11th Seminar for Homogenization and Quality Control in Climatological Databases and 6th Interpolation Conference jointly organized with the 14th EUMETNET Data Management Workshop

Sensitivity of Change-Point Detection and Trend Estimates to GNSS IWV Time Series Properties

PhD candidate Ninh Nguyen (Univ. Paris Cité, IPGP, IGN)
Dr. Olivier Bock (IPGP, IGN)
Pr. Emilie Lebarbier (Univ. Paris Nanterre)
Dr. Annarosa Quarello (Capgemini Engineering)

9 May 2023

Context and objectives

- Study global and regional water cycle in the changing climate
  - Use observations, atmospheric/ocean reanalyses and GCMs
- Ground-based GNSS IWV observations are very accurate with low bias
  - However, even small changes in bias (inhomogeneities) are detrimental to trend and decadal variability analysis
- Inhomogeneities in GNSS data are mainly due to:
  - Equipement changes, changes in the data processing, changes in the environment.
- Reanalyses may also have inhomogeneities, mainly due to:
  - Changes in the global observing system (e.g. start/end of satellite mission)
- Segmentation/homogenization methods help to:
  - Detect and correct inhomogeneities

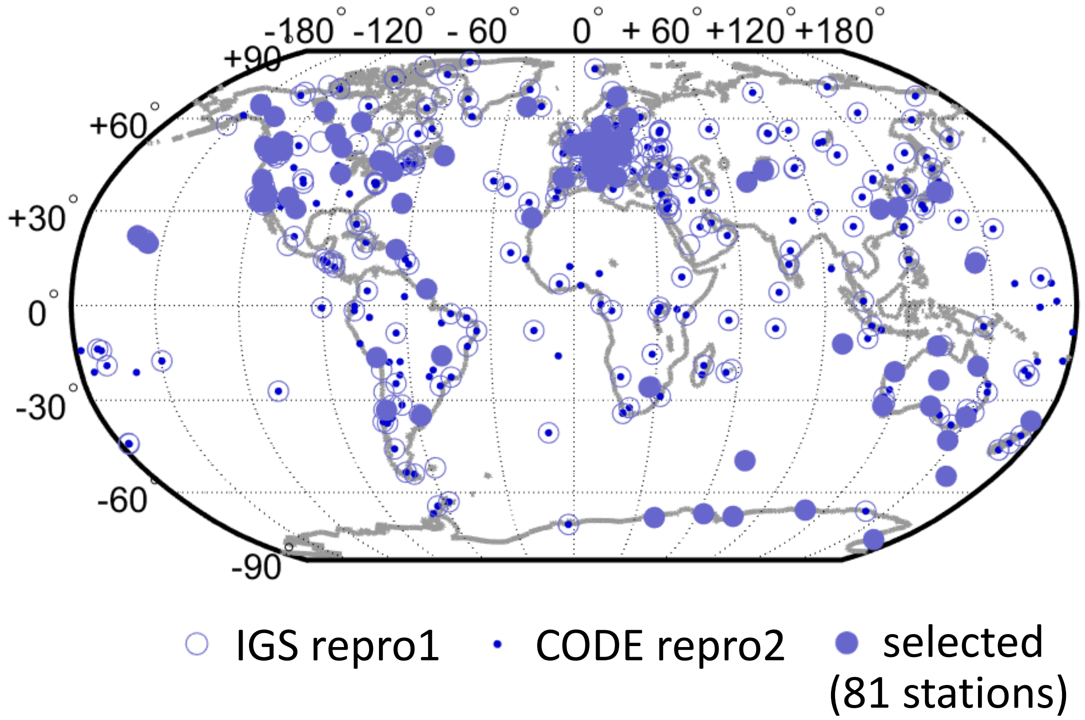
Context and objectives

- This talk discusses the sensitivity of segmentation results to:
    1. Change in GNSS data set version (repro1 vs. repro2)
    2. Length of time series: 16 years (repro1) vs. 25 years (repro2)
    3. Reference data set used in the target – reference series for the segmentation (reanalyses: ERA-Interim vs. ERA5)
    4. Auxiliary data set used for the conversion of GNSS ZTD data to IWV
- In addition we study the impact of the different segmentation results and data sets on the long-term linear trend estimates

GNSS data sets (daily IWV times series)

- IGS repro1
  - Software: GIPSY OASIS II
  - Released in 2010/2011
  - Covers period 1995-2010
- CODE (REPRO2015) repro2
  - Software: Bernese
  - Released in 2015/2019
  - Covers period 1994-2018

(*) repro2 used more recent satellite products and models => this data set should be more accurate



Analysis procedure

(1) GNSS data is ZTD (propag. delay)

- Conversion to IWV needs aux. data

  $IWV = k(T_m) \times \{ ZTD - ZHD(P_s) \}$

- We use a reanalysis for T_m and P_s

(2) Relative homogenization method works on differenced data: target – reference

- We use a reanalysis as reference

(3) Segmentation method is GNSSseg

- Here we use only the BM1 results

(4) Outlier screening

- As described in previous talk (E. Lebarbier)

(5) Validation is done wrt GNSS metadata

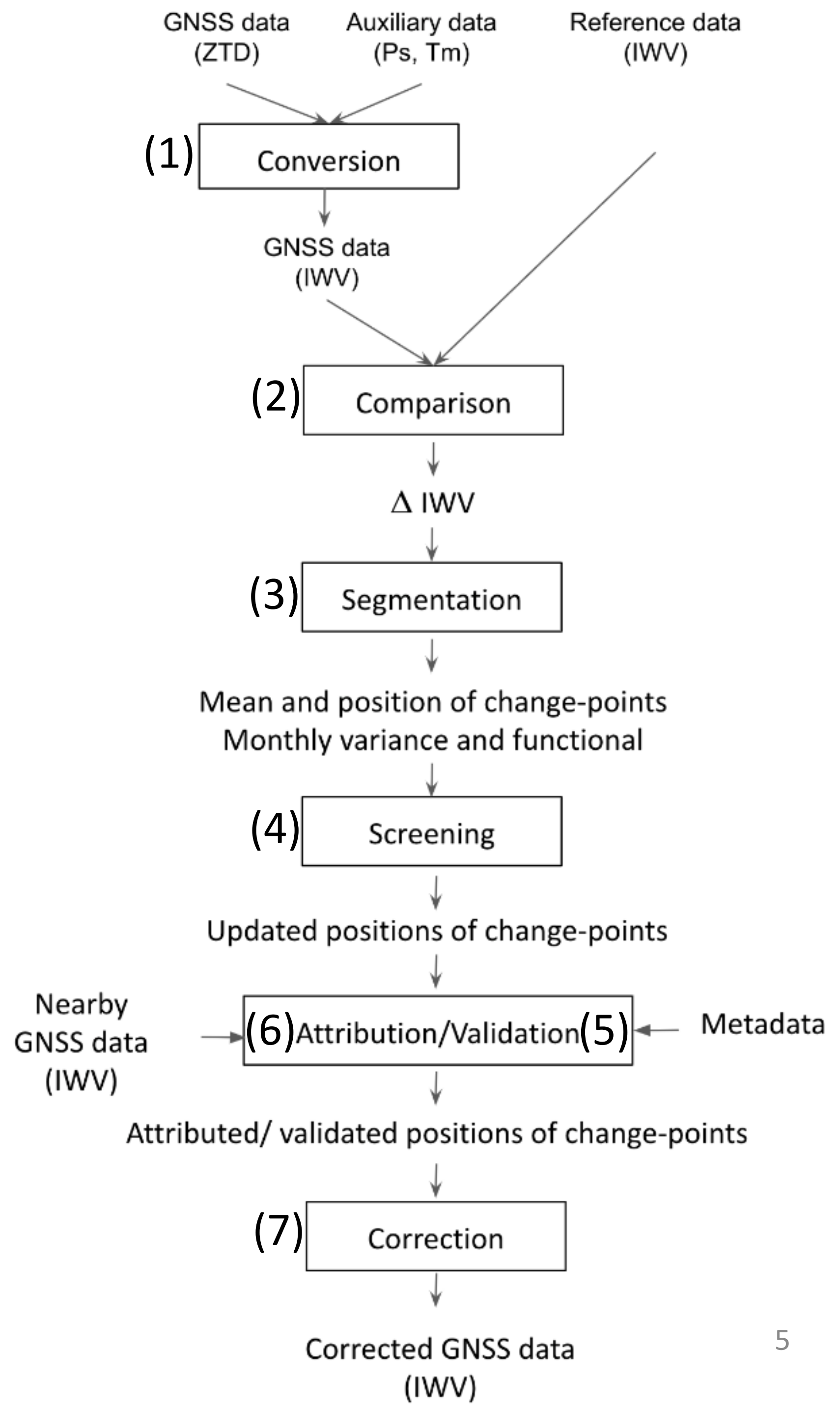
- Eq changes are well documented, no relocation issue.

(6) Attribution is not applied here.

(7) Correction : piece-wise bias correction, with the most recent segment taken as reference, and using:

i) only change-points validated by GNSS metadata,

ii) all detected change-points (assumed due to GNSS).

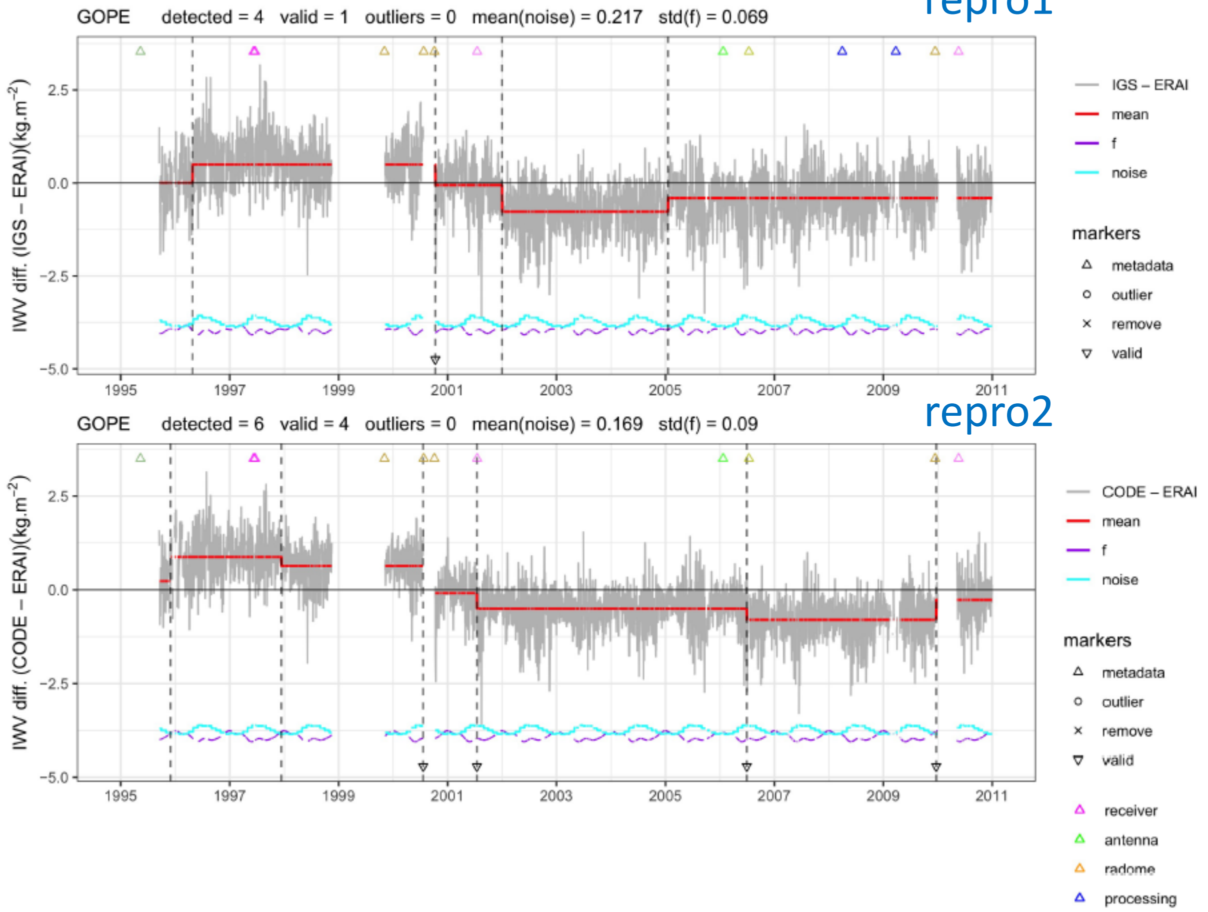


Segmentation results

	repro1		repro2
	(1) Impact of Processing		
Data Set	IGS—ERAI Time-Matched		CODE—ERAI Time-Matched
Time span	1995–2010		1995–2010
Mean of the monthly variances ($kg\ m^{-2}$)	0.68	>	0.62
Standard deviation of the functional ($kg\ m^{-2}$)	0.26	>	0.24
No. detections	231	<	257
No. outliers	36	≈	38
No. detections after screening	211	<	235
Validations after screening	63	<	68
Validations after screening (%)	29.9	≈	28.9
Similar detections	103~48.8%		

less noisy

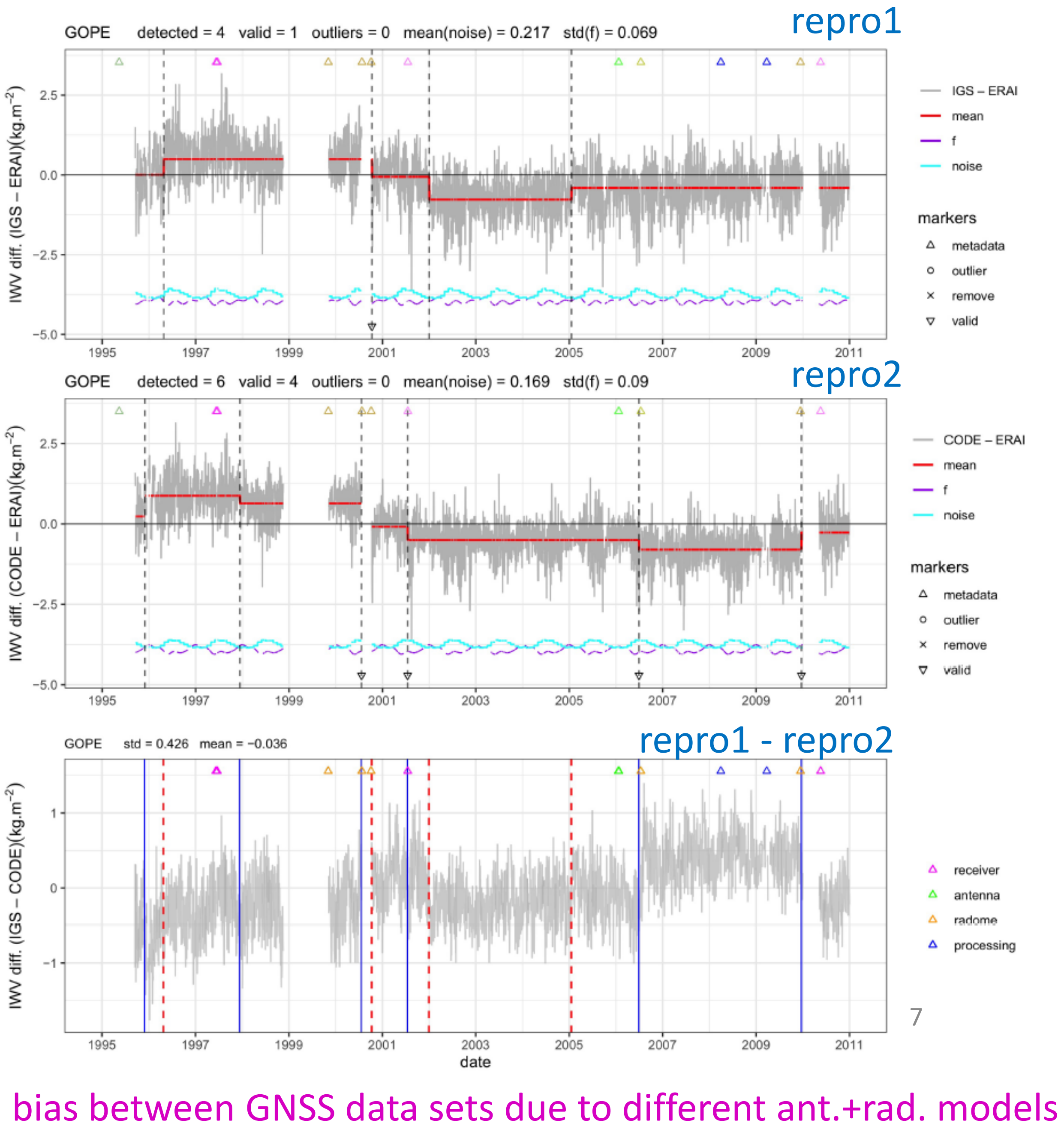
more detections



repro2 less noisy => higher detection power

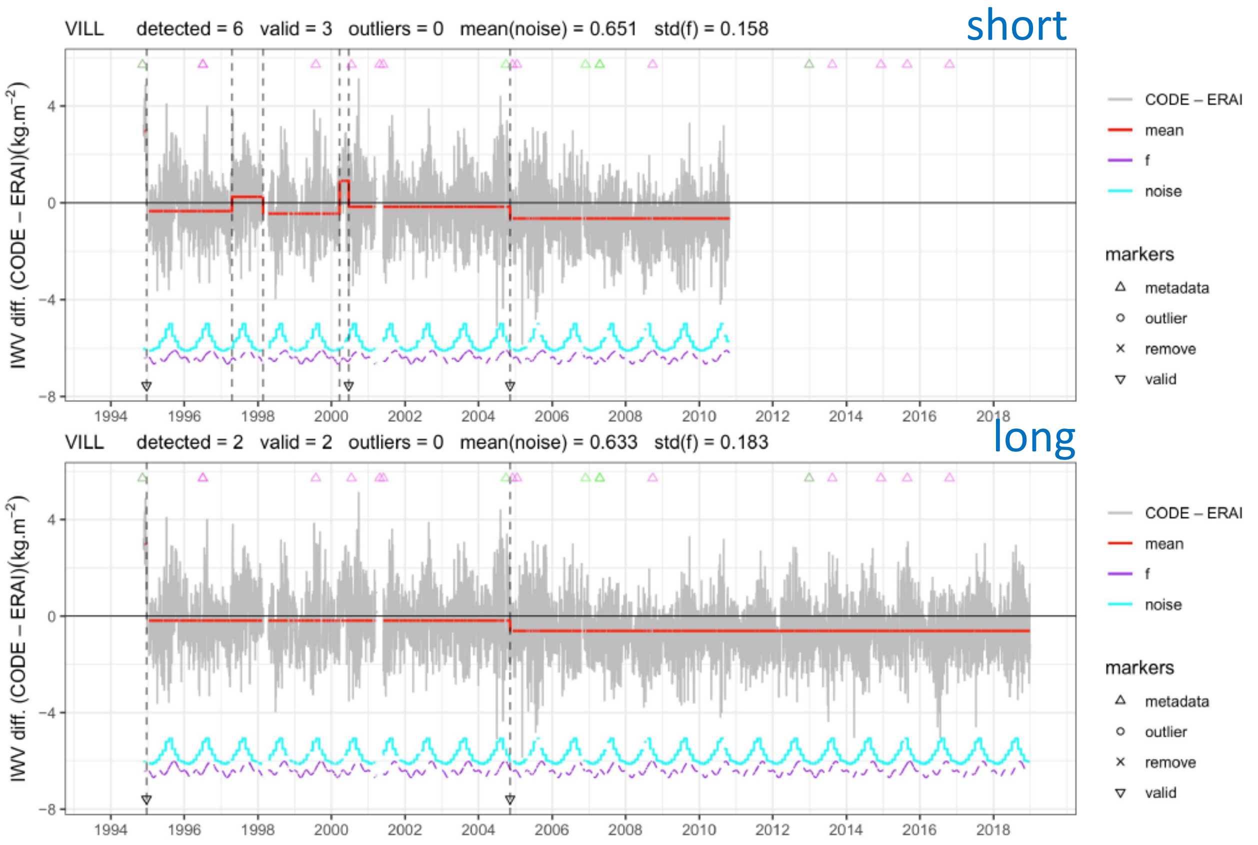
Segmentation results

	repro1		repro2
	(1) Impact of Processing		
Data Set	IGS—ERAI Time-Matched		CODE—ERAI Time-Matched
Time span	1995–2010		1995–2010
Mean of the monthly variances ($kg\ m^{-2}$)	0.68	>	0.62
Standard deviation of the functional ($kg\ m^{-2}$)	0.26	>	0.24
No. detections	231	<	257
No. outliers	36	≈	38
No. detections after screening	211	<	235
Validations after screening	63	<	68
Validations after screening (%)	29.9	≈	28.9
Similar detections	103~48.8%		



Segmentation results

	short		long
	(2) Impact of Time Length		
Data Set	CODE—ERAI Time-Limited		CODE —ERAI
Time span	1994–2010	<	1994–2018 (a)
Mean of the monthly variances ($kg\ m^{-2}$)	0.62	≈	0.63
Standard deviation of the functional ($kg\ m^{-2}$)	0.24	≈	0.23
No. detections	296	>	249 fewer detections
No. outliers	73	>	40
No. detections after screening	252	>	227
Validations after screening	77	≈	78
Validations after screening (%)	30.6	<	34.4 more validations
Similar detections	185~81.5%		



fewer detections on longer period : segmentation is conservative

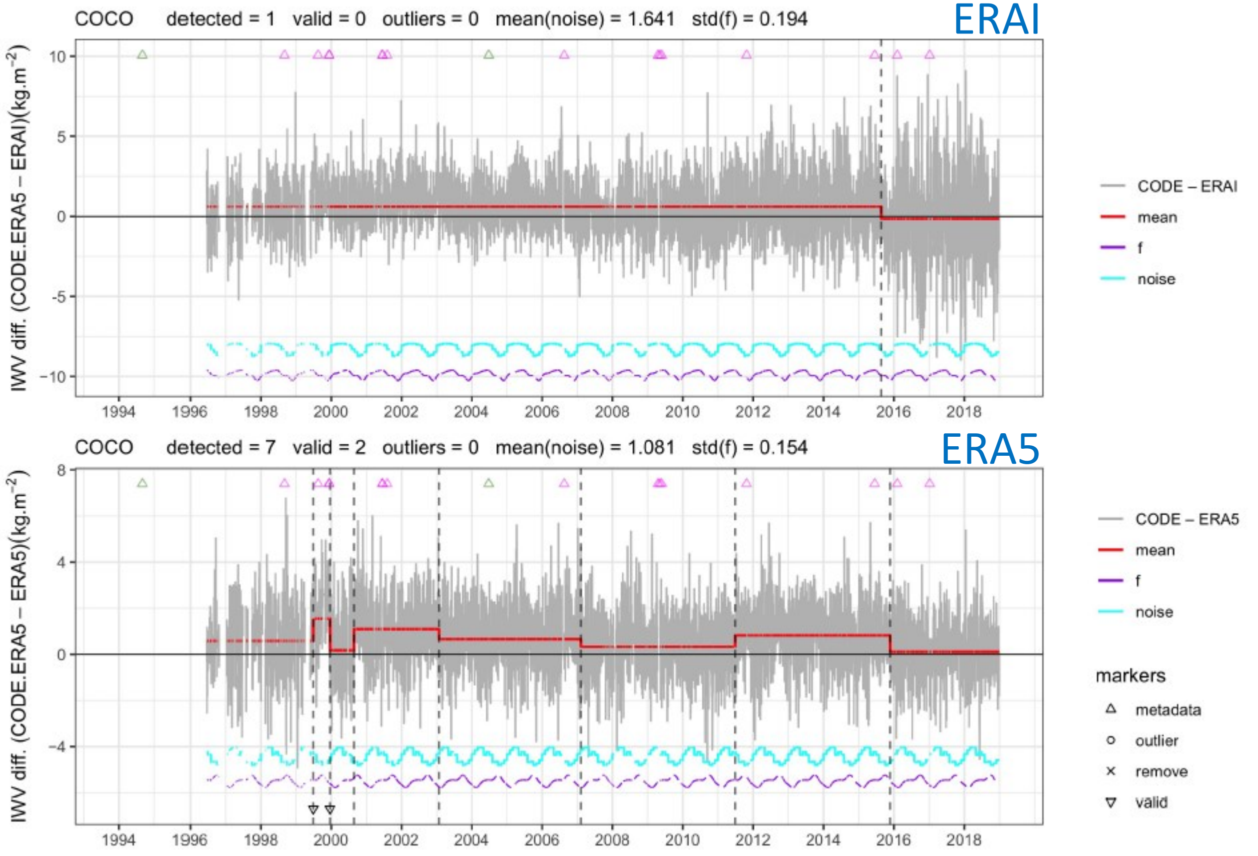
Segmentation results

Reference: ERAI ERA5

(3) Impact of Reference			
Data Set	CODE (b) —ERAI		CODE (b) —ERA5
Time span	1994–2018		1994–2018
Mean of the monthly variances ($kg\ m^{-2}$)	0.61	>	0.46
Standard deviation of the functional ($kg\ m^{-2}$)	0.23	>	0.17
No. detections	364	<	398
No. outliers	60	<	71
No. detections after screening	333	<	359
Validations after screening	114	<	131
Validations after screening (%)	34.2	≈	36.5
Similar detections	151 ~45.3%		

less noisy

more detections



ERA5 less noisy => higher detection power

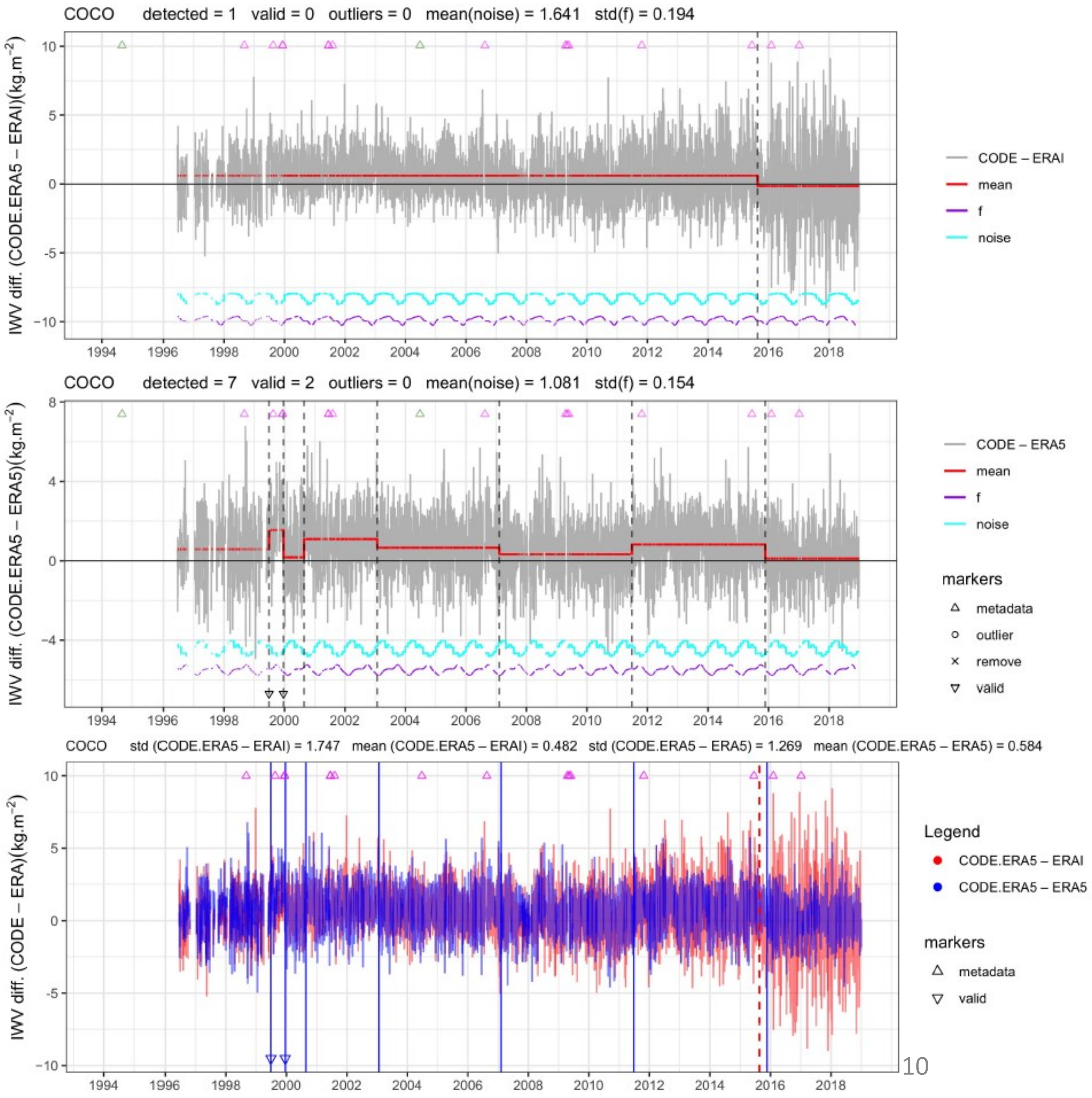
ERAI more noisy in recent years

Segmentation results

Reference: ERAI ERA5

	(3) Impact of Reference		
Data Set	CODE (b) —ERAI		CODE (b) —ERA5
Time span	1994–2018		1994–2018
Mean of the monthly variances ($kg\ m^{-2}$)	0.61	>	0.46
Standard deviation of the functional ($kg\ m^{-2}$)	0.23	>	0.17
No. detections	364	<	398
No. outliers	60	<	71
No. detections after screening	333	<	359
Validations after screening	114	<	131
Validations after screening (%)	34.2	≈	36.5
Similar detections	151 ∼45.3%		

ERA5 less noisy => higher detection power

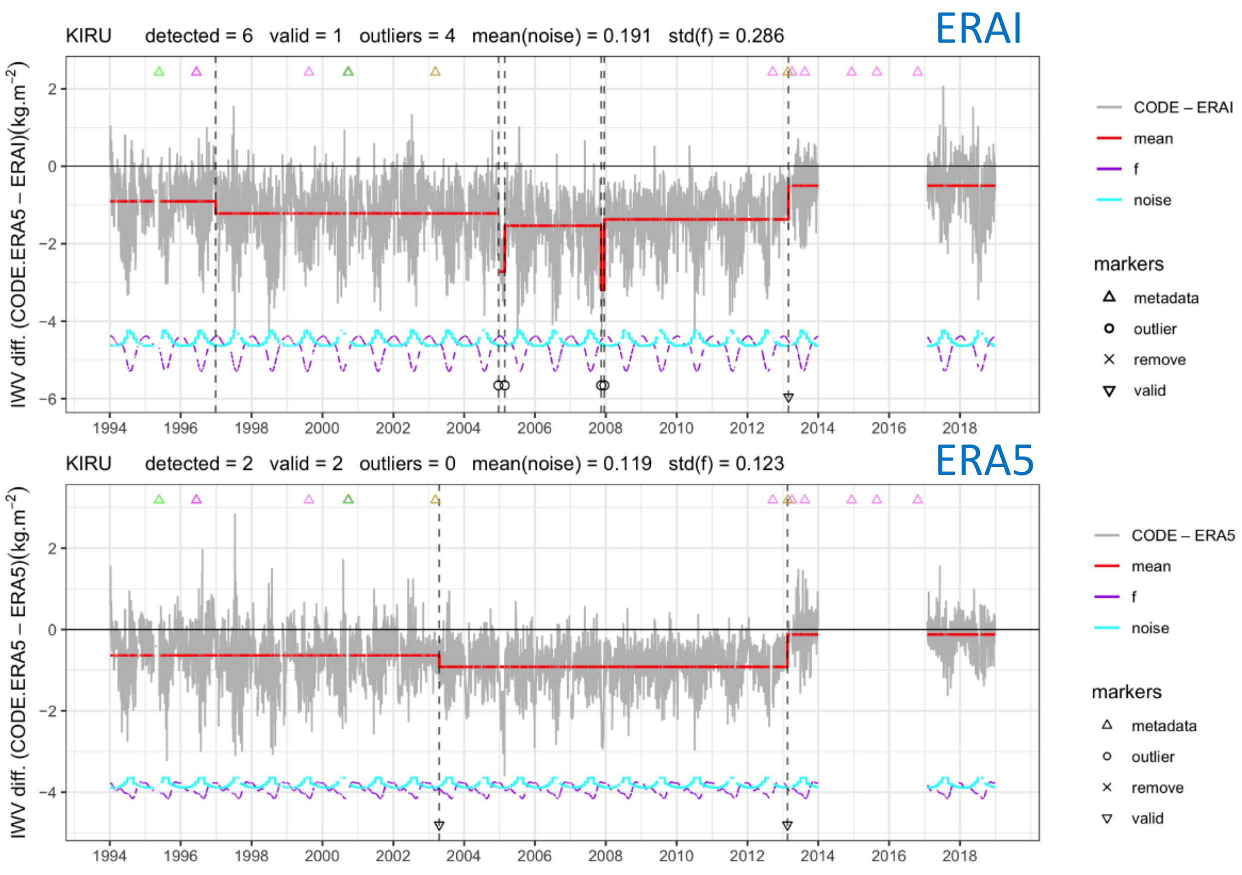


Segmentation results

Auxiliary: ERA5 ERAI

(4) Impact of Auxiliary			
Data Set	CODE aux. ERA5		CODE aux. ERAI
Time span	1994–2018		1994–2018
Mean of the monthly variances (kg m^{-2})	0.46	=	0.46
Standard deviation of the functional (kg m^{-2})	0.17	=	0.17
No. detections	398	≈	392
No. outliers	71	<	87
No. detections after screening	359	>	343
Validations after screening	131	>	125
Validations after screening (%)	36.5	≈	36.4
Similar detections	243 ~70.9%		

more outliers



ERAI has larger periodic bias (representativeness error)

Trend estimation procedure

- Linear Regression Model: $y_t = \mu + \omega x_t + s_t + \varepsilon_t$

y_t IWV time series
x_t linear trend function
μ mean IWV
ω trend slope
s_t seasonal function : 4th order Fourier Series
ε_t noise term : AR(1) $\varepsilon_t = \phi\varepsilon_{t-1} + w_t$

β coefficients of deterministic model (μ, ω, s_t)
ϕ, σ_w^2 coefficients of stochastic model (ε_t)

- Estimation method: Feasible-Generalized Least Squares (FGLS)

$y = X\beta + \varepsilon$
$\varepsilon \sim N(0, \Sigma_0)$

$$\hat{\beta}_{FGLS} = (X'\hat{\Sigma}_n^{-1}X)^{-1}X'\hat{\Sigma}_n^{-1}y$$

$$Var[\hat{\beta}_{FGLS}] = (X'\hat{\Sigma}_n^{-1}X)^{-1}$$

the coefficients of deterministic and stochastic models are estimated iteratively

=> $\hat{\omega}$, $\hat{\sigma}_\omega$ trend slope and standard error estimates

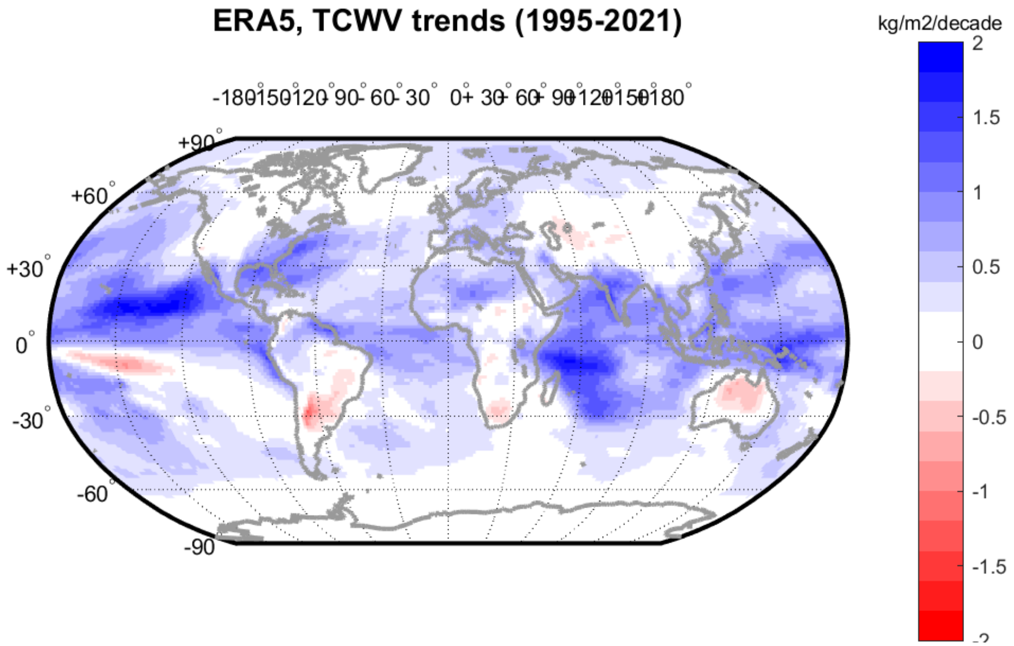
Trend results: reanalyses

Time Span		1995–2010	1994–2010	1994–2018
Std error ($kg\,m^{-2}\,year^{-1}$)		0.035	0.033	0.018
ERAI ($kg\,m^{-2}\,year^{-1}$)	mean ± std (signif)	0.018 ± 0.055 (9)	0.013 ± 0.049 (10)	0.027 ± 0.034 (37)
ERA5 ($kg\,m^{-2}\,year^{-1}$)		0.011± 0.052 (8)	0.008 ± 0.047 (8)	0.027±0.031 (35)

(81 stations)

<u>Increasing length</u>

- std. error decreases
- mean trend (positive) increased in recent years (linked to surface temperature increase ≈ 7% / 1K)
- std. trend (spatial variability) decreases
- more significant trends



Trend results: raw GNSS vs. reanalyses

Time Span		1995–2010		1994–2010	1994–2018
Std error ($kg\ m^{-2}\ year^{-1}$)		0.035		0.033	0.018
ERAI ($kg\ m^{-2}\ year^{-1}$)		0.018 ± 0.055 (9)		0.013 ± 0.049 (10)	0.027 ± 0.034 (37)
ERA5 ($kg\ m^{-2}\ year^{-1}$)		0.011 ± 0.052 (8)		0.008 ± 0.047 (8)	0.027 ± 0.031 (35)
GPS		IGS time-matched	CODE time-matched	CODE time-limited	CODE (c)
Raw data	IWV trend ($kg\ m^{-2}\ year^{-1}$)	0.024 ± 0.059 (20)	0.018 ± 0.060 (18)	0.016 ± 0.060 (23)	0.030 ± 0.031 (41)
	RMSE wrt ERA5 ($kg\ m^{-2}\ year^{-1}$)	0.044	0.046	0.046	0.033

IGS vs. CODE

- Stronger trends (positive) in IGS repro1

GNSS vs. ERA5

- mean : GNSS larger than reanalyses
- std (spatial variability): GNSS larger than reanalyses
- more significant trends
- RMSE (GNSS – ERA5): quite large

} Differences decrease with time

Trend results: GNSS homogenized (validated)

Time Span		1995–2010		1994–2010	1994–2018
Std error (kg m^{-2} year^{-1})		0.035		0.033	0.018
ERAI (kg m^{-2} year^{-1})		0.018 ± 0.055 (9)		0.013 ± 0.049 (10)	0.027 ± 0.034 (37)
ERA5 (kg m^{-2} year^{-1})		0.011± 0.052 (8)		0.008 ± 0.047 (8)	0.027±0.031 (35)
GPS		IGS time-matched	CODE time-matched	CODE time-limited	CODE (c)
Raw data	IWV trend (kg m^{-2} year^{-1})	0.024 ± 0.059 (20)	0.018 ± 0.060 (18)	0.016 ± 0.060 (23)	0.030±0.031 (41)
	RMSE wrt ERA5 (kg m^{-2} year^{-1})	0.044	0.046	0.046	0.033
corrected IWV by validations	IWV trend (kg m^{-2} year^{-1})	0.015± 0.052 (12)	0.014± 0.052 (11)	0.011± 0.052 (15)	0.027 ± 0.026 (34)
	RMSE wrt ERA5 (kg m^{-2} year^{-1})	0.038	0.039	0.040	0.019

<u>GNSS raw vs. homogenized</u>

- mean decreases
- Std (spatial variability) decreases
- Nb significant trends decreases
- RMSE (GNSS – ERA5): decreass

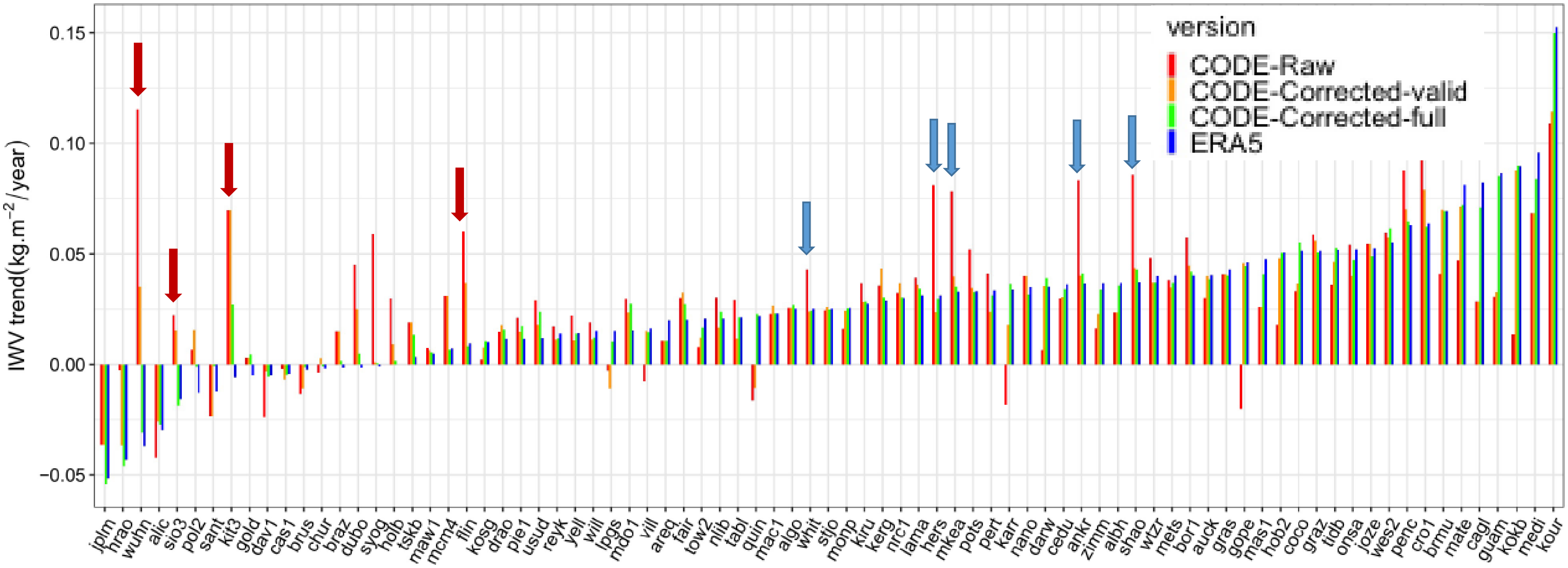
Homogenized GNSS trends are more similar to reanalyses

Trend results: GNSS homogenized (all)

Time Span		1995–2010		1994–2010	1994–2018
Std error ($kg\ m^{-2}\ year^{-1}$)		0.035		0.033	0.018
ERAI ($kg\ m^{-2}\ year^{-1}$)		0.018 ± 0.055 (9)		0.013 ± 0.049 (10)	0.027 ± 0.034 (37)
ERA5 ($kg\ m^{-2}\ year^{-1}$)		0.011± 0.052 (8)		0.008 ± 0.047 (8)	0.027±0.031 (35)
GPS		IGS time-matched	CODE time-matched	CODE time-limited	CODE (c)
Raw data	IWV trend ($kg\ m^{-2}\ year^{-1}$)	0.024 ± 0.059 (20)	0.018 ± 0.060 (18)	0.016 ± 0.060 (23)	0.030±0.031 (41)
	RMSE wrt ERA5 ($kg\ m^{-2}\ year^{-1}$)	0.044	0.046	0.046	0.033
corrected IWV by validations	IWV trend ($kg\ m^{-2}\ year^{-1}$)	0.015± 0.052 (12)	0.014± 0.052 (11)	0.011± 0.052 (15)	0.027 ± 0.026 (34)
	RMSE wrt ERA5 ($kg\ m^{-2}\ year^{-1}$)	0.038	0.039	0.040	0.019
corrected IWV by all breakpoints	IWV trend ($kg\ m^{-2}\ year^{-1}$)	0.017 ± 0.053 (9)	0.016 ± 0.054 (9)	0.012 ± 0.048 (13)	0.027 ± 0.030 (34)
	RMSE wrt ERA5 ($kg\ m^{-2}\ year^{-1}$)	0.021	0.022	0.022	0.006

GNSS homogenized with all change-points gets very close to ERA5 (this is expected)

Trend results: at 81 stations



- In many cases both corrected trends agree with each other and are more consistent with reanalysis
- In some cases the corrected GNSS trends are different and the valided trends don’t agree with reanalysis

Conclusions

- More recent GNSS data set and reanalysis are less noisy
  - segmentation has more detection power
    - helps detect biases in GNSS antenna+radome models => next check GNSS repro3
  - only small impact on trend estimates
- Segmentation is more conservative on longer period
  - fewer detections (only the most signigicant offsets are dectected)
- Trend estimates are more precise on longer period
  - more trends are significant and spatial variability decreases
    - trends increased in recent years (atmosphere gets warmer and moister)
- Trends from homogenized and validated GNSS data
  - more similar to ERA5 on average
  - evidence that some changepoints are undocumented or due to reanalysis
    - need a more effective change-point validation strategy => attribution method

Reference

- Nguyen, K.N.; Quarello, A.; Bock, O., Lebarbier, E. (2021) Sensitivity of Change-Point Detection and Trend Estimates to GNSS IWV Time Series Properties. Atmosphere, 12, 1102. https://doi.org/10.3390/atmos12091102