a general overview
    - time-series show seasonality
    - scatter plots / heatmaps of daily values are honest but look bad
    - scatter plots / heatmaps of annual values look good but are not really honest
        - scatter plot somewhat implies individual data points
    - aggregate stastical metrics can be a nice tool to summarize
- It is hard to not get lost in all the options … what to choose best?

References

K. F. Boersma, G. C. M. Vinken, and H. J. Eskes. Representativeness errors in comparing chemistry transport and chemistry climate models with satellite uv-vis tropospheric column retrievals. *Geoscientific Model Development*, 9(2):875–898, 2016. DOI:10.5194/gmd-9-875-2016.

H. J. Eskes and K. F. Boersma. Averaging kernels for DOAS total-column satellite retrievals. *Atmospheric Chemistry and Physics*, 3(5):1285–1291, 2003. DOI:10.5194/acp-3-1285-2003.

A. Richter, A. Hilboll, A.-M. Blechschmidt, and L. K. Behrens. On the use of UV/vis satellite tropospheric data products. PANDA Deliverable D5.6, University of Bremen, December 2015. URL https://doi.org/10.6084/m9.figshare.3593361.

K. E. Taylor. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7):7183–7192, 2001. DOI:10.1029/2000jd900719.

S. Yu, B. Eder, R. Dennis, S.-H. Chu, and S. E. Schwartz. New unbiased symmetric metrics for evaluation of air quality models. *Atmospheric Science Letters*, 7(1): 26–34, 2006. DOI:10.1002/asl.125.