RMI

Koninklijk Meteorologisch Instituut

Institut Royal Météorologique

Königlische Meteorologische Institut

Royal Meteorological Institute

Break detection in integrated water vapour benchmark datasets

R. Van Malderen, E. Pottiaux, A. Klos, P. Domonkos, M. Elias, T. Ning,
O. Bock, J. Guijarro, F. Alshawaf, M. Hoseini, A. Quarello, E. Lebarbier,
B. Chimani, V. Tornatore, S. Zengin Kazancı, and J. Bogusz

.be

Outline

1. Introduction
2. Synthetic benchmark dataset generation
3. Break detection methods
4. Performance
    1. Accuracy of break point positions
    2. Centered RMSE
    3. Trend differences
5. Conclusions

Introduction

- surface warming
- warm air can contain more water vapour than cold air (Clausius Clapeyron)
- Integrated Water Vapour (IWV) or Precipitable Water Vapour (PWV) amounts are increasing?
- GPS IWV retrievals are providing a worldwide, long-term dataset

Introduction

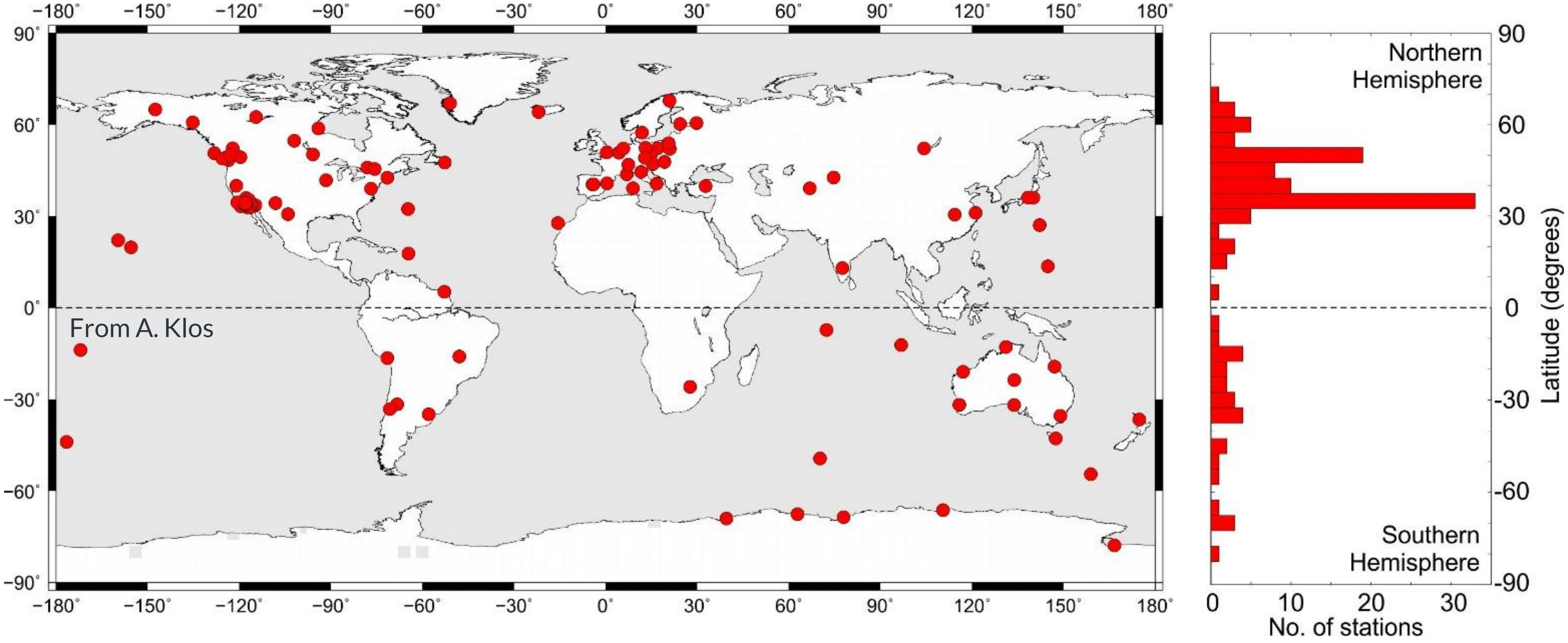
ZTD → IWV

Daily Observations

120 Stations

Period: 1995-2010

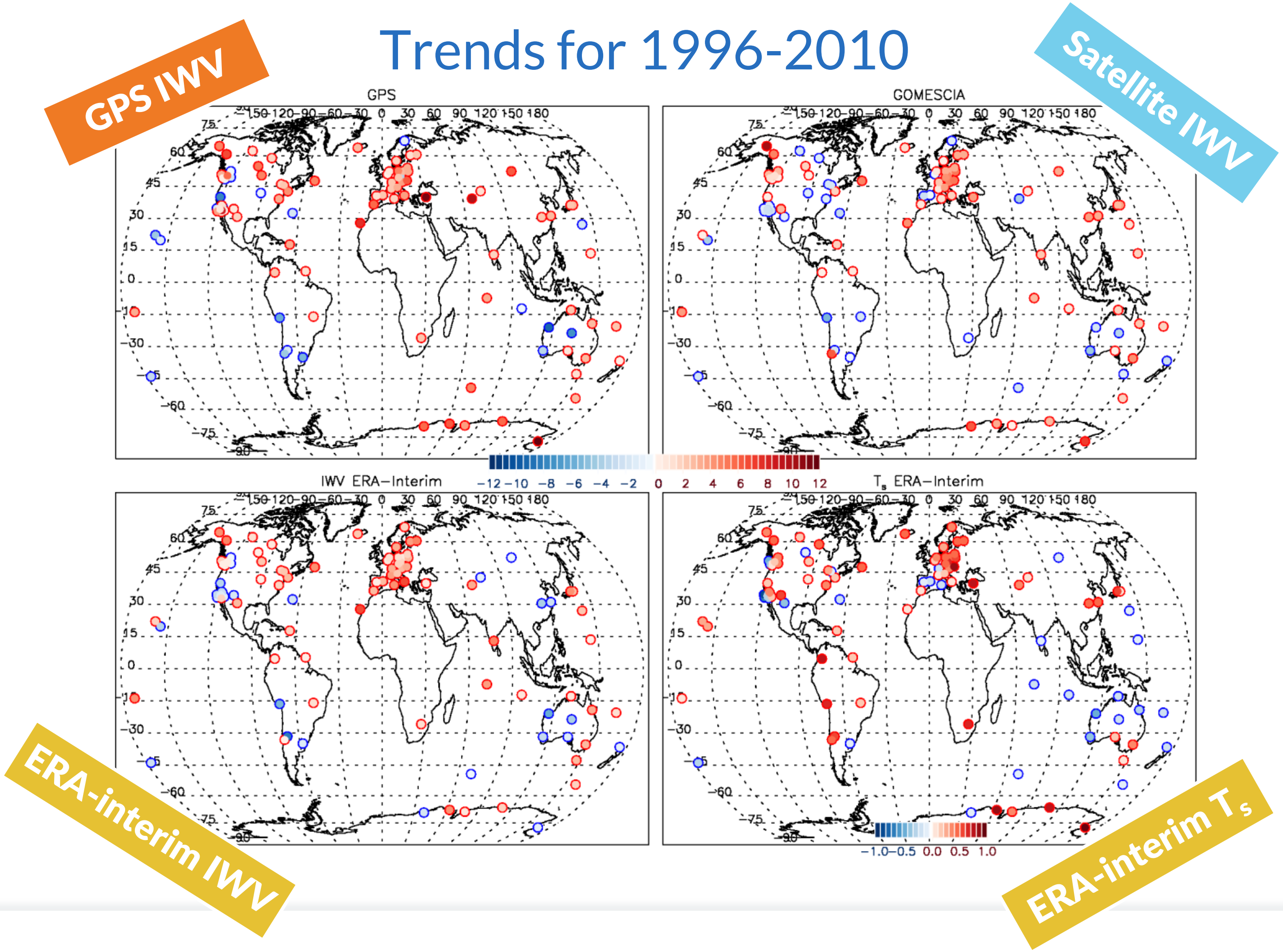
Screened and converted to Integrated Water Vapor (IWV) by O. Bock.



From A. Klos

Introduction

Trends for 1996-2010



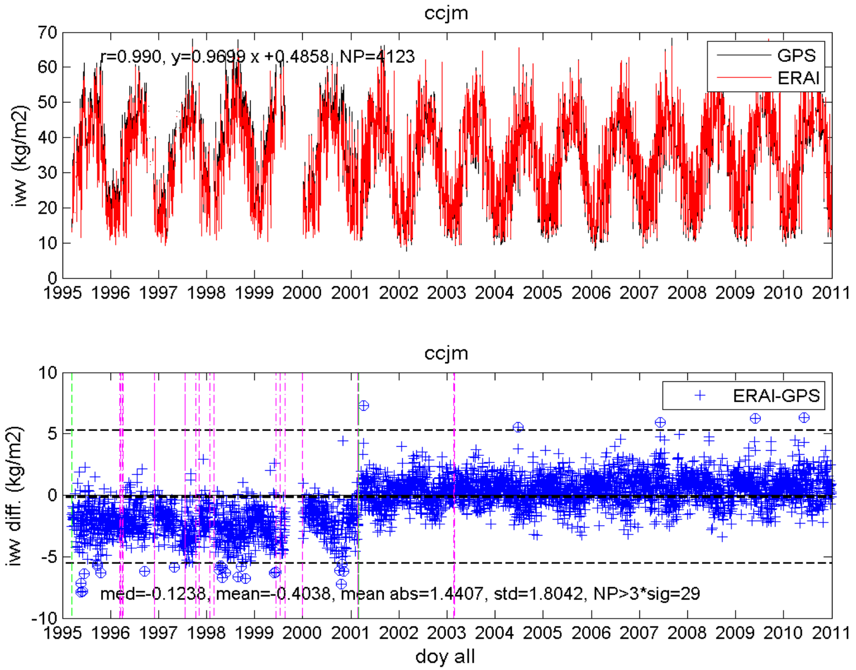
- IWV trends follow T_s trends globally
- differences between different datasets (but GPS and ERA-Interim not too different)
- due to inhomogeneities in datasets?
- COST action

GPS IWV

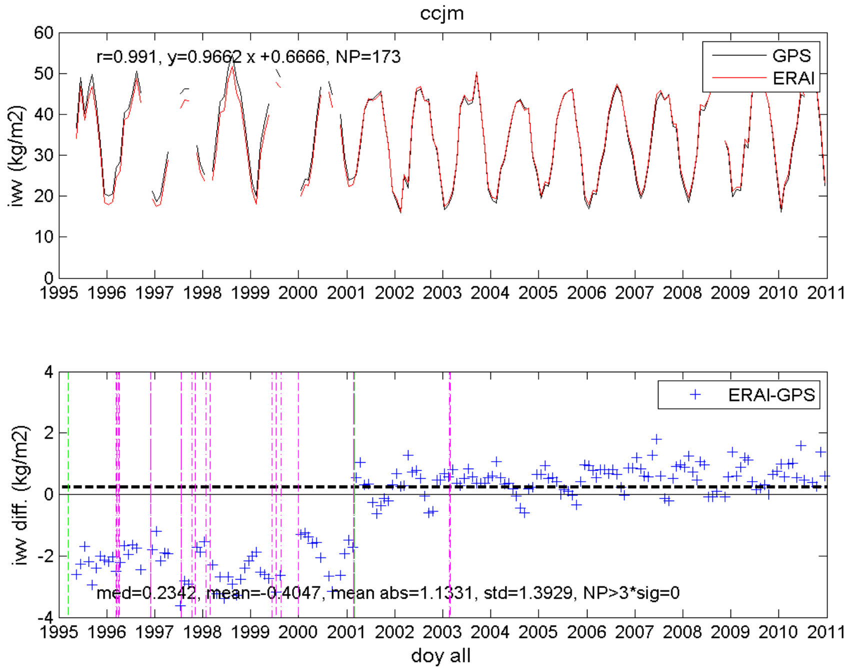
Introduction

CCJM: Ogasawara, Japan

Daily



Monthly



From O. Bock

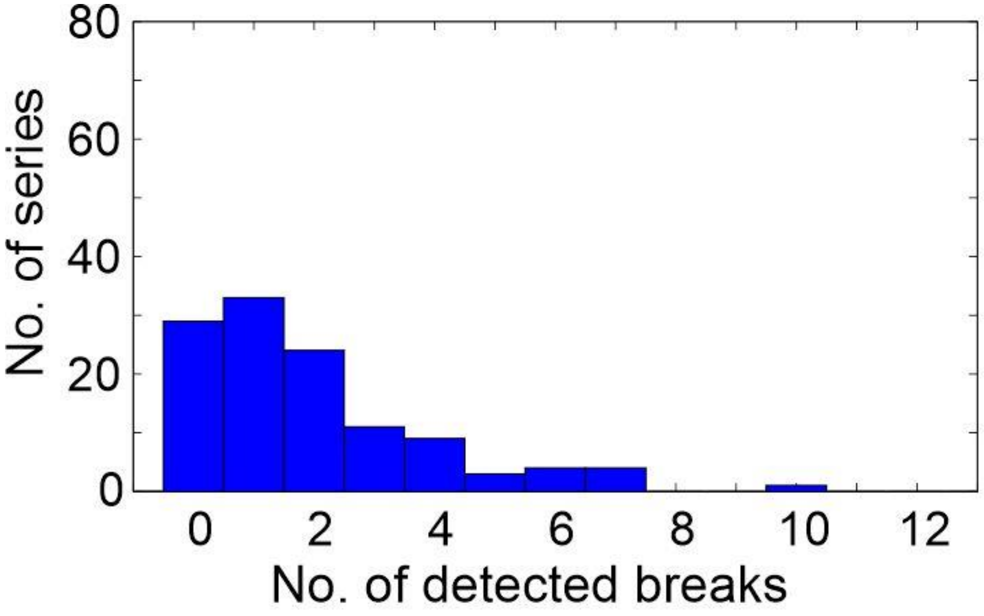
→ We will look for break points in the ERA-interim-GPS IWV differences time series

ERA-interim IWV **GPS IWV**

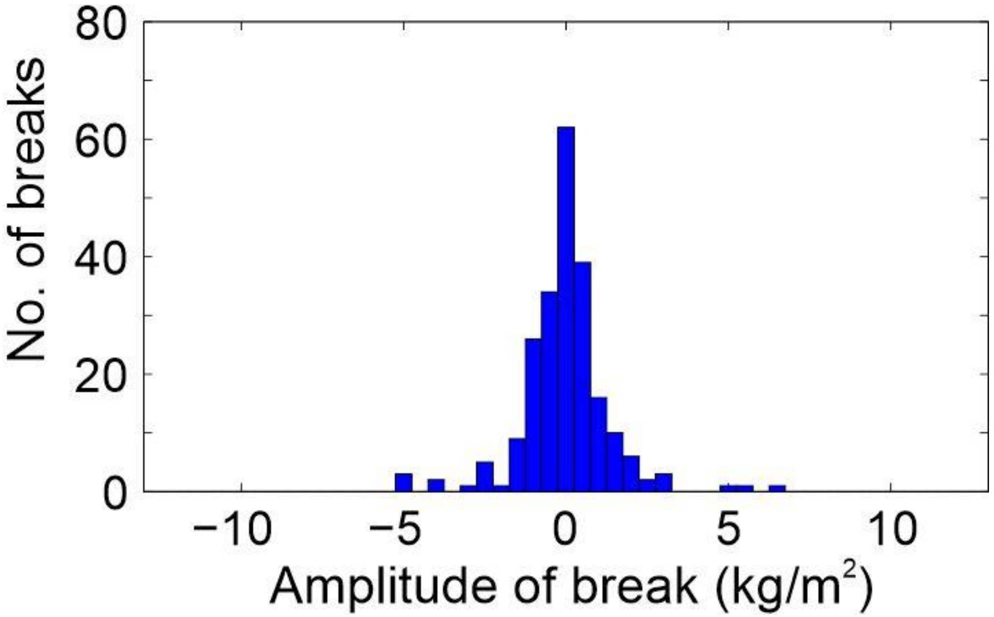
Synthetic benchmark dataset generation

Real IWV Diff. (ERAI-GPS)	Synthetic IWV Diff.
manual homogenization GPS log files (metadata)	characterization of the number and amplitude of breaks (randomly inserted)
power spectra density analysis	significant frequencies (annual, semi-annual…)
noise analysis	noise model: AR(1) + WN
non-climatic trend analysis	characterization of non-climatic trends (reference series)

Synthetic benchmark dataset generation



on average: 1.93 breaks/time series



most breaks have amplitudes between ±1 kg/m^2

Synthetic benchmark dataset generation

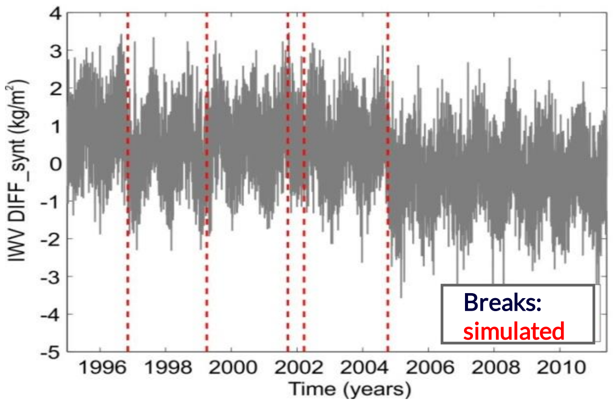
Real IWV Diff. (ERAI-GPS)		**Synthetic IWV Diff.**
manual homogenization GPS log files (metadata)	→	characterization of the number and amplitude of breaks (randomly inserted)
power spectra density analysis	→	significant frequencies (annual, semi-annual…)
noise analysis	→	noise model: AR(1) + WN
non-climatic trend analysis	→	characterization of non-climatic trends (reference series)

Synthetic benchmark dataset generation

3 variants: assess the performances of break detection methods w.r.t. dataset characteristics

EASY

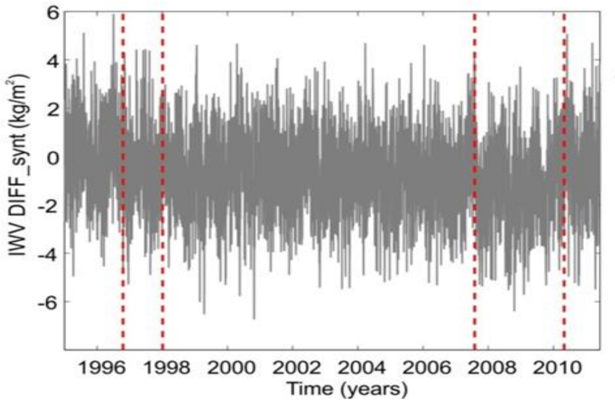
- seasonal signals
- breaks
- white noise (WN)



MODERATE

= EASY but +

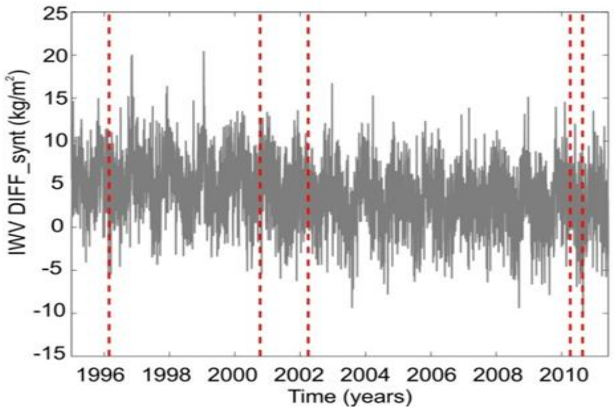
- autoregressive process of the first order (noise model = AR(1)+WN)



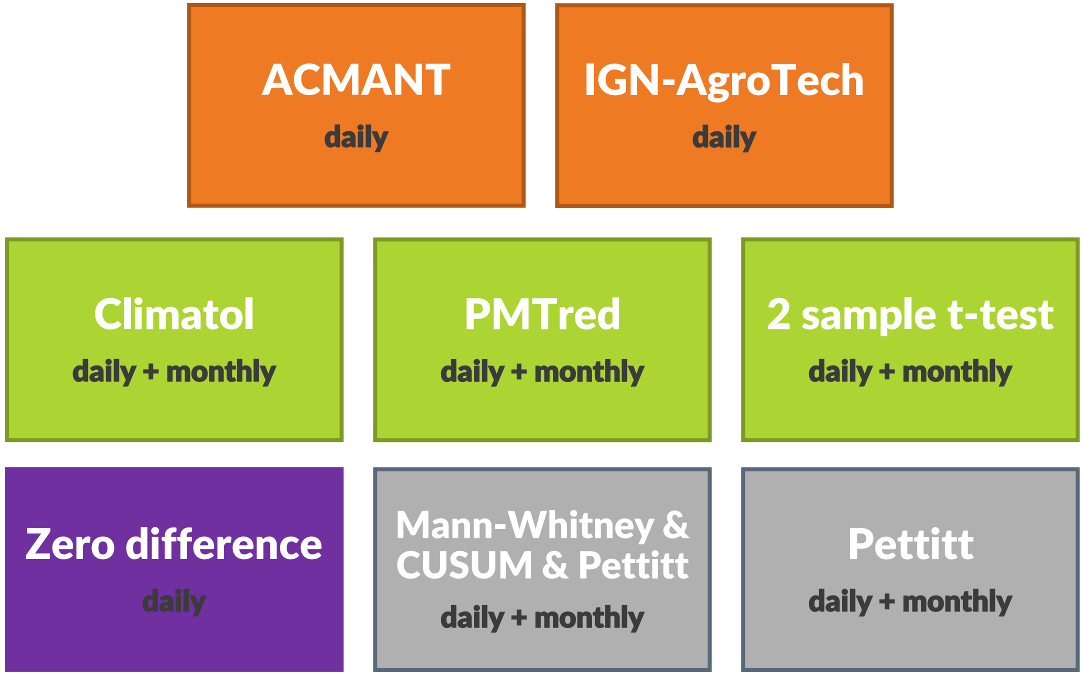
COMPLEX

= MODERATE but +

- gaps (up to 20% of missing data)
- non-climatic trend (Ref. Series)



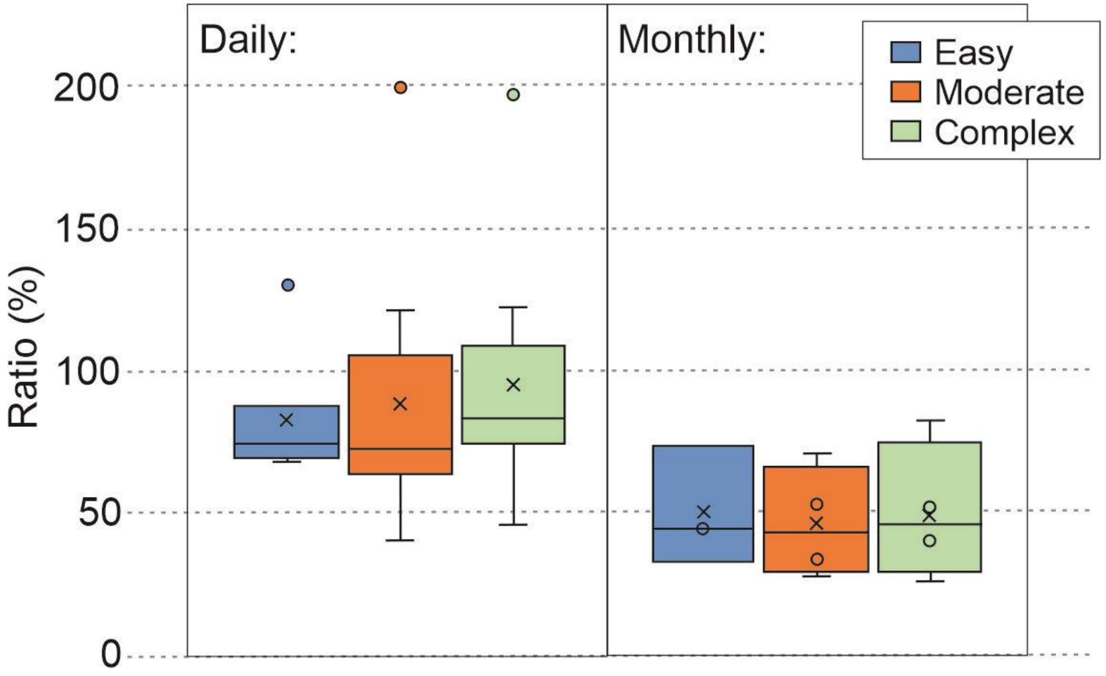
Break detection methods



- 8 break detection methods (7/8 different operators)
- 13 break detection methods (daily+monthly)
- not all of them applied on EASY/MODERATE/COMPLEX datasets
- 4 main types of break detection methods:
  - 🟧 **Maximum Likelihood (ML) multiple break methods**
  - 🟩 **Standard Normal Homogeneity Test methods**
  - 🟪 **Singular Spectrum Analysis (SSA)**
  - ⬜ **Non-parametric methods**

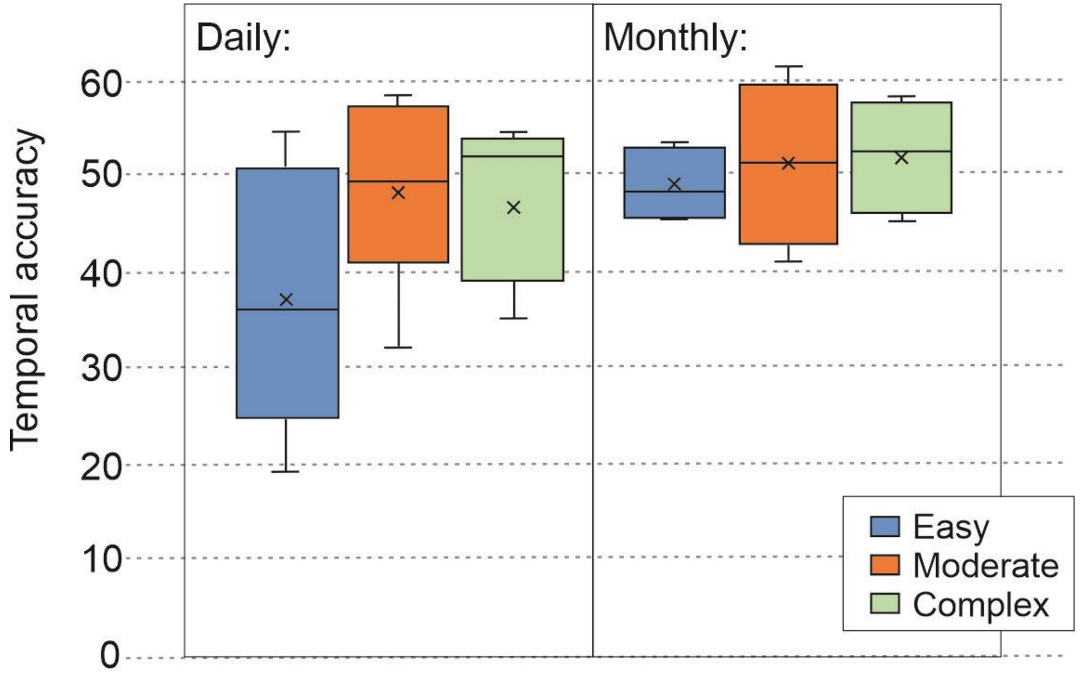
Performance: 1. Accuracy of break point positions

Number of detected breaks



- except NP2d: methods find less breaks than inserted.
- larger ratio of detected breakpoints for daily methods

Time window for break detection

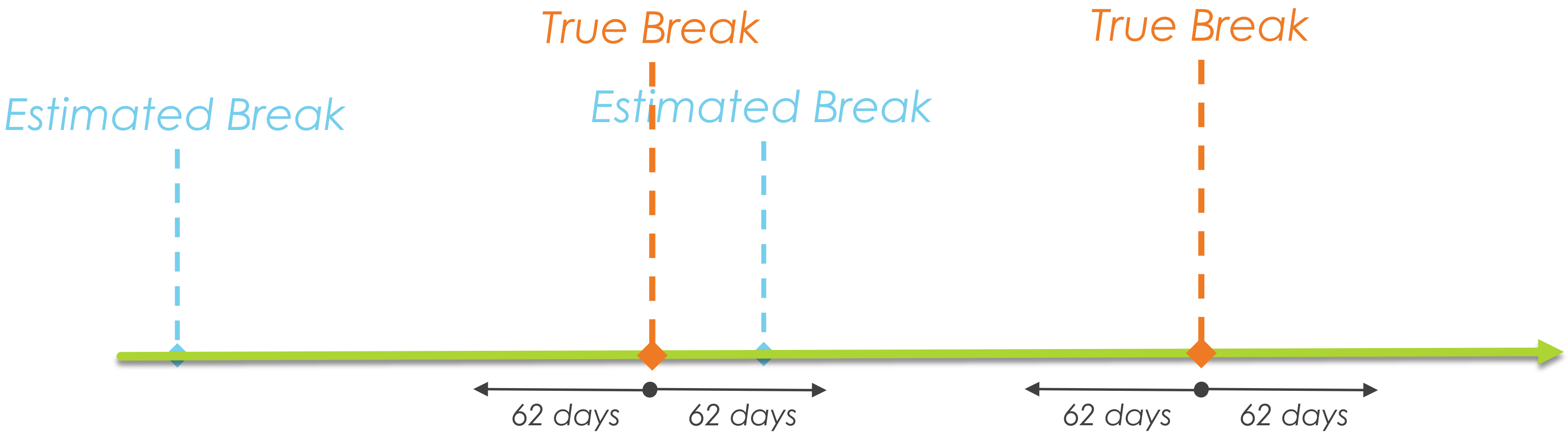


➔ a time window of 62 days seems appropriate

Performance: 1. Accuracy of break point positions

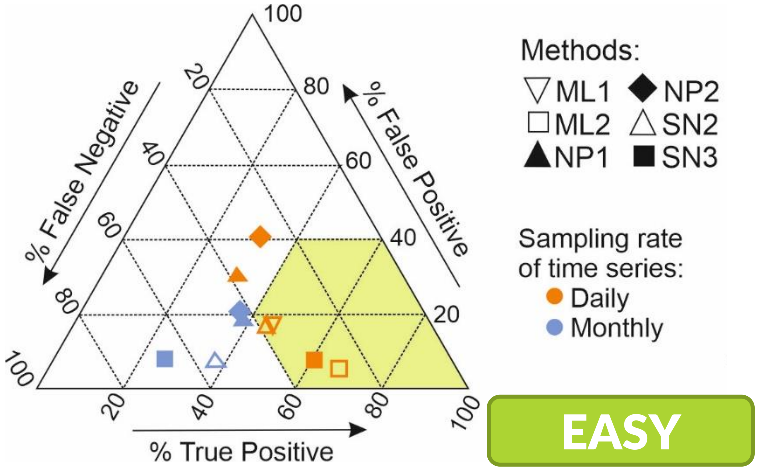
Skill scores

False Alarms	Matches	Misses	
👎 False Positive	👍 True Positive	👎 False Negative	👍 True Negative

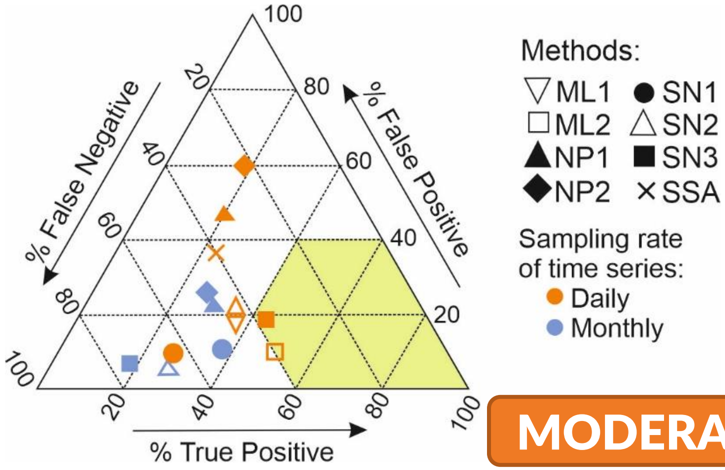


Performance: 1. Accuracy of break point positions

Skill scores

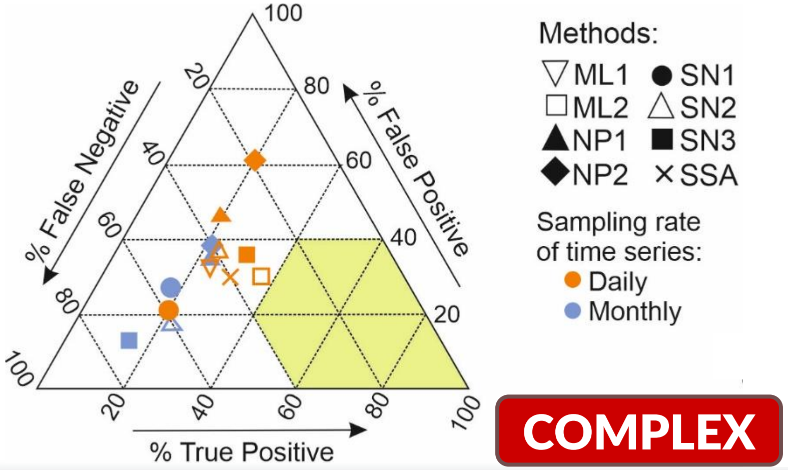


EASY



MODERATE

- with 62d: large number of hits (TP) compensated by large number of false alarms (FP)
- performance decreases with increasing complexity
- best: ML2 and SN3



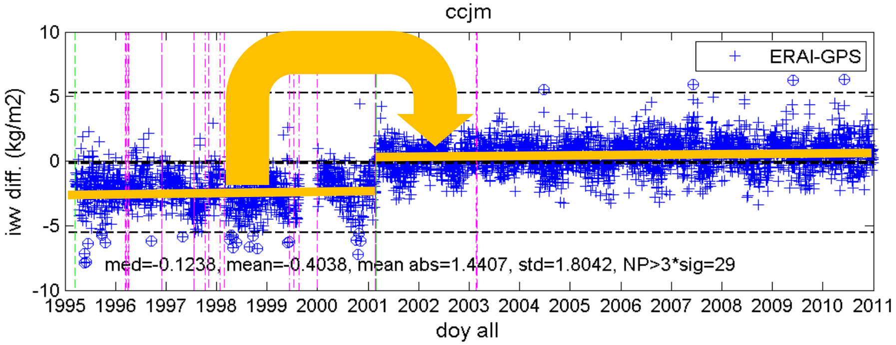
COMPLEX

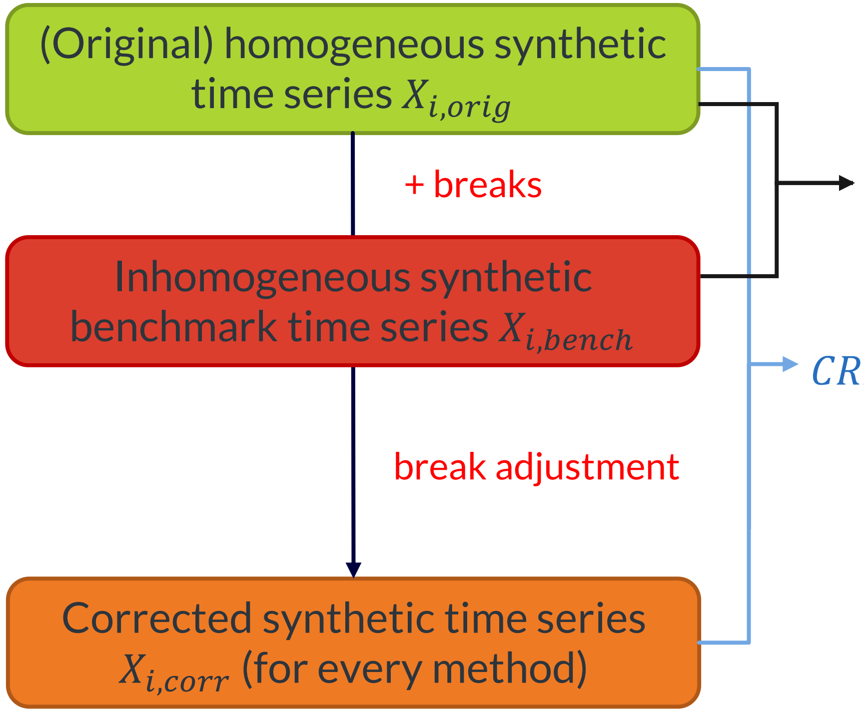
- better scores for N=183 days

Performance: 2. Centered RMSE

We use one common method for adjustment of the time series for the detected inhomogenities:

adjustment to mean of last segment → $X_{i,corr}$





(Original) homogeneous synthetic time series $X_{i,orig}$

+ breaks

Inhomogeneous synthetic benchmark time series $X_{i,bench}$

break adjustment

Corrected synthetic time series $X_{i,corr}$ (for every method)

$$CRMSE \equiv \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\left(X_{i,orig}-\bar{X}_{orig}\right)-\left(X_{i,bench}-\bar{X}_{bench}\right)\right]^2}$$

$$CRMSE \equiv \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\left(X_{i,orig}-\bar{X}_{orig}\right)-\left(X_{i,corr}-\bar{X}_{corr}\right)\right]^2}$$

improvement?

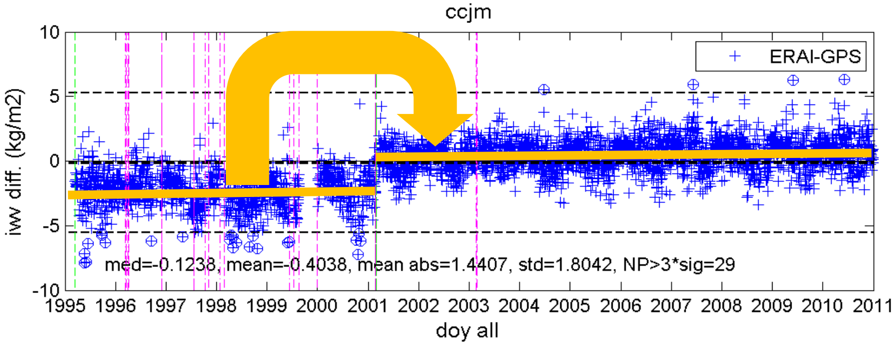
Performance: 2. Centered RMSE

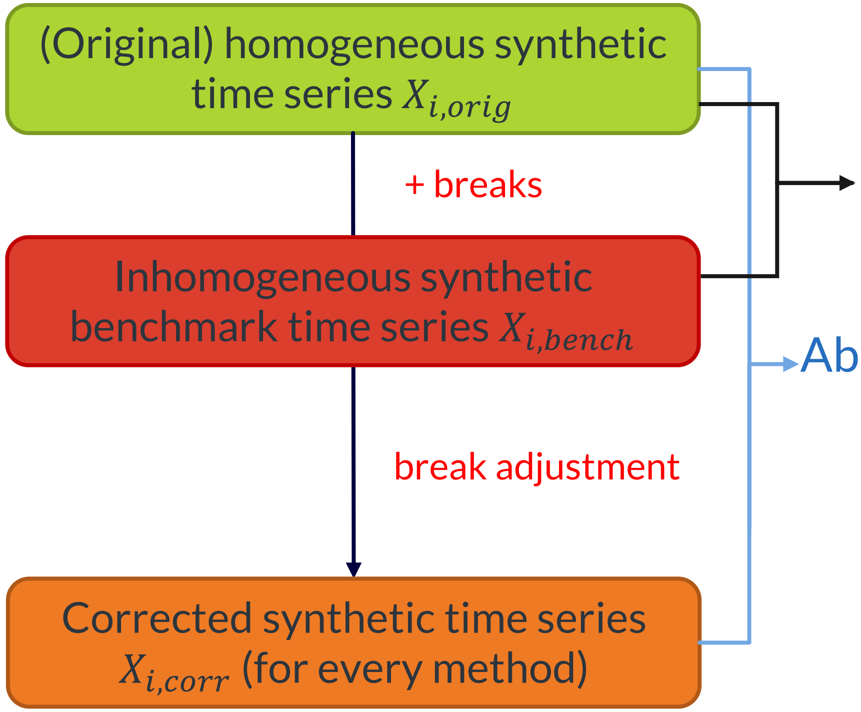
- all break detection methods (+ adjustment) give an improvement w.r.t. the inhomogeneous benchmark time series (“raw data”)!
- improvement decreases with increasing complexity of the datasets
    - Easy: 71% (55-85%)
    - Moderate: 63% (45-75%)
    - Complex: 28% (19-35%)
    - largest decrease for Moderate→Complex
- CRMSE (& improvements) very similar for adjusted daily and monthly time series for a given method
- best methods remain ML2d and SN3d

Performance: 3. Trend differences

We use one common method for adjustment of the time series for the detected inhomogenities:

adjustment to mean of last segment → $X_{i,corr}$





(Original) homogeneous synthetic time series $X_{i,orig}$

+ breaks

Inhomogeneous synthetic benchmark time series $X_{i,bench}$

break adjustment

Corrected synthetic time series $X_{i,corr}$ (for every method)

Abs. Trend Bias $\equiv abs(Trend_{orig} - Trend_{bench})$

Abs. Trend Bias $\equiv abs(Trend_{orig} - Trend_{corr})$

improvement?

Performance: 3. Trend differences

- all break detection methods (+ adjustment) give smaller trend biases as compared to the inhomogeneous benchmark time series ("raw data")!

- trend bias improvement decreases with increasing complexity of the datasets
    - Easy: 91% (84-95%)
    - Moderate: 87% (72-94%)
    - Complex: 27% (17-36%)
    - largest decrease for Moderate → Complex

- different methods are best performing

Conclusions

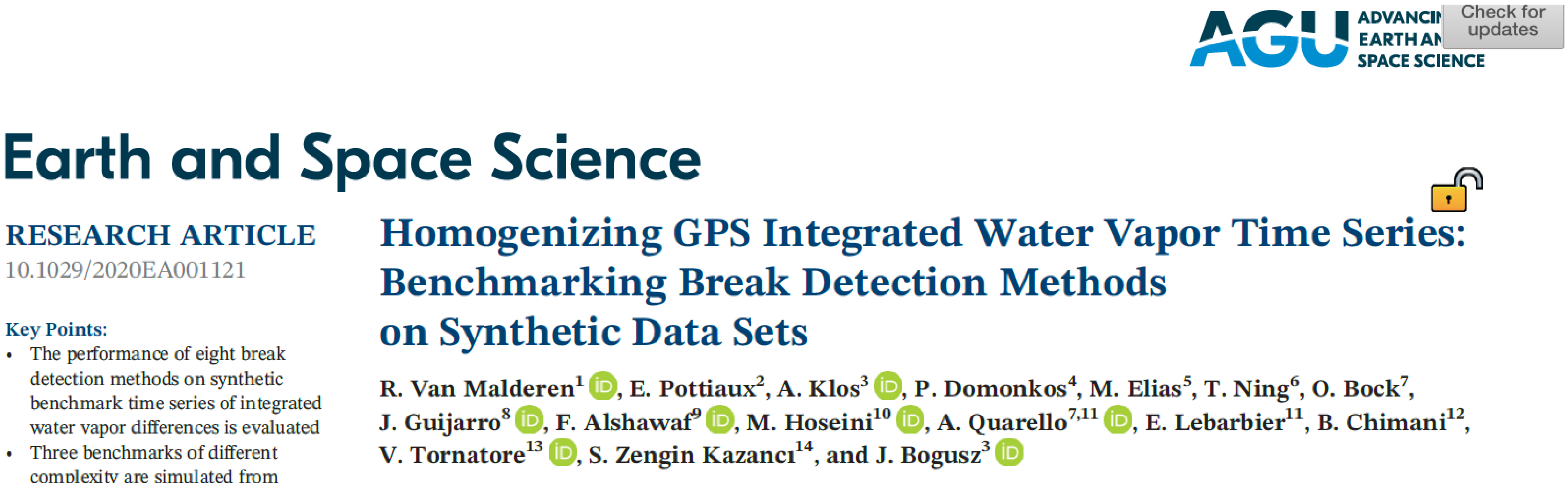
- methods perform well in detecting the inserted breaks, but rather high number of false break detections (especially for Complex)
- after adjustment, significant improvement in time series, both in terms of CRMSE and trend errors (especially for Easy + Moderate)
- poorer performance for Complex due to gaps or **trends**?
- Complex: closest to the real IWV homogenization task, but higher improvement expected in the real IWV homogenization due to
    - both the ERAI trend bias and data gap problems are overshot in Complex
    - the use of metadata in real IWV homogenization.

Conclusions

- 2 best methods represent 2 different classes of methods (ML and SNHT), so no best performing class
- those 2 best methods have been applied on the daily series, but differences of efficiencies between daily and monthly versions of the same method is often surprisingly high.

Conclusions

This research has been published in *Earth and Space Science* (AGU journal).



AGU ADVANCING EARTH AND SPACE SCIENCE

Check for updates

Earth and Space Science

RESEARCH ARTICLE

10.1029/2020EA001121

Homogenizing GPS Integrated Water Vapor Time Series: Benchmarking Break Detection Methods on Synthetic Data Sets

R. Van Malderen[1], E. Pottiaux[2], A. Klos[3], P. Domonkos[4], M. Elias[5], T. Ning[6], O. Bock[7], J. Guijarro[8], F. Alshawaf[9], M. Hoseini[10], A. Quarello[7,11], E. Lebarbier[11], B. Chimani[12], V. Tornatore[13], S. Zengin Kazancı[14], and J. Bogusz[3]

Key Points:

- The performance of eight break detection methods on synthetic benchmark time series of integrated water vapor differences is evaluated
- Three benchmarks of different complexity are simulated from

THANK YOU

Het Koninklijk Meteorologisch Instituut

L’Institut Royal Météorologique

The Royal Meteorological Institute

Het KMI verleent een betrouwbare dienstverlening aan het publiek en de overheid gebaseerd op onderzoek, innovatie en continuïteit.

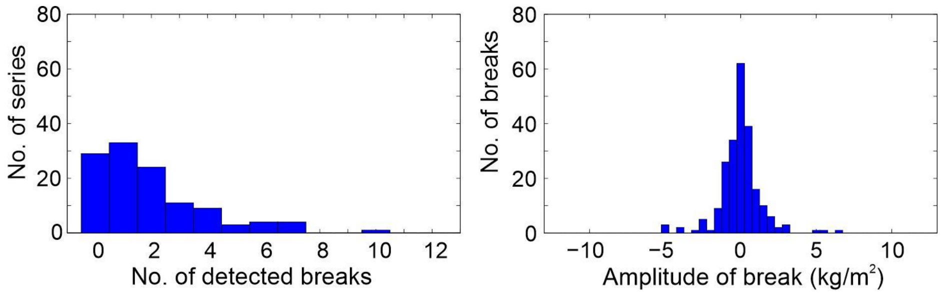
L'IRM fournit un service fiable basé sur la recherche, l'innovation et la continuité au public et aux autorités.

The RMI provides reliable public service realized by empowered staff and based on research, innovation and continuity.

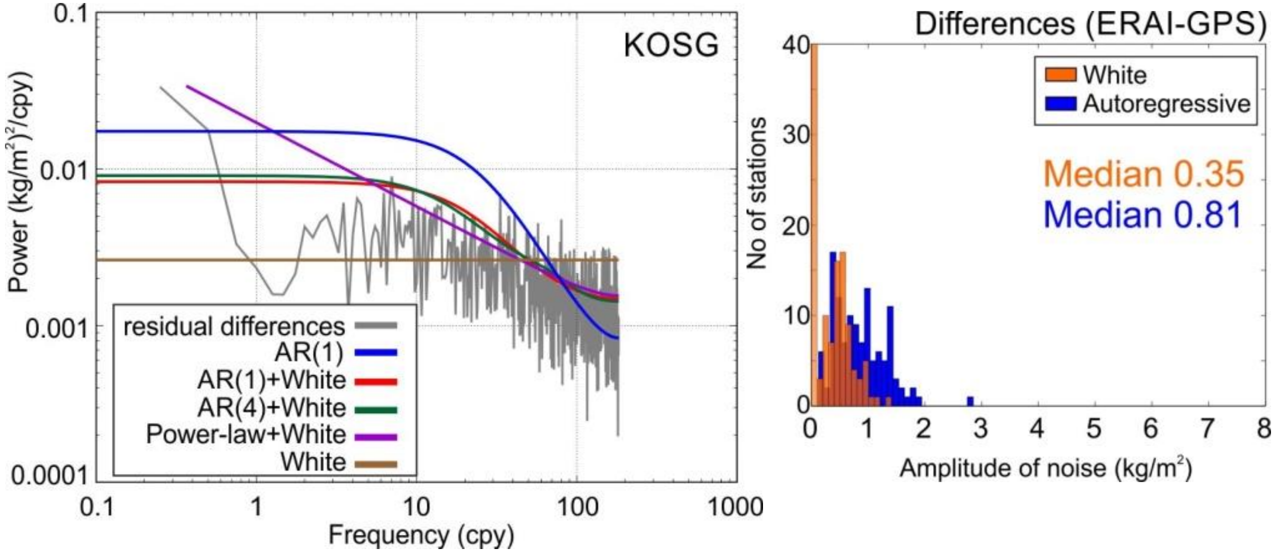
KMI - IRM

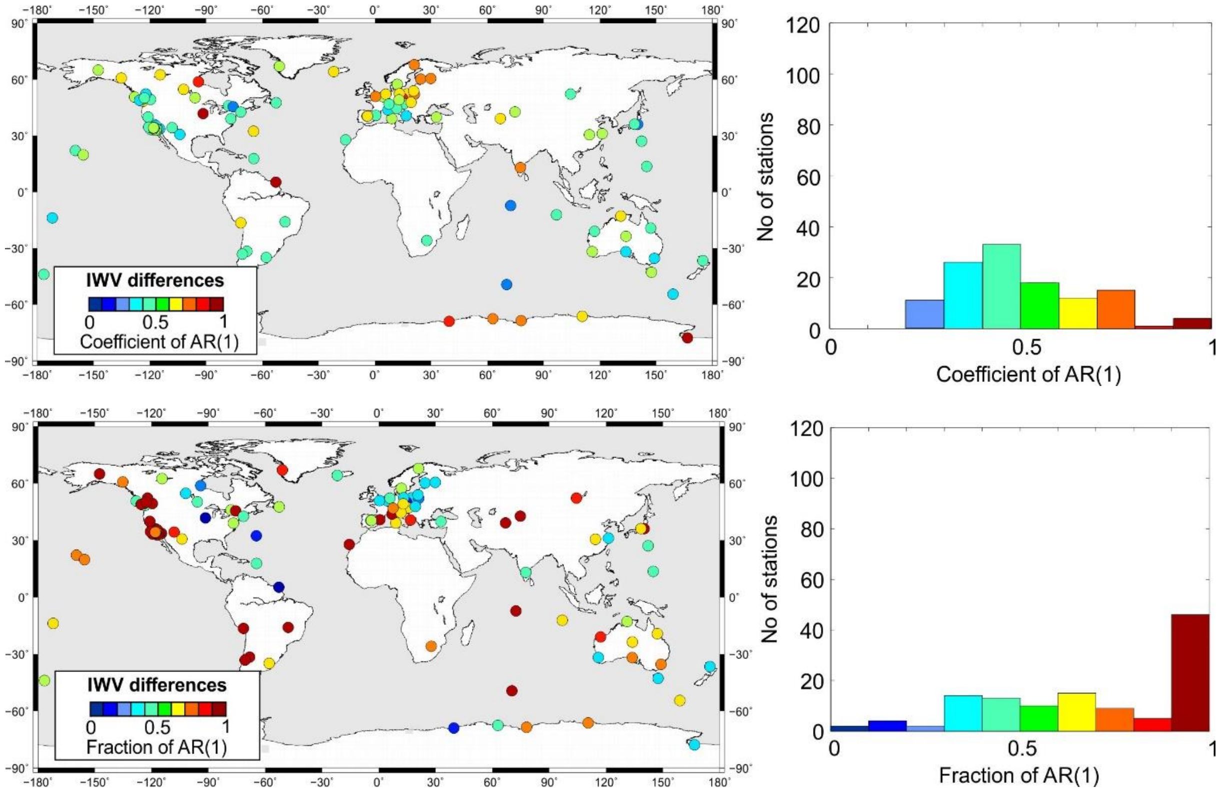
Synthetic benchmark dataset generation

Statistics of detected breaks



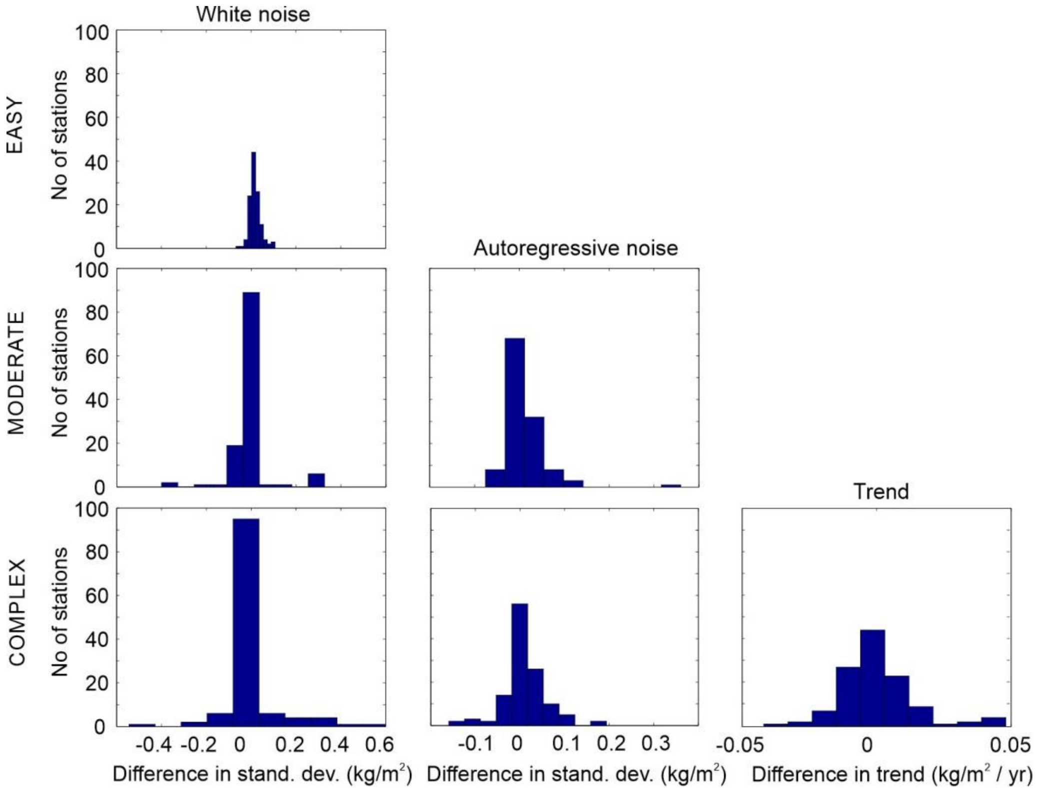
Noise model fitting



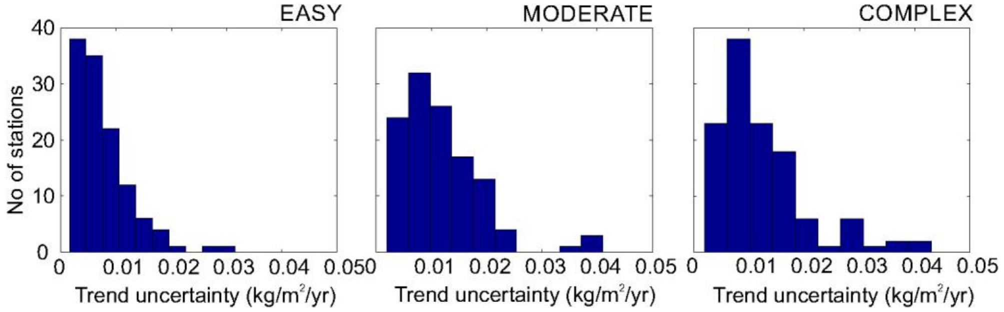


Coefficients (up) and fractions (bottom) of first order autoregressive process for IWV differences.

Synthetic benchmark dataset generation



Verification of the synthetic benchmark time series by comparison with the real IWV differences



Expected trend uncertainty estimations