Photography encouraged

ROYAL METEOROLOGICAL INSTITUTE OF BELGIUM
1

ROYAL OBSERVATORY OF BELGIUM
2

Interpreting the time variability of integrated water vapour retrievals using local meteorological data and teleconnection indices.

Roeland Van Malderen[1], Eric Pottiaux[2], Gintautas Stankunavicius[3], Steffen Beirle[4], Thomas Wagner[4], Hugues Brenot[5], and Carine Bruyninx[2]

VILNIUS UNIVERSITY
3

MAX PLANCK INSTITUTE FOR CHEMISTRY
4

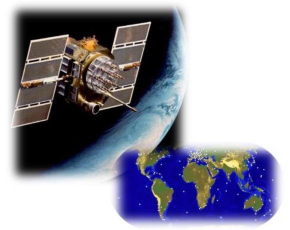
ROYAL BELGIAN INSTITUTE FOR SPACE AERONOMY
5

ABSTRACT

Being the most important natural greenhouse gas and responsible for the largest known feedback mechanism for amplifying climate change, the role of **water vapour** is crucial in a warming climate. Atmospheric water vapour is highly variable, both in space and in time. Therefore, measuring it remains a demanding and challenging task. As a consequence, in this study, **three different datasets** have been taken into account. At **118 globally distributed Global Positioning System (GPS) sites**, Integrated Water Vapour (IWV) is retrieved from a homogeneous data reprocessing from **1995-2010**. At those site locations, also UV/VIS IWV satellite retrievals by GOME, SCIAMACHY and GOME-2 (= GOMESCIA), and ERA-Interim reanalysis output is used to study the time variability of the IWV.

The IWV seasonal behaviour and the inter-annual variability are fitted together by means of a **stepwise multiple linear regression** of the station's time series, with a selection of regionally dependent candidate explanatory variables. Overall, the variables that are most frequently used and explain the largest fractions of the IWV variability are the **surface temperature** and **precipitation**. Also the surface pressure and tropopause pressure (in particular for higher latitude sites) are important contributors to the IWV time variability. All these variables also seem to account for the sign of long-term trend in the IWV time series to a large extent, when considered as explanatory variable. Furthermore, the multiple linear regression linked the IWV variability at some particular regions to **teleconnection patterns or climate/oceanic indices** like the North Oscillation index for West USA, the El Niño Southern Oscillation (ENSO) for East Asia, the East Atlantic (associated with the North Atlantic Oscillation, NAO) index for Europe.

1. DATASETS



ECMWF

ground-based GPS

- 118 sites worldwide with homogeneous data processing from 1995-Mar 2011 (IGS repro 1)
- time delay measurement, ZTD → IWV needs P_{surf} and T_{mean}
- Δt = 5', but in practice: every 6h

ERA-Interim

- worldwide reanalysis from the ECMWF
- IWV from surface fields, corrected to GNSS station height, horizontal interpolation
- 0.75° × 0.75°
- Δt = 6h

GOME/SCIAMACHY/GOME-2

- nadir spectroscopy of back-scattered light (around 700 nm for water vapour)
- global coverage, here: GOMESCIA satellite overpass measurements at GPS stations
- available as 1° × 1° grid
- mid 1995 to present
- Δt = 1 month (climate product, Beirle et al., 2018)

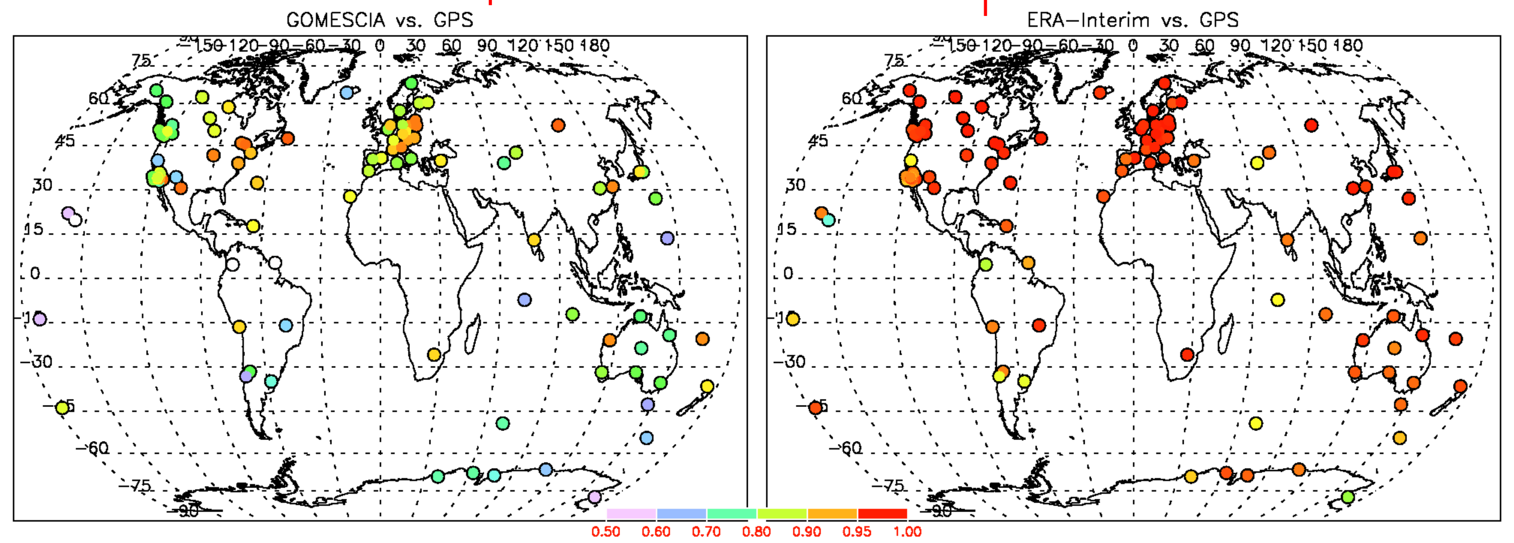


Fig. 1: Correlation coefficients between the monthly means of GOMESCIA and GPS (left) and ERA-Interim and GPS (right).

- GOMESCIA differs most from the other 2 datasets (average R^2 with GPS is 0.865, while the average R^2 between ERA-Interim and GPS is equal to 0.975): coarser horizontal resolution!
- worst correlations with GPS for island and coastal sites: spatial representation of the IWV field at the GPS site by GOMESCIA & ERA-Interim can be questioned here.

2. METHODOLOGY: STEPWISE MULTIPLE LINEAR REGRESSION

MULTIPLE LINEAR REGRESSION

The monthly mean IWV time series are fitted by a multiple linear regression model:

$$Y(t) = A_0 + A_1 t + \sum_{i=2}^{n} A_{seas,i} X_{seas,i}(t) + \sum_{j=0}^{m} B_j X_j(t) + \epsilon(t)$$

with

- t the timestep (one month),
- A_0 the intercept,
- A_1 the annual trend,
- $A_{seas,i}\, X_{seas,i}\,(t)$ describing the seasonal cycle of the IWV, either by using the long-term monthly means or by harmonic functions,
- $X_j(t)$ the monthly mean time series from explanatory variables like surface temperature, surface pressure, tropopause pressure, precipitation, teleconnection indices & the monthly mean time series generated from teleconnection indices shifted with 1 to 6 months back in time,
- B_j their respective coefficients,
- m denoting the number of candidate explanatory variables, depending on the geographical region (ranges between 103 and 194),
- $\epsilon(t)$ representing the residuals.

STEPWISE

Regional:

1. correlation analysis between IWV (from NCEP/NCAR reanalysis) and different teleconnection indices for different regions to determine potential explanatory variables per region.
2. check if those potential explanatory variables are independent (also with the local meteorological variables).

For every site (IWV time series Y(t)):

3. rank the explanatory variables by decreasing impact on Y(t), i.e. explained variability by single linear regression for IWV.
4. step-by-step, add those ranked explanatory variables to the multiple linear regression and test the statistical significance of an included variable by means of a t-test of the regression coefficient.

EXAMPLES

In Fig.2, we show 2 examples of a good and bad fit to the IWV time series.

- **(left)** for the GPS IWV time series at DUBO (Lac Du Bonnet, Canada), the multiple linear regression fit explains 98.64% of the variability or a correlation coefficient of 0.993 is obtained. Despite the very high percentage of explained variance, a significant positive trend is still present in the residual time series (although the annual trend was not retained as a significant explanatory variable in the multiple linear regression). Next to the 5 explanatory variables shown in Fig. 3, also P_{surf}, NP, EAWR (preceding 1 month), AO (preceding 5 months), NOI (preceding 6 months), WP, NAO, SOI (preceding 5 months) have been used.
- **(right)** for the GOMESCIA IWV time series at HOB2 (Hobart, Australia), the multiple linear regression results in an explained variability of 54.55% and a correlation coefficient of 0.739. Additionally to the 5 explanatory variables shown in Fig. 3, also the surface pressure was included.

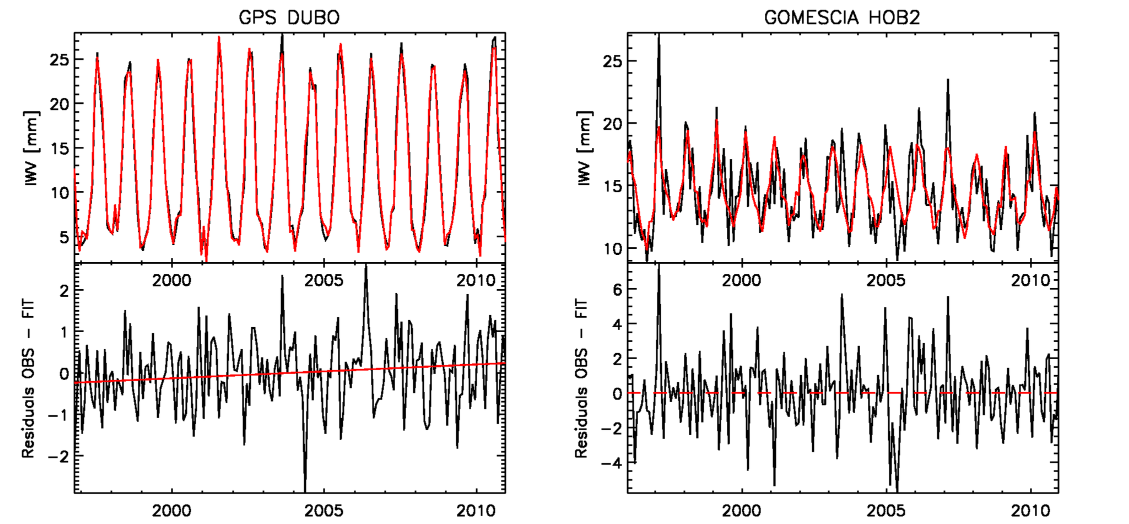


Fig. 2: Examples of the stepwise multiple linear regression fits (in red) to (left) the GPS monthly mean IWV time series of DUBO (Lac Du Bonnet, Canada) and (right) the GOMESCIA IWV time series of HOB2 (Hobart, Australia). The lower panels show the residuals between the observations and fitted time series of the upper panels (black minus red), with a linear fit to the residuals in red (positive trend in both cases, but only significant (full line) for DUBO).

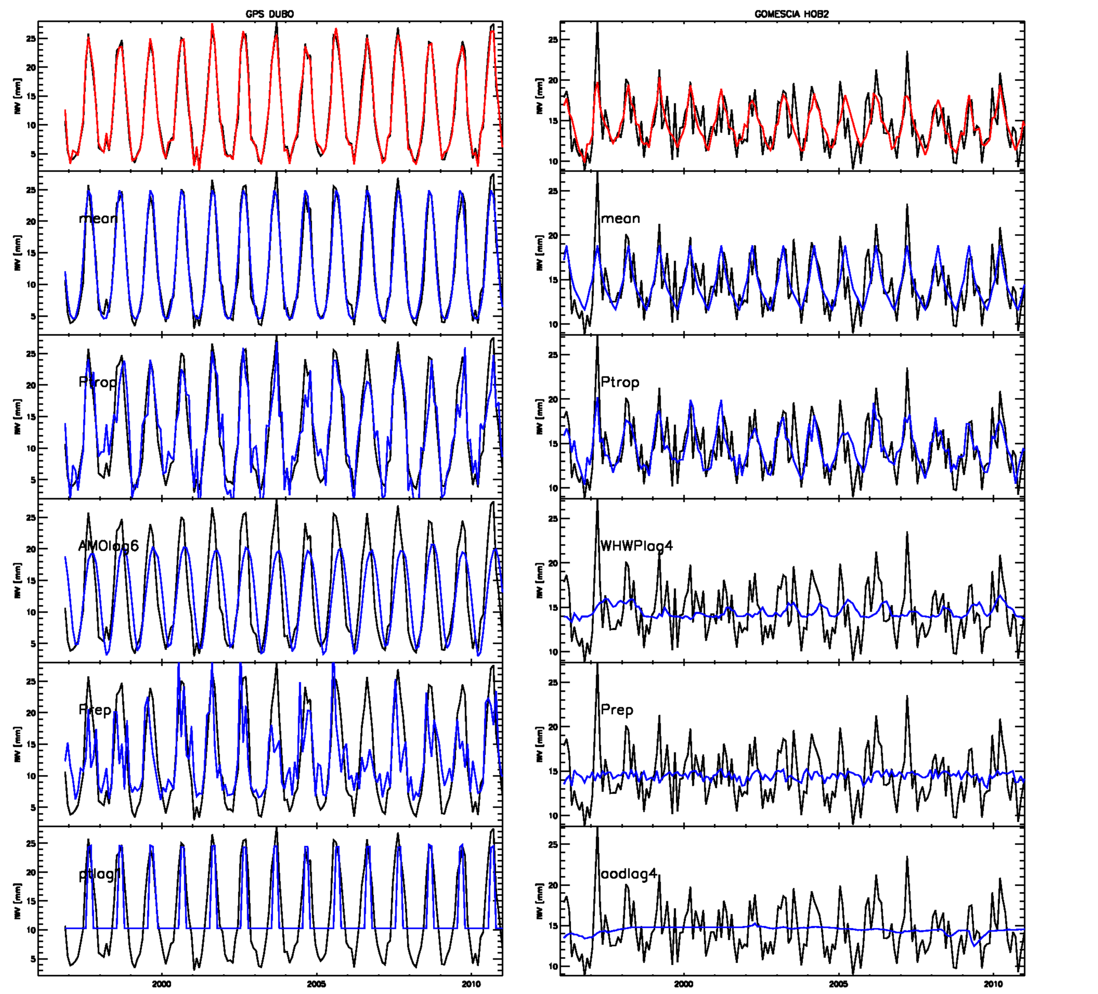


Fig. 3: The time series of the 5 most important explanatory variables used for the multiple linear regression model fit of DUBO (left) and HOB2 (right). The coefficients used here to scale the time series to the IWV observations (black) are determined from the single linear regression (just for illustration here). The explanatory variables used are the long-term means, tropopause pressure, AMO (especially preceding with 6 months), precipitation, and PT (preceding 1 month) for DUBO, and the long-term means, tropopause pressure, WHWP (preceding 4 months), precipitation, and AOD (preceding 4 months) for HOB2.

3. RESULTS

Overall:

- the best fits are obtained for the ERA-Interim output (see Fig 3, on average explaining 90% of the IWV variability), and the "worst" fits for the GOMESCIA dataset (on average 85.5%), with the GPS time series there in between (88.9%).
- The higher ERA-Interim explained variance is also obtained by including a higher number of explanatory variables (on average 8.2) in the linear regression than for GPS (8.0) and GOMESCIA (6.8).
- the most important explanatory variables (next to the long-term means) are the surface temperature, precipitation, surface pressure and tropopause pressure, both in their frequency of occurences and in the explained variability.
- The lower mean explained variability for the GOMESCIA dataset seems to be linked with the significant lower percentage of stations for which the precipitation and the surface pressure are included in the linear regression (resp. around 40% and 25%), in comparison with the other two datasets (resp. above 70% and around 40%).

Regional:

- A summary of the dominant explanatory variables for different regions (i.e. explanatory variable significantly contributes to IWV variability for at least half of the stations of this region, and this for the three IWV datasets) is given in Fig. 4.
- Consistent patterns arise: the **North Pacific** index for West USA, Canada and East Asia, the **West Pacific** index for a majority of the West USA sites, the **Pacific Transition** index arises for Australia, and to a lesser extent for Latin America, **ENSO** is well established over East Asia, but also present in Australian and Latin American sites. For the West USA, the **North Oscillation Index** is very dominant, and the **Arctic Oscillation** in Canada. The **NAO index** could be linked to the IWV variability at Canada as well. In Europe. In this last continent, the **East Atlantic**, **Polar/Eurasia**, **Tropical/Northern Hemisphere** and **Atlantic Multidecadal Oscillation** indices are important contributors to the IWV variability.

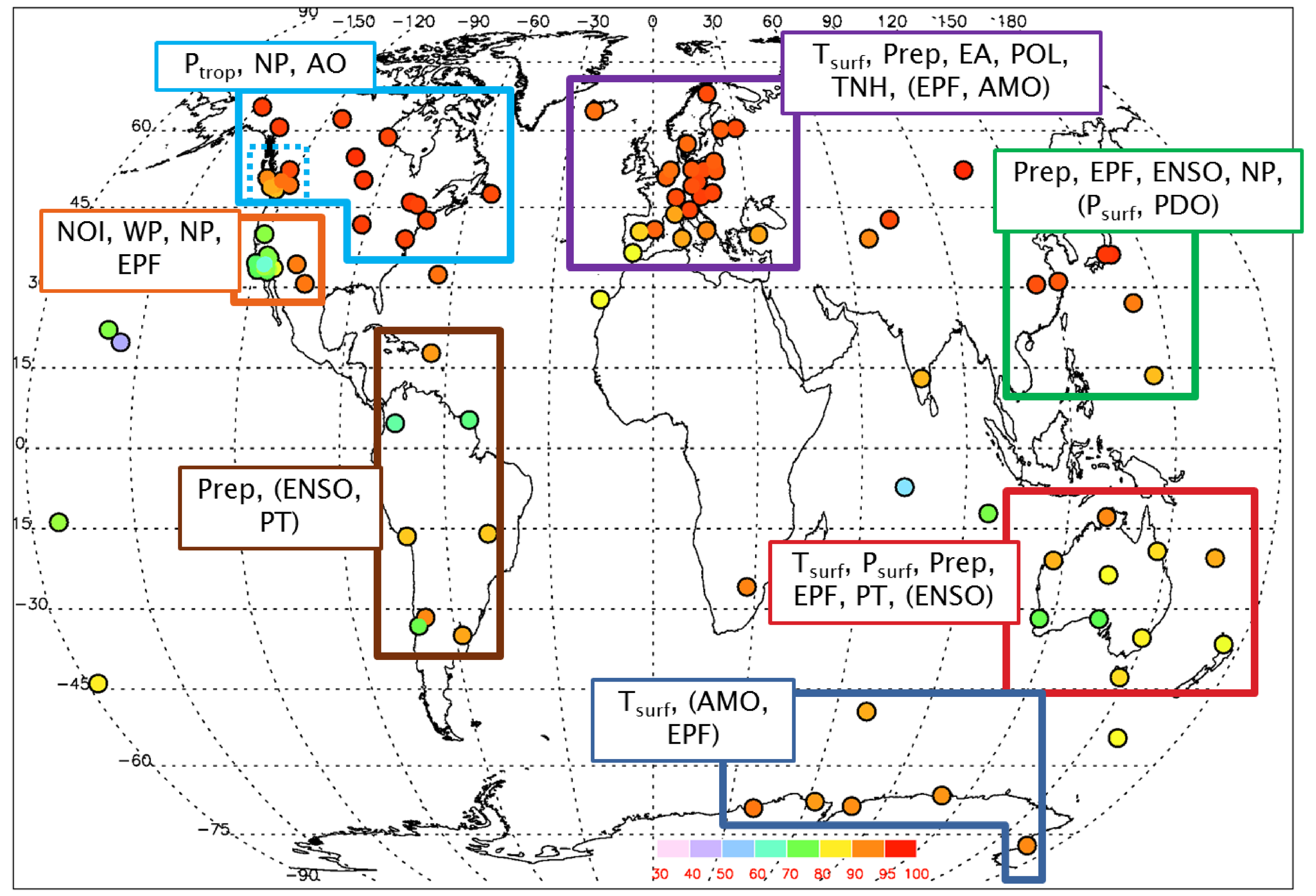


Fig. 4: Summary of the statistically significant explanatory variables that account for the variability of the IWV time series of the three datasets for at least half of the stations for different geographical regions. Between brackets: explanatory variables not entirely meeting these criteria, but still dominant in two datasets and/or in a considerable amount of the stations. The colours of the dots indicate the explained variability by the multiple linear regression of the ERA-Interim IWV time series.

4. CONCLUSIONS

- With the (empirical) stepwise multiple linear regression of the IWV monthly mean time series, we aimed at finding the most relevant variables to explain the IWV variability for different regions. As we started from a large selection of potential explanatory variables (i.e. 100 to 200 variables if allowing their time series to precede the IWV time series by 1 to 6 months), but only end up with a considerable small amount of effectively included explanatory variables (around 7 to 8 on average), with a reasonable geographical footprint (see the summary Fig. 4), and mostly consistent among the three datasets used, we are confident in our approach.
- In particular, it could be confirmed that, apart from the harmonics or long-term means that account for a large part of the seasonal variability, the bulk of the remaining IWV time variability linked with the surface temperature and precipitation. The precipitation is included by a statistical t-test in the multiple linear regression for the largest number of sites, but with a lower explained fraction of the IWV variability than the surface temperature. The other site specific explanatory variables, the surface pressure and tropopause pressure, are also important contributors to the multiple linear regression, the latter seems to be more common at high-latitude sites.
- To identify which of those explanatory variables accounts most for the trend sign of the IWV time series is not straightforward: as a matter of fact, we found that each of those variables results in a regression term with equal trend sign as the IWV trend sign for about 70% of the cases, and this for the three datasets.
- This analysis could be repeated for the whole globe, making use of the gridded IWV datasets (GOMESCIA and ERA-Interim/ERA5).

REFERENCES AND ACKNOWLEDGEMENTS

This research has been carried out in the framework of the Solar-Terrestrial Centre of Excellence (STCE). We are grateful to all colleagues and data providers below:

SOLAR-TERRESTRIAL CENTRE OF EXCELLENCE
INTERNATIONAL GNSS SERVICE
ECMWF

REFERENCES:

- Beirle, S., Lampel, J., Wang, Y., Mies, K., Dörner, S., Grossi, M., Loyola, D., Dehn, A., Danielczok, A., Schröder, M., and Wagner, T.: The ESA GOME-Evolution "Climate" water vapor product: a homogenized time series of H2O columns from GOME, SCIAMACHY, and GOME-2, Earth Syst. Sci. Data, 10, 449-468, https://doi.org/10.5194/essd-10-449-2018, 2018.

FURTHER READING:

- Van Malderen, R., Pottiaux, E., Stankunavicius, G., Beirle, S., Wagner, T., Brenot, H., and Bruyninx, C.: Interpreting the time variability of world-wide GPS and GOME/SCIAMACHY integrated water vapour retrievals, using reanalyses as auxiliary tools, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2018-1170, in review, 2018.