



*Supplement of*

**Assessment of pluri-annual and decadal changes in terrestrial water storage predicted by global hydrological models in comparison with the GRACE satellite gravity mission**

**Julia Pfeffer et al.**

*Correspondence to:* Julia Pfeffer ([julia.pfeffer@magellium.fr](mailto:julia.pfeffer@magellium.fr))

The copyright of individual parts of the supplement might differ from the article licence.

## **Table of contents:**

### **Supplementary S1: Comparison of TWS anomalies from satellite gravimetry and land surface models 5**

**Figure S1.1** Comparison of TWS anomalies estimated from an ensemble of nine GRACE solutions, one land surface model (NOAH) and two global hydrological models (ISBA-CTRIP and WGHM). a) Residual TWS anomalies calculated as the difference between GRACE and NOAH. The amplitude of the residual TWS anomalies is expressed as the range at 95% CL, calculated as the difference between the 97.5 and 2.5 percentiles of the TWS anomalies obtained in each grid cell over the entire study period. b) Determination coefficients between the average GRACE solution and NOAH land surface model. Negative coefficients appear in white. c) Model with smallest TWS residuals. d) Model with smallest determination coefficient. e) Model with largest TWS residuals. f) Model with largest determination coefficient. **6**

**Figure S1.2** Comparison of TWS and precipitation anomalies averaged over the Central Amazon Corridor (box A in Fig. B1 - Appendix B). a) Average precipitation anomalies for the GPCC (gauge-based) and IMERG (satellite-based) products. b) Power Spectral Density (PSD) of average precipitation anomalies. c) TWS anomalies average over the central Amazon for two global hydrological models (ISBA-CTRIP in blue and WGHM in black), one land surface model (NOAH in green) and 9 GRACE solutions (mascons in red, spherical harmonic in magenta). The solid line corresponds to the average of the sub-ensemble, the shaded area to the minimum to maximum envelope. d) PSD of the averaged TWS anomalies shown in (c). e) Residual TWS anomalies averaged over the central Amazon corridor and calculated as the difference between GRACE and ISBA-CTRIP (blue when the difference is calculated with mascons, cyan with spherical harmonics), WGHM (black when the difference is calculated with mascons, grey with spherical harmonics) or NOAH (green when the difference is calculated with mascons, yellow with spherical harmonics). **8**

**Fig. S1.3** Same as Fig. S1.2 but for the Upper Sao Francisco (box B in Fig. B1 - Appendix B). **9**

**Fig. S1.4** Same as Fig. S1.2 but for the Zambezi (box C in Fig. B1 - Appendix B). **9**

**Fig. S1.5** Same as Fig. S1.2 but for the Congo (box D in Fig. B1 - Appendix B). **10**

**Fig. S1.6** Same as Fig. S1.2 but for North Australia (box E in Fig. B1 - Appendix B). **10**

**Fig. S1.7** Same as Fig. S1.2 but for the central US (box F in Fig. B1 - Appendix B). **11**

**Fig. S1.8** Same as Fig. S1.2 but for North India (box G in Fig. B1 - Appendix B). **11**

**Fig. S1.9** Same as Fig. S1.2 but for the North of the Black Sea (box H in Fig. B1 - Appendix B). **12**

### **Supplementary S2 : Comparison of global hydrological models with a GRACE-based mascon ensemble and a GRACE-based spherical harmonic ensemble 13**

**Figure S2.1:** Comparison of TWS anomalies estimated from an ensemble of three GRACE mascon solutions and two global hydrological models. The amplitude of the non-seasonal TWS variability is expressed as the range at 95% CL, calculated as the difference between the 97.5 and 2.5 percentiles of the TWS anomalies obtained in each grid cell over the entire study period. a) Range of TWS anomalies estimated as the average of three GRACE mascons solutions. b) Dispersion of the range of TWS anomalies among three GRACE mascons solutions. Range of TWS anomalies estimated with ISBA-CTRIP (c) and WGHM (d). Range of residual TWS anomalies estimated as the difference between the average of three GRACE mascon solutions and ISBA-CTRIP (e) or WGHM (f). **14**

**Figure S2.2:** Comparison of TWS anomalies estimated from an ensemble of six GRACE spherical harmonic solutions and two global hydrological models. The amplitude of the non-seasonal TWS variability is expressed as the range at 95% CL, calculated as the difference between the 97.5 and 2.5 percentiles of the TWS anomalies obtained in each grid cell over the entire study period. a) Range of TWS anomalies estimated as the average of six GRACE spherical harmonics solutions. b) Dispersion of the range of TWS anomalies among six GRACE spherical harmonics solutions. Range of TWS anomalies estimated with ISBA-CTRIP (c) and WGHM (d). Range of residual TWS anomalies estimated as the difference between the average of six GRACE spherical harmonic solutions and ISBA-CTRIP (e) or WGHM (f). **15**

**Figure S2.3:** Range ratios between the average of three GRACE mascon solutions and the hydrological models ISBA-CTRIP (a) and WGHM (b). Determination coefficients between the average GRACE mascon solution and the hydrological models ISBA-CTRIP (c) and WGHM (d). Regions, where the coefficient of determination is negative, are shown in white **16**

**Figure S2.4:** Range ratios between the average of six GRACE spherical harmonic solutions and the hydrological models ISBA-CTRIP (a) and WGHM (b). Determination coefficients between the average GRACE spherical harmonic solution and the hydrological models ISBA-CTRIP (c) and WGHM (d). Regions, where the coefficient of determination is negative, are shown in white **16**

**Figure S2.5:** Characteristic time scales in residual TWS anomalies calculated as the differences between the average of three GRACE mascon solutions and ISBA-CTRIP (a) or WGHM (b). Subannual, pluriannual and decadal contributions have been computed with high-pass (cut-off period at 1.5 years), band-pass (cut-off periods at 1.5 and 10 years) and low-pass (cut-off period at 10 years) filters respectively. The percentage of variance explained by one contribution has been calculated as the coefficient of determination with respect to the full residual signal. **17**

**Figure S2.6:** Characteristic time scales in residual TWS anomalies calculated as the differences between the average of six GRACE spherical harmonic solutions and ISBA-CTRIP (a) or WGHM (b). Subannual, pluriannual and decadal contributions have been computed with high-pass (cut-off period at 1.5 years), band-pass (cut-off periods at 1.5 and 10 years) and low-pass (cut-off period at 10 years) filters respectively. The percentage of variance explained by one contribution has been calculated as the coefficient of determination with respect to the full residual signal. **18**

**Figure S2.7:** a) Linear trends in residual TWS anomalies calculated as the difference between the average of three GRACE mascon solutions and ISBA-CTRIP. b) Same as (a) with WGHM. c) Amplitude of non-linear signals in residual TWS anomalies calculated as the difference between the average of three GRACE mascon solutions and ISBA-CTRIP. The amplitude is calculated as the difference between the 97.5 and 2.5 percentiles. d) Same as (c) with WGHM. e) Coefficient of determination calculated for non-linear signals with respect to TWS anomalies calculated as the difference between the average GRACE mascon solution and ISBA-CTRIP. f) Same as (e) with WGHM. **19**

**Figure S2.8:** a) Linear trends in residual TWS anomalies calculated as the difference between the average of six GRACE spherical harmonic solutions and ISBA-CTRIP. b) Same as (a) with WGHM. c) Amplitude of non-linear signals in residual TWS anomalies calculated as the difference between the average of six GRACE spherical harmonic solutions and ISBA-CTRIP. The amplitude is calculated as the difference between the 97.5 and 2.5 percentiles. d) Same as (c) with WGHM. e) Coefficient of determination calculated for non-linear signals with respect to TWS anomalies calculated as the difference between the average GRACE mascon solution and ISBA-CTRIP. f) Same as (e) with WGHM. **20**

**Supplementary S3 Comparison of TWS anomalies from GRACE and global hydrological models over large river basins **21****

**Figure S3.1:** Map of the 40 largest river basins considered in this study: 1) Dnieper, 2) Brahmaputra, 3) Sao Francisco, 4) Kolyma, 5) Colorado, 6) Columbia, 7) Rio Grande, 8) Okavango, 9) Tocantins, 10) Mekong, 11) Danube, 12) Jubba, 13) Yukon, 14) Indus, 15) Shatt Al Arab, 16) Orinoco, 17) Yellow River, 18) Orange, 19) Ganges, 20) Saint Lawrence, 21) Murray, 22) Nelson, 23) Lake Eyre, 24) Zambezi, 25) Volga, 26) Tarim He, 27) Aral Sea, 28) Yangtze, 29) Mackenzie, 30) Niger, 31) Amur, 32) Lena, 33) Chad, 34) Yenisei, 35) Parana, 36) Ob, 37) Mississippi, 38) Nile, 39) Congo, 40) Amazon. **22**

**Figure S3.2:** Comparison of TWS and precipitation anomalies averaged over Amazon basin. a) Average precipitation anomalies for the GPCC (gauge-based) and IMERG (satellite-based) products. b) Power Spectral Density (PSD) of average precipitation anomalies. c) TWS anomalies average over the central Amazon for two global hydrological models (ISBA-CTRIP in blue and WGHM in black) and 9 GRACE solutions (mascons in red, spherical harmonic in magenta). The solid line corresponds to the average of the sub-ensemble, the shaded area to the minimum to maximum envelope. d) PSD of the averaged TWS anomalies shown in (c). e) Residual TWS anomalies averaged over the central Amazon corridor and calculated as the difference between GRACE and ISBA-CTRIP (blue when the difference is calculated with mascons, cyan with spherical harmonics) or WGHM (black when the difference is calculated with mascons, grey with spherical harmonics). **22**

**Figure S3.3:** Same as S3.2 for the Amur Basin. Non-seasonal precipitation anomalies are only estimated with GPCC, as a significant part of the basin is not covered by IMERG satellites due to the high latitude of the Amur basin. **23**

**Figure S3.4:** Same as S3.2 for the Aral Sea basin. **23**

**Figure S3.5:** Same as S3.2 for the Brahmaputra basin. **24**

**Figure S3.6:** Same as S3.2 for the Chad basin. **24**

**Figure S3.7:** Same as S3.2 for the Colorado basin. **25**

**Figure S3.8:** Same as S3.2 for the Columbia basin. **25**

**Figure S3.9:** Same as S3.2 for the Congo basin. **26**

**Figure S3.10:** Same as S3.2 for the Danube basin. **26**

**Figure S3.11:** Same as S3.2 for the Dnieper basin. **27**

**Figure S3.12:** Same as S3.2 for the Ganges basin. **27**

<b>Figure S3.13:</b> Same as S3.2 for the Indus basin.	<b>28</b>
<b>Figure S3.14:</b> Same as S3.2 for the Jubba basin.	<b>28</b>
<b>Figure S3.15:</b> Same as S3.2 for the Kolyma basin. Non-seasonal precipitation anomalies are only estimated with GPCC, as a significant part of the river basin is not covered by IMERG satellites due to its high latitude.	<b>29</b>
<b>Figure S3.16:</b> Same as S3.2 for the Lake Eyre basin.	<b>29</b>
<b>Figure S3.17:</b> Same as S3.2 for the Lena basin. Non-seasonal precipitation anomalies are only estimated with GPCC, as a significant part of the river basin is not covered by IMERG satellites due to its high latitude.	<b>30</b>
<b>Figure S3.18:</b> Same as S3.2 for the Mackenzie basin. Non-seasonal precipitation anomalies are only estimated with GPCC, as a significant part of the river basin is not covered by IMERG satellites due to its high latitude.	<b>30</b>
<b>Figure S3.19:</b> Same as S3.2 for the Mekong basin.	<b>31</b>
<b>Figure S3.20:</b> Same as S3.2 for the Mississippi basin.	<b>32</b>
<b>Figure S3.21:</b> Same as S3.2 for the Murray basin.	<b>32</b>
<b>Figure S3.22:</b> Same as S3.2 for the Nelson basin.	<b>33</b>
<b>Figure S3.23:</b> Same as S3.2 for the Niger basin.	<b>33</b>
<b>Figure S3.24:</b> Same as S3.2 for the Nile basin.	<b>34</b>
<b>Figure S3.25:</b> Same as S3.2 for the Ob basin. Non-seasonal precipitation anomalies are only estimated with GPCC, as a significant part of the river basin is not covered by IMERG satellites due to its high latitude.	<b>34</b>
<b>Figure S3.26:</b> Same as S3.2 for the Okavango basin.	<b>35</b>
<b>Figure S3.27:</b> Same as S3.2 for the Orange basin.	<b>35</b>
<b>Figure S3.28:</b> Same as S3.2 for the Orinoco basin.	<b>36</b>
<b>Figure S3.29:</b> Same as S3.2 for the Parana basin.	<b>36</b>
<b>Figure S3.30:</b> Same as S3.2 for the Rio Grande basin.	<b>37</b>
<b>Figure S3.31:</b> Same as S3.2 for the Saint Lawrence basin.	<b>37</b>
<b>Figure S3.32:</b> Same as S3.2 for the Sao Francisco basin.	<b>38</b>
<b>Figure S3.33:</b> Same as S3.2 for the Shatt al Arab basin.	<b>38</b>
<b>Figure S3.34:</b> Same as S3.2 for the Tarim He basin.	<b>39</b>
<b>Figure S3.35:</b> Same as S3.2 for the Tocantins basin.	<b>39</b>
<b>Figure S3.36:</b> Same as S3.2 for the Volga basin.	<b>40</b>
<b>Figure S3.37:</b> Same as S3.2 for the Yangtze basin.	<b>40</b>
<b>Figure S3.38:</b> Same as S3.2 for the Yellow River basin.	<b>41</b>
<b>Figure S3.39:</b> Same as S3.2 for the Yenisei basin. Non-seasonal precipitation anomalies are only estimated with GPCC, as a significant part of the river basin is not covered by IMERG satellites due to its high latitude.	<b>41</b>
<b>Figure S3.40:</b> Same as S3.2 for the Yukon basin. Non-seasonal precipitation anomalies are only estimated with GPCC, as a significant part of the river basin is not covered by IMERG satellites due to its high latitude.	<b>42</b>
<b>Figure S3.41:</b> Same as S3.2 for the Zambezi basin	<b>42</b>
<b>Supplementary S4 Comparison of TWS anomalies from GRACE and global hydrological models over Southern India</b>	<b>43</b>
<b>Figure S4.1</b> Comparison of TWS and precipitation anomalies averaged across the Indian Peninsular Plateau (latitudes 7°-23°N; longitudes 70-80°E). a) Average precipitation anomalies for the GPCC (gauge-based) and IMERG (satellite-based) products. b) Power Spectral Density (PSD) of average precipitation anomalies. c) TWS anomalies average over the central Amazon for two global hydrological models (ISBA-CTRIP in blue and WGHM in black) and 9 GRACE solutions (mascons in red, spherical harmonic in magenta). The solid line corresponds to the average of the sub-ensemble, the shaded area to the minimum to maximum envelope. d) PSD of the averaged TWS anomalies shown in (c). e) Residual TWS anomalies averaged over the central Amazon corridor and calculated as the difference between GRACE and ISBA-CTRIP (blue when the difference is calculated with mascons, cyan with spherical harmonics) or WGHM (black when the difference is calculated with mascons, grey with spherical harmonics).	<b>43</b>

## **Supplementary S1: Comparison of TWS anomalies from satellite gravimetry and land surface models**

In addition to ISBA-CTRIP and WGHM, the GRACE-based TWS anomalies are compared to the Global Land Data Assimilation System (GLDAS) NOAH v3.3 land surface model (e.g. Rodell et al., 2004; Landerer et al., 2012; Rodell et al., 2020).

GLDAS is a global land modelling system, uncoupled to the atmosphere, integrating in-situ and satellite observations to constrain meteorological and surface fluxes (e.g. precipitation, longwave and shortwave radiation) and states (e.g. temperature, pressure, vegetation, etc.). NOAH is a one-dimensional land surface model, simulating vertical water and energy fluxes at the surface and in the two first metres of soil. NOAH V3.3 uses the meteorological forcing fields from GLDAS 2.1, including the Global Data Assimilation System (GDAS) atmospheric analysis fields (Derber et al., 1991), the disaggregated Global Precipitation Climatology Project (GPCP) precipitation fields (Adler et al., 2003), and the Air Force Weather Agency's AGRicultural METeorological modelling system (AGRMET) radiation fields. Monthly total water storage anomalies are calculated as the sum of soil moisture, snow and canopy water sampled the same days as GRACE and GRACE-FO over a regular  $1^\circ \times 1^\circ$  grid. To be consistent with WGHM and ISBA-CTRIP, the same corrections have been applied. Lake water storage anomalies have been added to the TWS anomalies predicted with NOAH. Seasonal signals have been removed by least square adjustment of a semi-annual and annual sinusoid. Regions (i.e. ice sheets, arctic islands and regions impacted by Sumatra, Tohoku and Maule earthquakes) dominated by non-hydrological processes have been masked. A diffusive filter with an isotropic Daley length of 250 km has been applied.

The Fig. S1.1 shows the residual TWS anomalies, calculated as the difference between GRACE and NOAH based TWS anomalies. To be consistent with ISBA-CTRIP and WGHM, the residual TWS anomalies have been calculated over the same period of time, i.e. April 2002 - December 2016. Larger residual TWS anomalies are obtained with NOAH than ISBA-CTRIP or WGHM over most continental areas (Fig. S1.1), except in central North America, around the Caspian Sea and in central Australia (Fig. S1.1). A large part of the variance in GRACE-based TWS anomalies cannot be explained by NOAH simulations, leading to large areas with negative determination coefficients (see areas where  $R^2 < 0$  in Fig. S1.1, i.e. the variance of TWS residuals is larger than GRACE-based TWS observations).

Over most of continental areas, NOAH is less performant than ISBA-CTRIP and WGHM when compared to GRACE. This is expressed by larger TWS residuals and smaller  $R^2$  values. The representation of hydrological processes is indeed much scarcer in NOAH than in ISBA-CTRIP or WGHM. For example, NOAH does not take into account lateral fluxes. Surface water storage and groundwater storage are not represented in the land surface model. Finally, similarly to ISBA-CTRIP and in opposition to WGHM, NOAH does not take into account anthropogenic influences on terrestrial water storage changes.

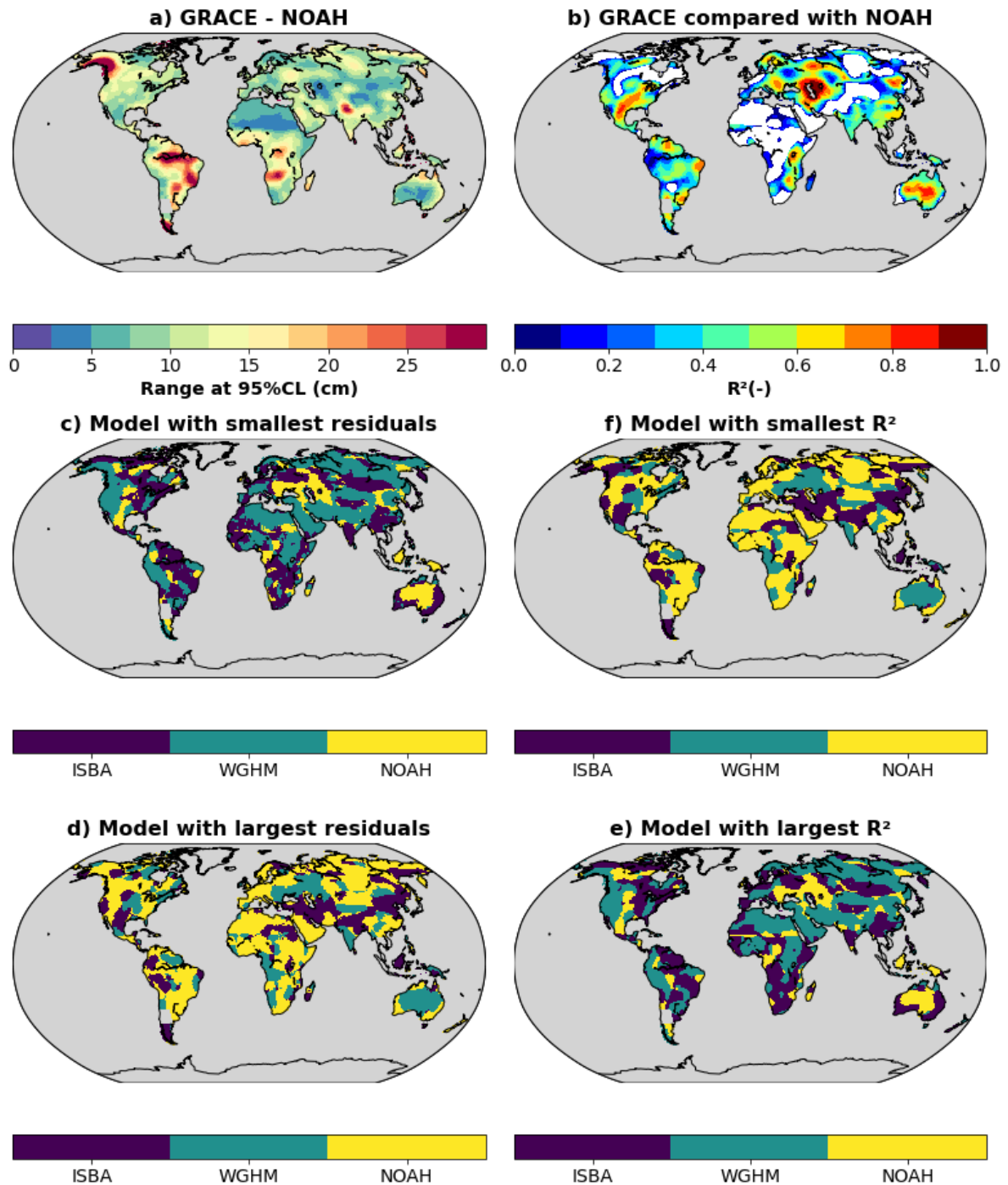


Figure S1.1 Comparison of TWS anomalies estimated from an ensemble of nine GRACE solutions, one land surface model (NOAH) and two global hydrological models (ISBA-CTRIP and WGHM). a) Residual TWS anomalies calculated as the difference between GRACE and NOAH. The amplitude of the residual TWS anomalies is expressed as the range at 95% CL, calculated as the difference between the 97.5 and 2.5 percentiles of the TWS anomalies obtained in each grid cell over the entire study period. b) Determination coefficients between the average GRACE solution and NOAH land surface model. Negative coefficients appear in white. c) Model with smallest TWS residuals. d) Model with smallest determination coefficient. e) Model with largest TWS residuals. f) Model with largest determination coefficient.

Regionally-averaged TWS anomalies estimated with GRACE, NOAH, ISBA-CTRIP and WGHM have been compared over the Amazon (Fig S1.2), Sao Francisco (Fig. S1.3), Zambezi (Fig. S1.4), Congo (Fig. S1.5), North Australia (Fig. S1.6), central US (Fig. S1.7), North India (S1.8) and Northern Black Sea (S1.9). For all of these regions, pluri-annual to decadal TWS changes are underestimated when comparing NOAH simulations to GRACE and GRACE-FO observations. Over the Amazon (Fig S1.2), Sao Francisco (Fig. S1.3), Zambezi (Fig. S1.4), Congo (Fig. S1.5) and North Australia (Fig. S1.6) regions, the underestimation of pluri-annual to decadal TWS changes is more pronounced for NOAH than for ISBA-CTRIP and WGHM, leading to larger residuals in TWS changes. Over the central US and North of the Black Sea, the underestimation of pluri-annual to decadal TWS changes is less severe for NOAH than for ISBA-CTRIP or WGHM. Over the North of India, NOAH and ISBA-CTRIP simulations fail to capture the decreasing trend in TWS, partly captured by WGHM.

Over the central US, non-seasonal TWS changes predicted with NOAH appear strongly correlated with GRACE ( $R=0.73$ ). The amplitude of TWS changes is slightly weaker for NOAH ( $\pm 5$  cm EWH) than GRACE ( $\pm 7.5$  cm EWH). The sharp decrease in TWS during the 2012 and 2020 droughts is well captured by NOAH (better than ISBA-CTRIP or WGHM), although the model does not take into account groundwater abstractions for irrigation. The decline in TWS occurs 6 months earlier in NOAH than in GRACE and GRACE-FO for both droughts. The good performance of the NOAH model is puzzling in this region undergoing significant agricultural water management, which is not taken into account in the model. At the North-East of the Black sea catchment (Fig S1.9), NOAH captures part of the interannual variability in TWS seen by GRACE, but does not capture the decadal trend.

Over most continental areas, NOAH provides TWS estimates that are less accurate than ISBA-CTRIP or WGHM when compared to GRACE. The pluri-annual and decadal changes in TWS are particularly underestimated in the Southern hemisphere (i.e South America, Africa, North Australia). Better performances are reached in the central US, Caspian Sea catchment and central Australia. The difficulty to accurately simulate pluri-annual to decadal changes in TWS may originate from overly simplistic assumptions in the land surface model (vertical fluxes only, no surface water compartment, no groundwater compartment, no anthropogenic influences). The NOAH land surface model is however the only model considered in this study providing TWS estimates in near real time. The availability of TWS anomalies in time is a limiting factor for many applications, making land surface models indispensable tools in spite of a lower accuracy in many regions of the world.

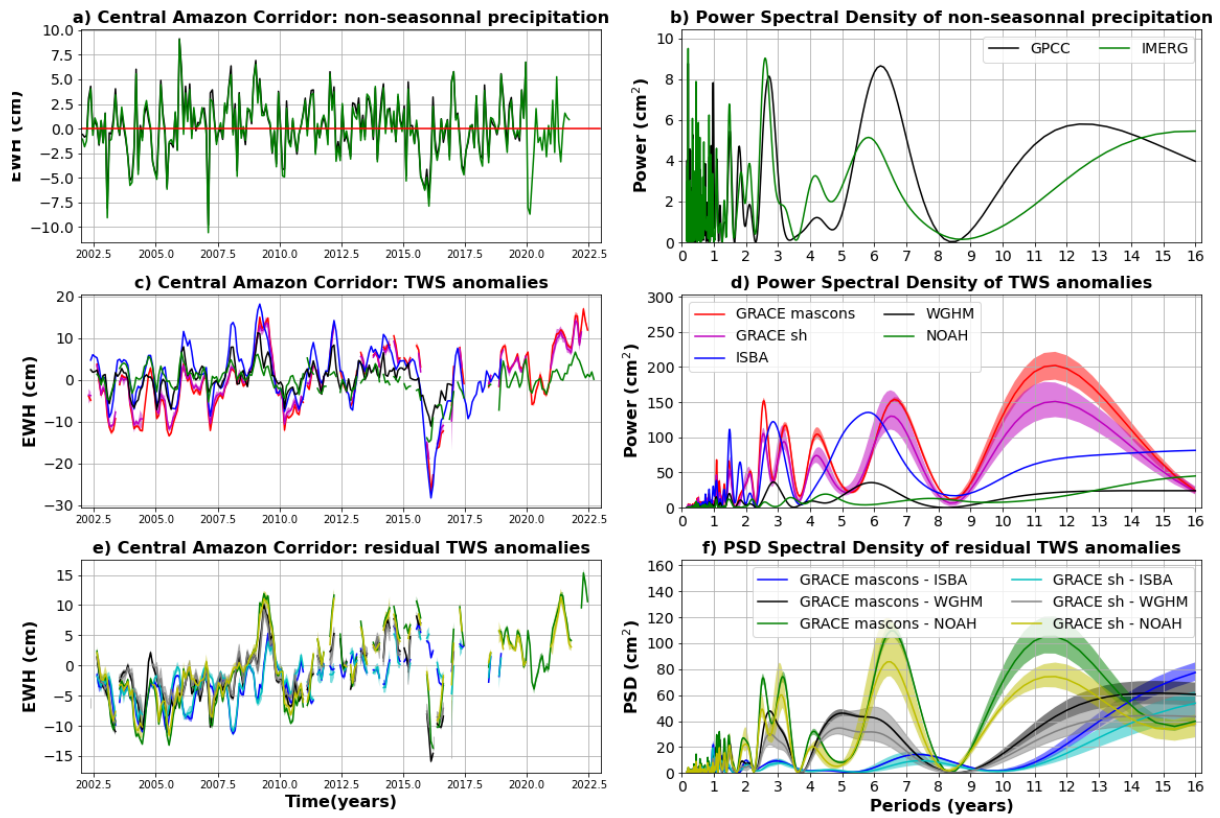


Figure S1.2 Comparison of TWS and precipitation anomalies averaged over the Central Amazon Corridor (box A in Fig. B1 - Appendix B). a) Average precipitation anomalies for the GPCC (gauge-based) and IMERG (satellite-based) products. b) Power Spectral Density (PSD) of average precipitation anomalies. c) TWS anomalies average over the central Amazon for two global hydrological models (ISBA-CTrip in blue and WGHM in black), one land surface model (NOAH in green) and 9 GRACE solutions (mascons in red, spherical harmonic in magenta). The solid line corresponds to the average of the sub-ensemble, the shaded area to the minimum to maximum envelope. d) PSD of the averaged TWS anomalies shown in (c). e) Residual TWS anomalies averaged over the central Amazon corridor and calculated as the difference between GRACE and ISBA-CTrip (blue when the difference is calculated with mascons, cyan with spherical harmonics), WGHM (black when the difference is calculated with mascons, grey with spherical harmonics) or NOAH (green when the difference is calculated with mascons, yellow with spherical harmonics).



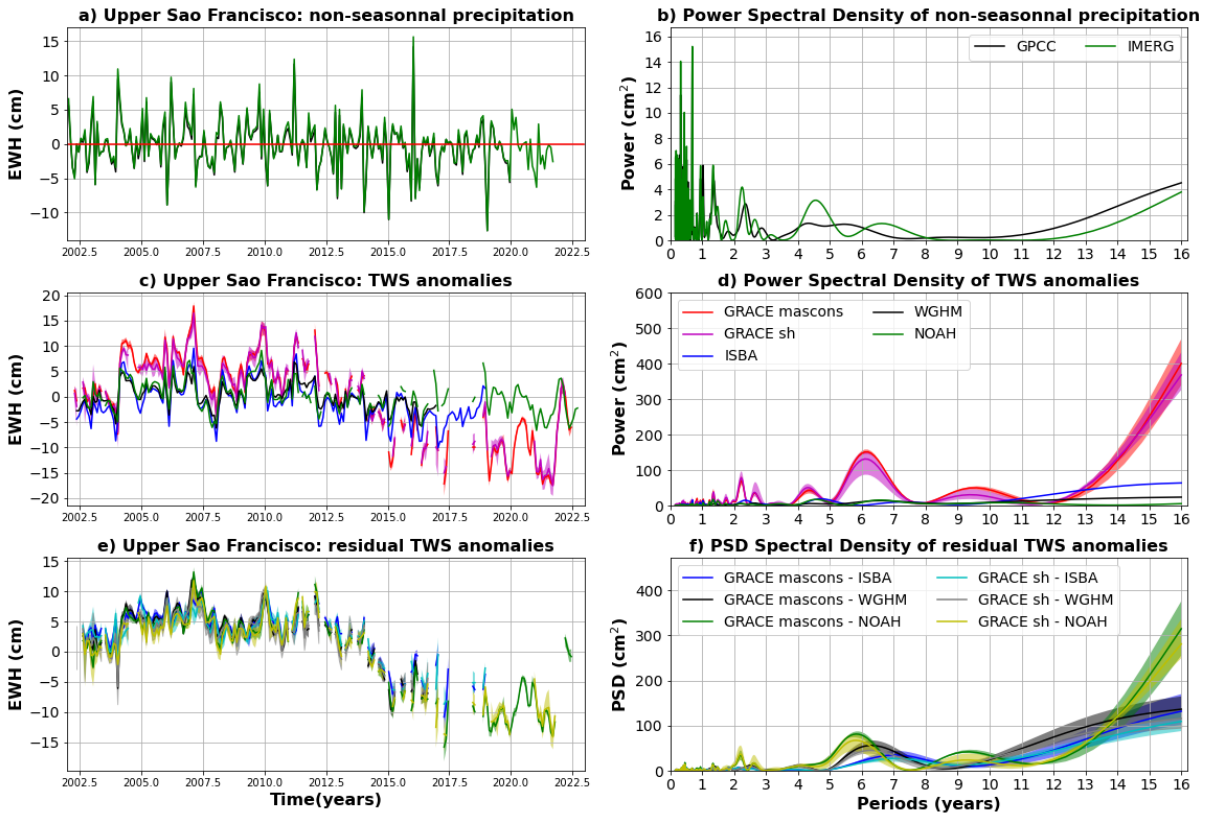


Fig. S1.3 Same as Fig. S1.2 but for the Upper Sao Francisco (box B in Fig. B1 - Appendix B).

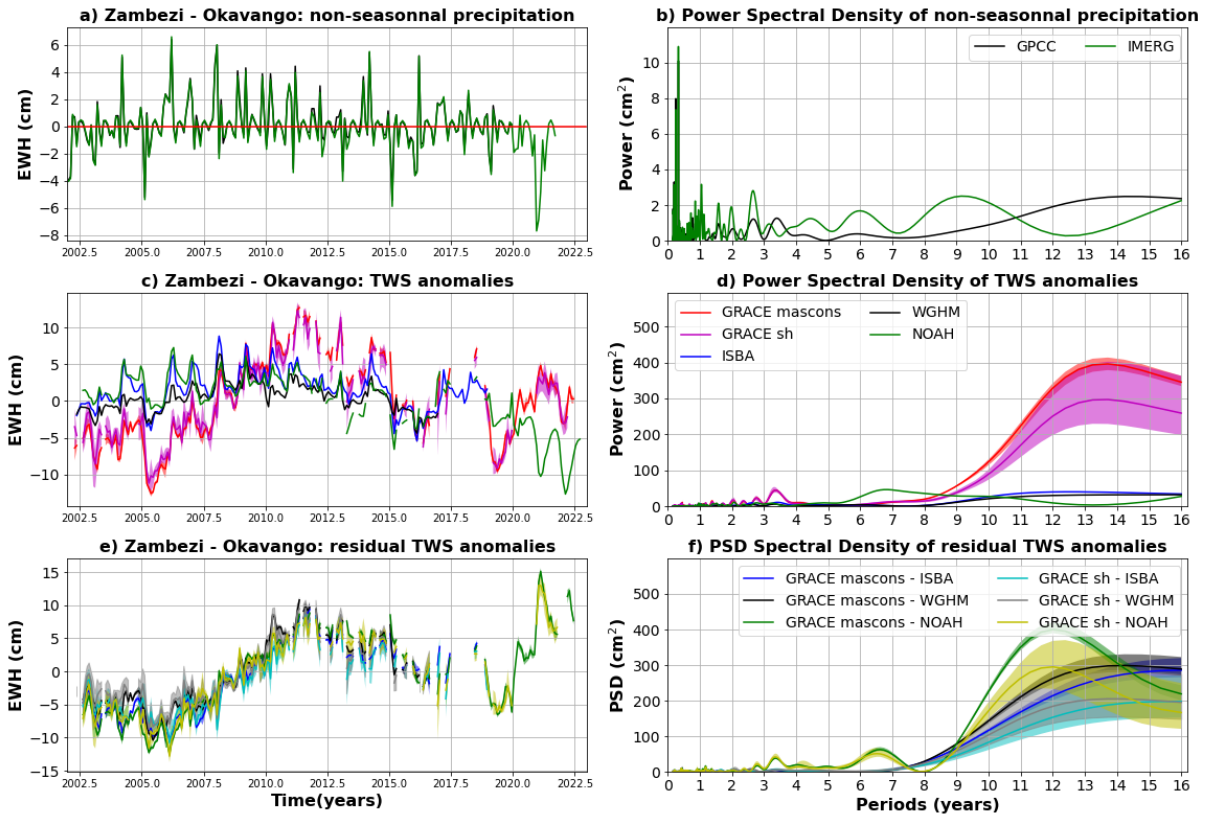


Fig. S1.4 Same as Fig. S1.2 but for the Zambezi (box C in Fig. B1 - Appendix B).

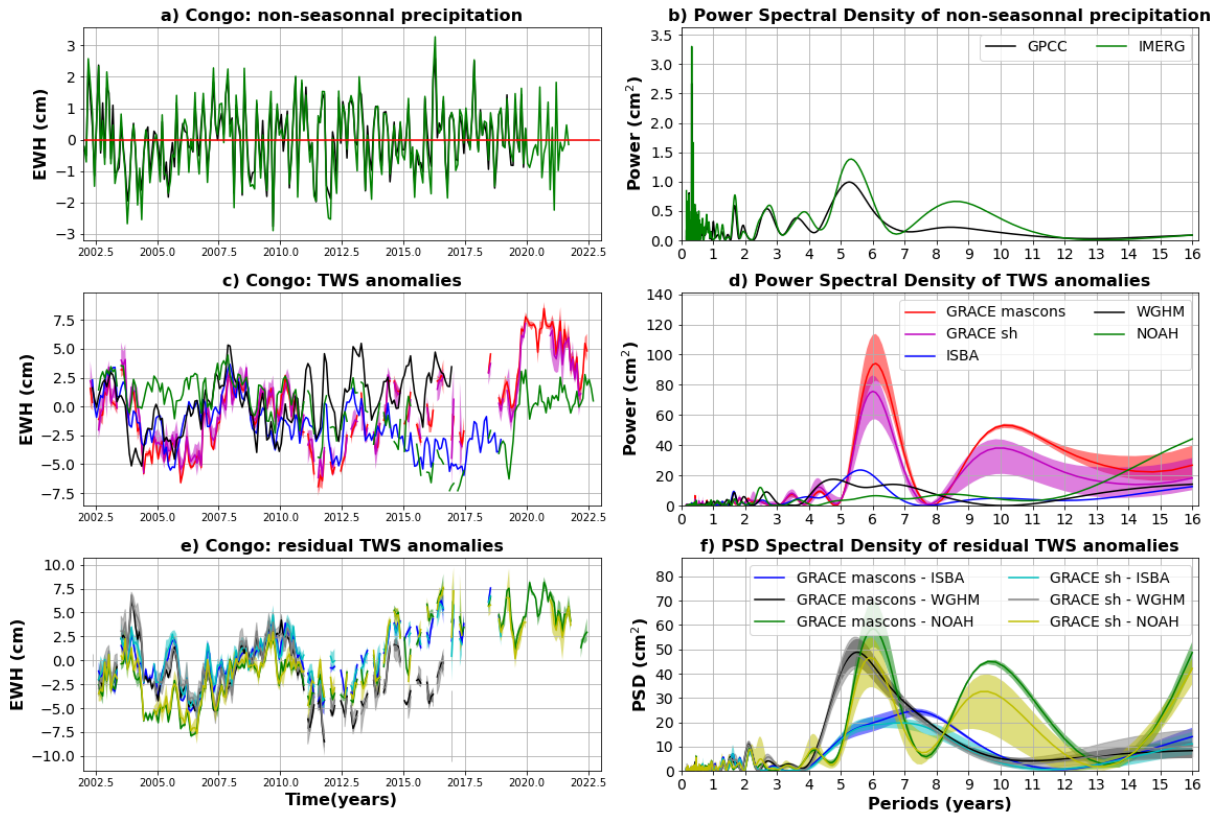


Fig. S1.5 Same as Fig. S1.2 but for the Congo (box D in Fig. B1 - Appendix B).

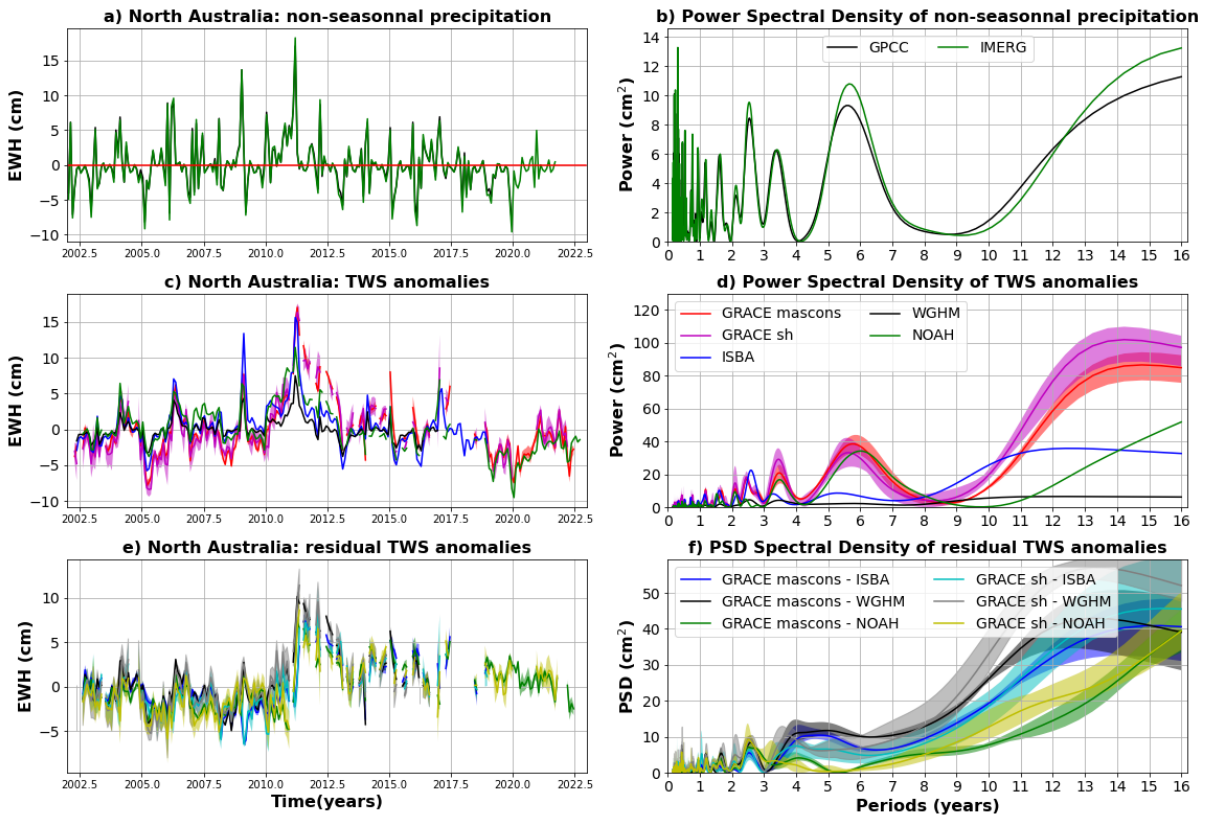


Fig. S1.6 Same as Fig. S1.2 but for North Australia (box E in Fig. B1 - Appendix B).

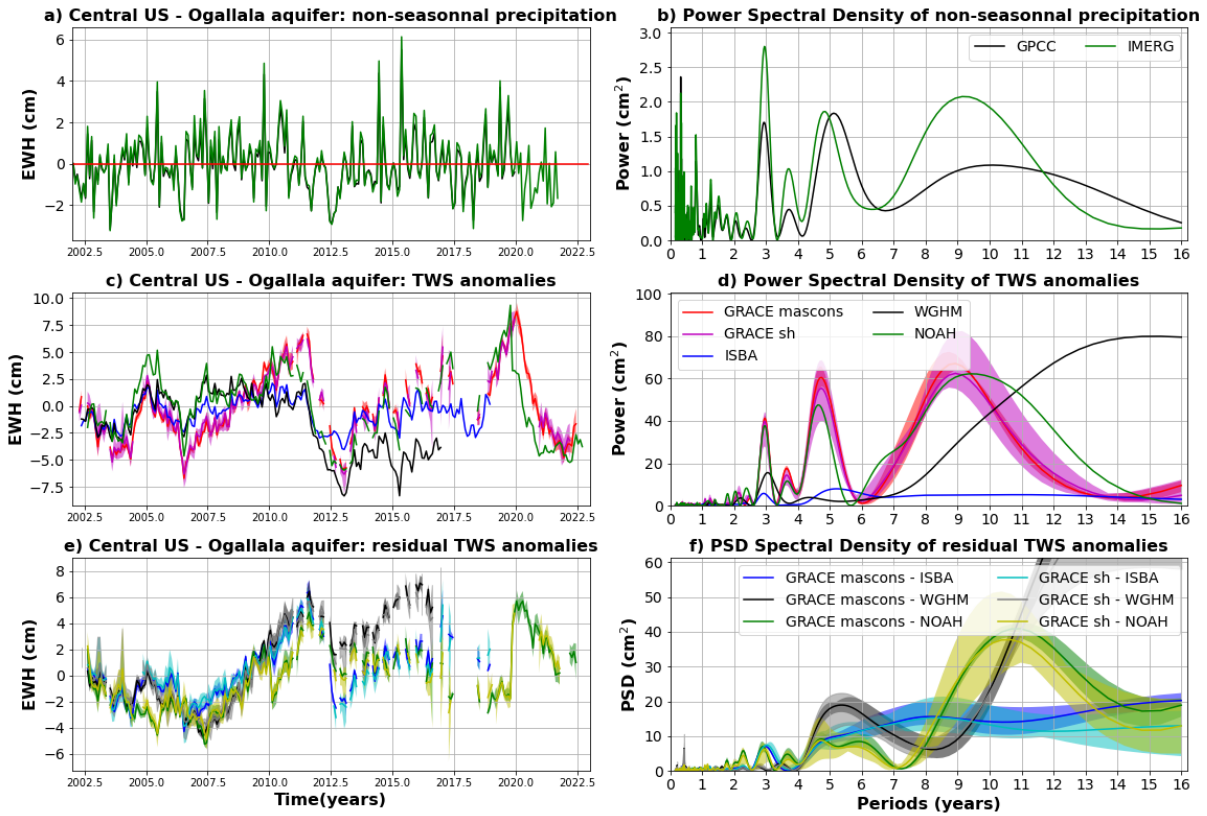


Fig. S1.7 Same as Fig. S1.2 but for the central US (box F in Fig. B1 - Appendix B).

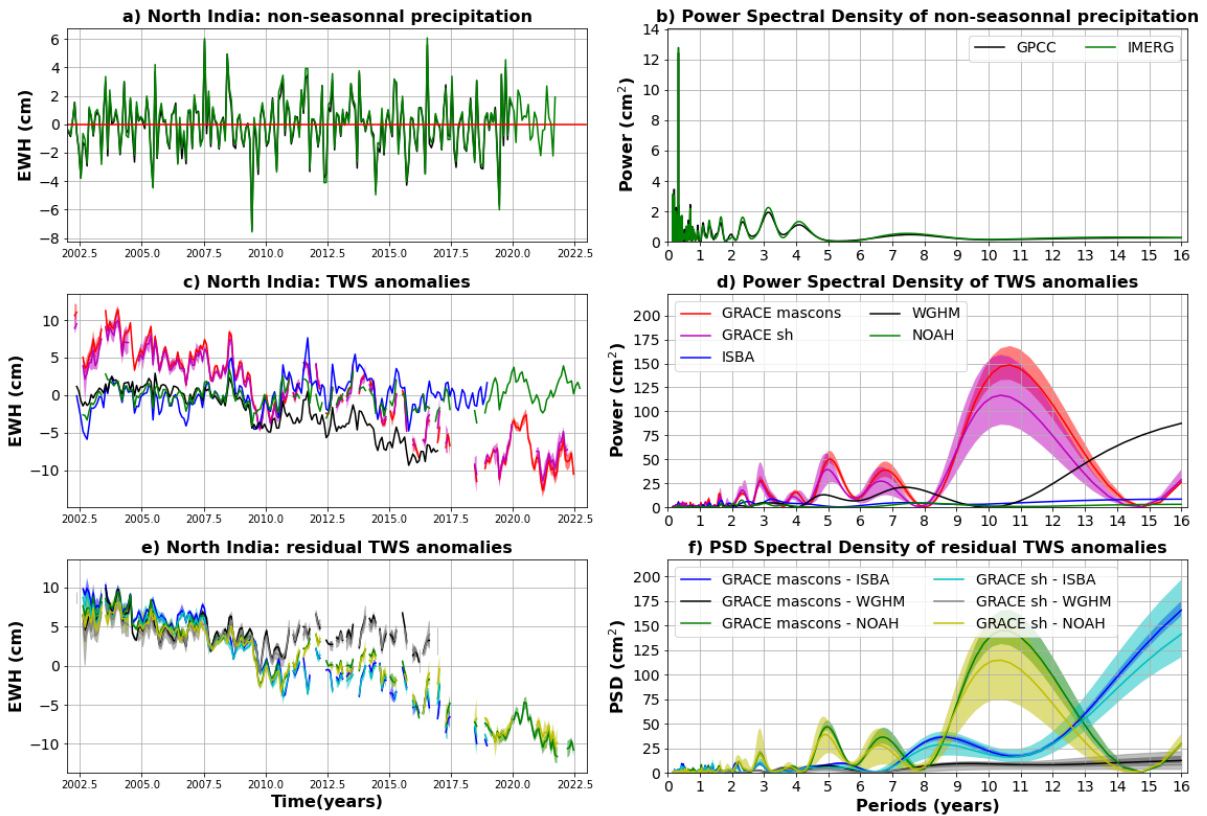


Fig. S1.8 Same as Fig. S1.2 but for North India (box G in Fig. B1 - Appendix B).

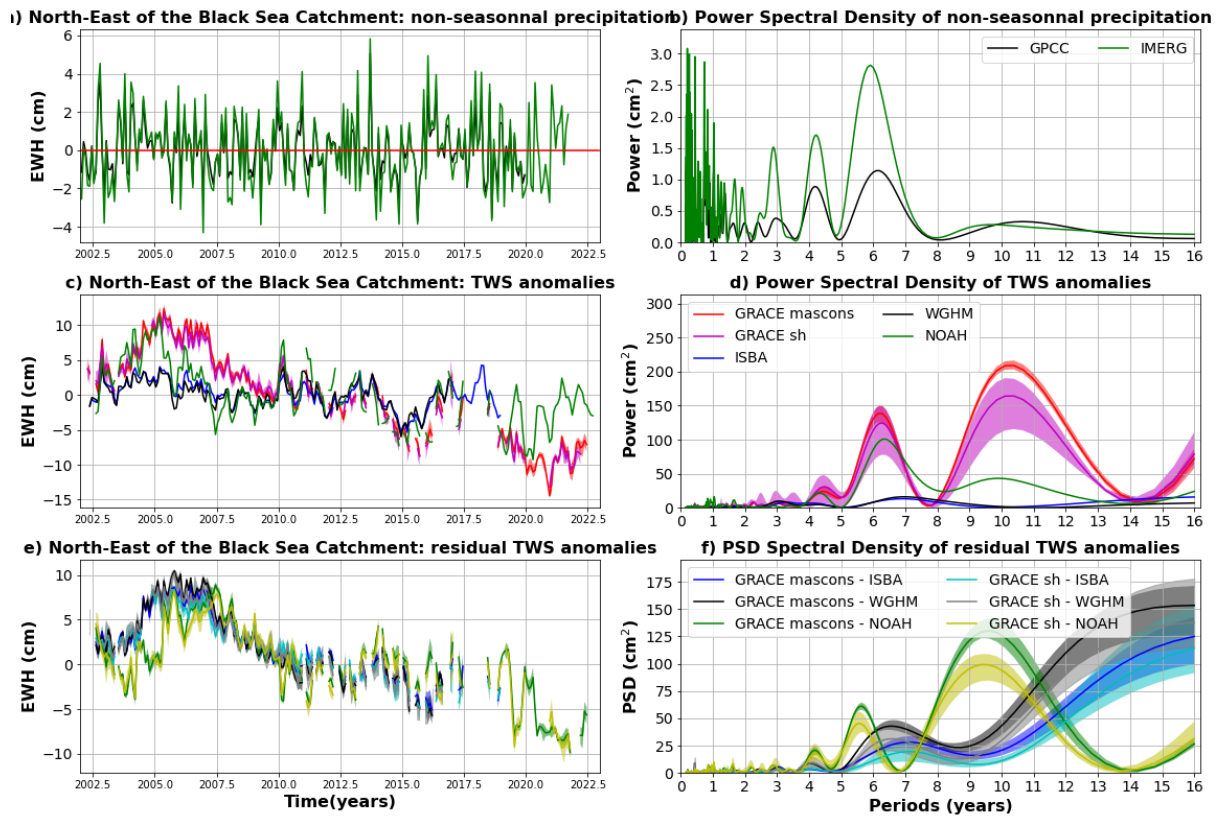


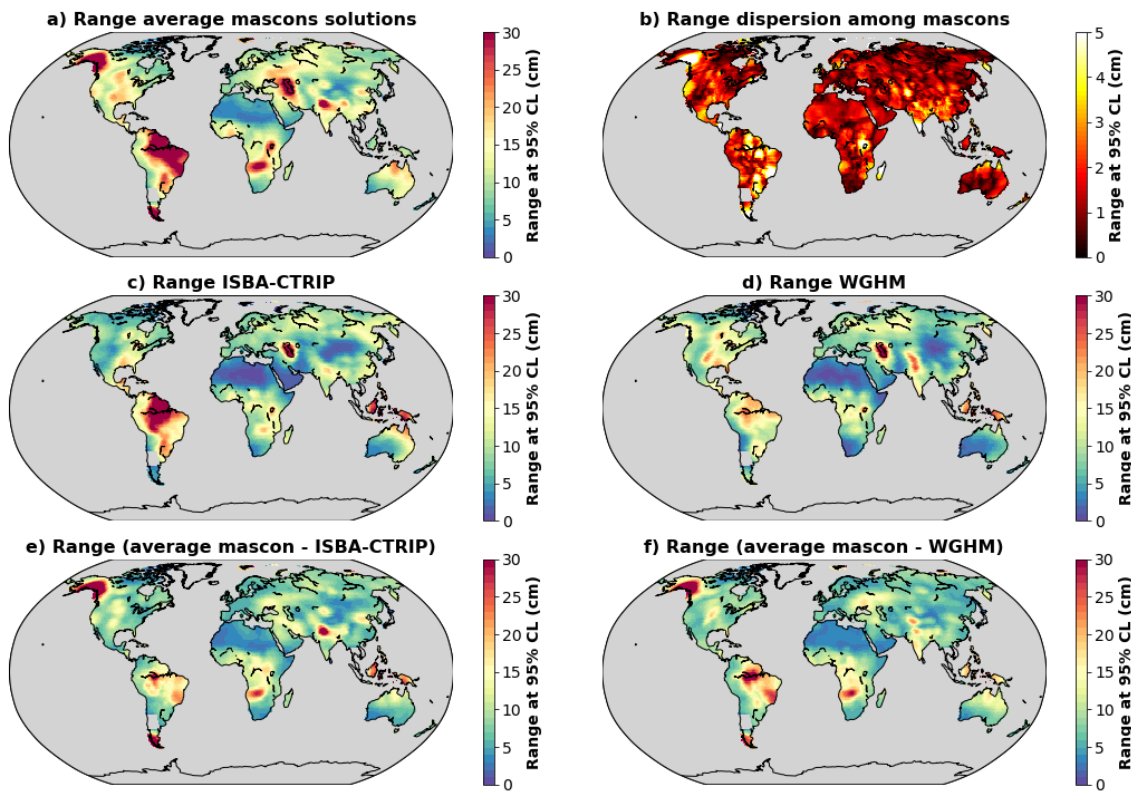
Fig. S1.9 Same as Fig. S1.2 but for the North of the Black Sea (box H in Fig. B1 - Appendix B).

## References

- Derber JC, Parrish DF, & Lord SJ (1991) The new global operational analysis system at the National Meteorological Center Weather Forecasting 6:538-547.
- Landerer F. W. and S. C. Swenson. 2012. Accuracy of scaled GRACE terrestrial water storage estimates, *Water Resources Research*, 48, 4. <https://doi.org/10.1029/2011WR011453>
- Rodell, M., et al., 2004. The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, <https://doi.org/10.1175/BAMS-85-3-381>.
- Rodell M, et al. 2020. TELLUS\_GLDAS-NOAH-3.3\_TWS-ANOMALY\_MONTHLY. Ver. 3.3. PO.DAAC, CA, USA. Dataset accessed 2023-07-26 at <https://doi.org/10.5067/GGDAS-3NH33>

## **Supplementary S2 : Comparison of global hydrological models with a GRACE-based mascon ensemble and a GRACE-based spherical harmonic ensemble**

We present here a comparison of TWS anomalies predicted with global hydrological models and observed with GRACE, for which we separate the mascon solutions from the spherical harmonic solutions. The GRACE solutions are remarkably similar one with another, whether we consider mascons or spherical harmonic solutions. The same conclusions can be drawn for the full GRACE ensemble presented in the manuscript and the mascons or spherical harmonic sub-ensembles. The amplitude of pluri-annual to decadal signals is underestimated in global hydrological models when compared to GRACE, regardless of the type of solution considered: mascons or spherical harmonics. It may be noted that the dispersion is larger among the sub-ensemble of 6 spherical harmonic solutions than among the 3 mascon solutions. The biases between GRACE solutions and global hydrological models is however much larger than the dispersion of spherical harmonic solutions.



**Figure S2.1: Comparison of TWS anomalies estimated from an ensemble of three GRACE mascon solutions and two global hydrological models. The amplitude of the non-seasonal TWS variability is expressed as the range at 95% CL, calculated as the difference between the 97.5 and 2.5 percentiles of the TWS anomalies obtained in each grid cell over the entire study period. a) Range of TWS anomalies estimated as the average of three GRACE mascons solutions. b) Dispersion of the range of TWS anomalies among three GRACE mascons solutions. Range of TWS anomalies estimated with ISBA-CTRIP (c) and WGHM (d). Range of residual TWS anomalies estimated as the difference between the average of three GRACE mascon solutions and ISBA-CTRIP (e) or WGHM (f).**

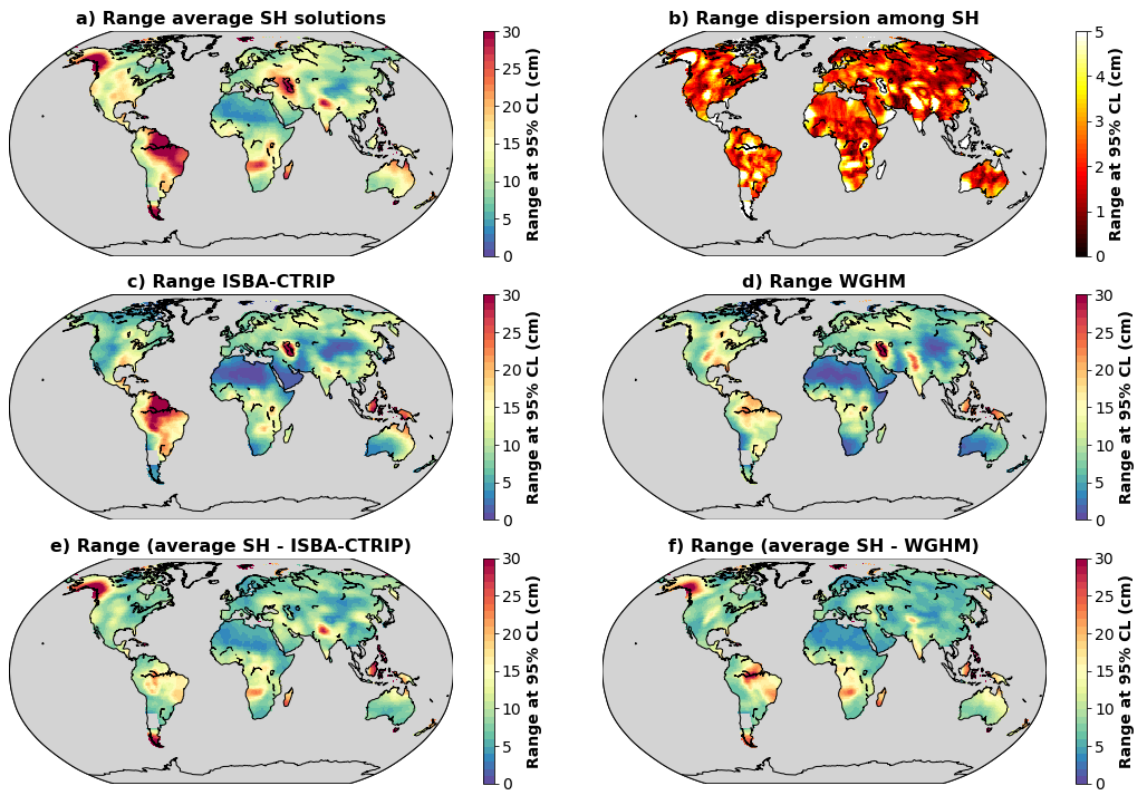


Figure S2.2: Comparison of TWS anomalies estimated from an ensemble of six GRACE spherical harmonic solutions and two global hydrological models. The amplitude of the non-seasonal TWS variability is expressed as the range at 95% CL, calculated as the difference between the 97.5 and 2.5 percentiles of the TWS anomalies obtained in each grid cell over the entire study period. a) Range of TWS anomalies estimated as the average of six GRACE spherical harmonics solutions. b) Dispersion of the range of TWS anomalies among six GRACE spherical harmonics solutions. Range of TWS anomalies estimated with ISBA-CTRIP (c) and WGHM (d). Range of residual TWS anomalies estimated as the difference between the average of six GRACE spherical harmonic solutions and ISBA-CTRIP (e) or WGHM (f).

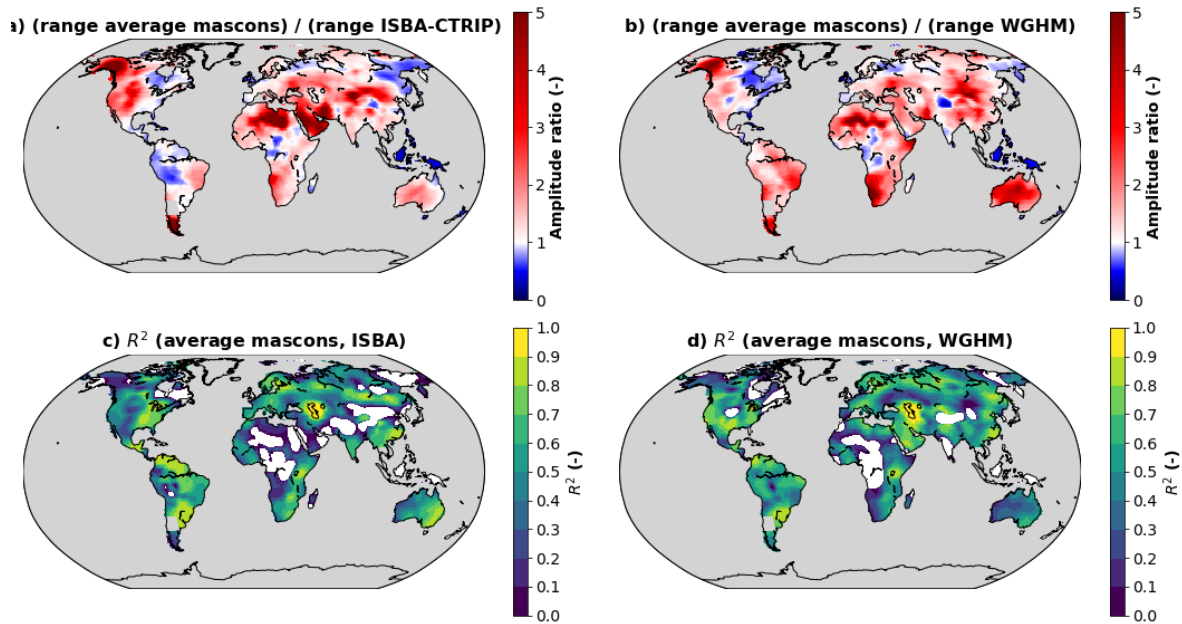


Figure S2.3: Range ratios between the average of three GRACE mascon solutions and the hydrological models ISBA-CTRIP (a) and WGHM (b). Determination coefficients between the average GRACE mascon solution and the hydrological models ISBA-CTRIP (c) and WGHM (d). Regions, where the coefficient of determination is negative, are shown in white

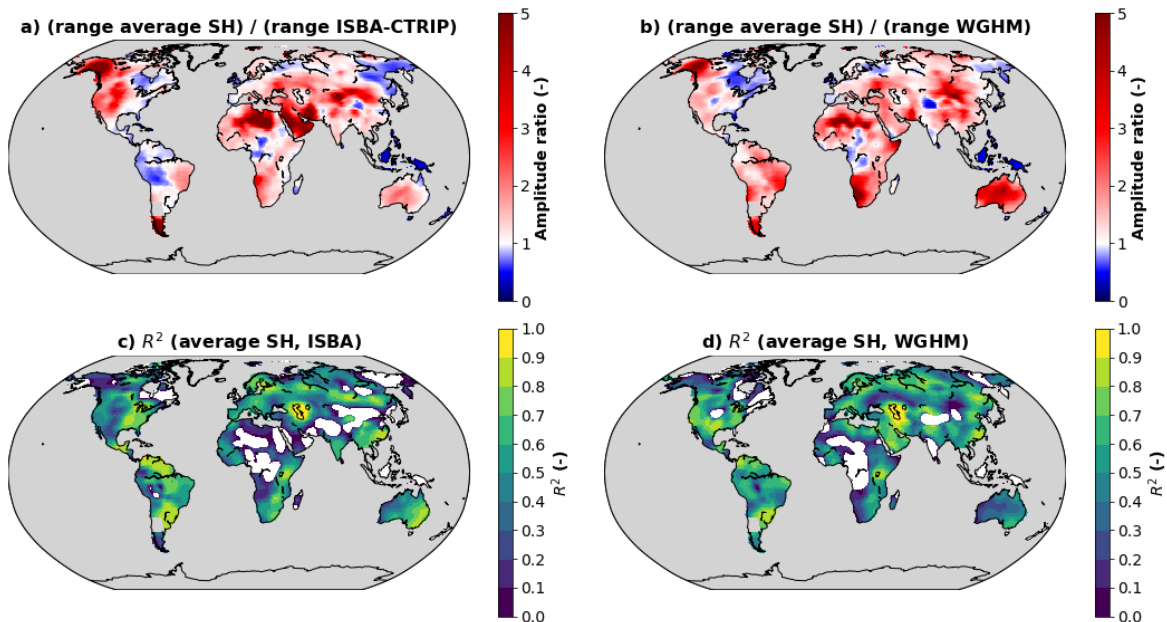
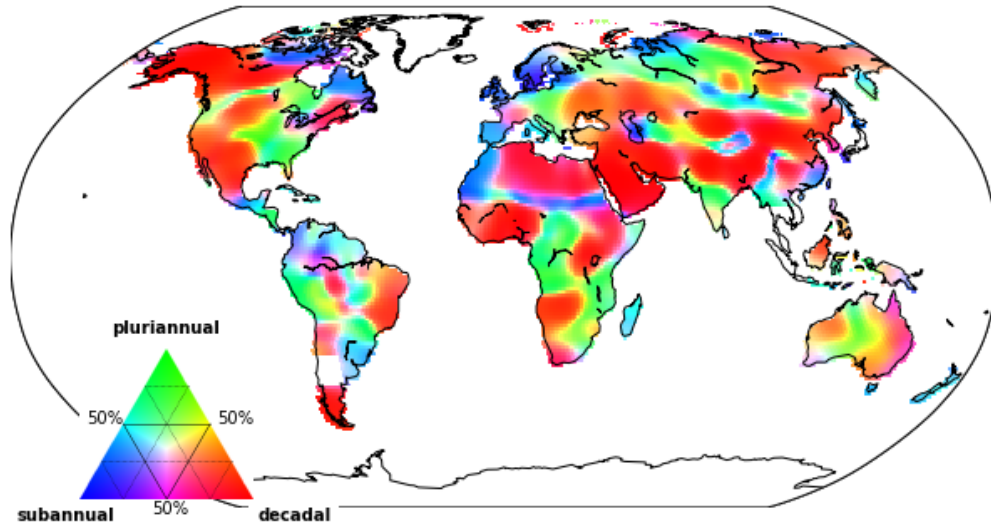


Figure S2.4: Range ratios between the average of six GRACE spherical harmonic solutions and the hydrological models ISBA-CTRIP (a) and WGHM (b). Determination coefficients between the average GRACE spherical harmonic solution and the hydrological models ISBA-CTRIP (c) and WGHM (d). Regions, where the coefficient of determination is negative, are shown in white

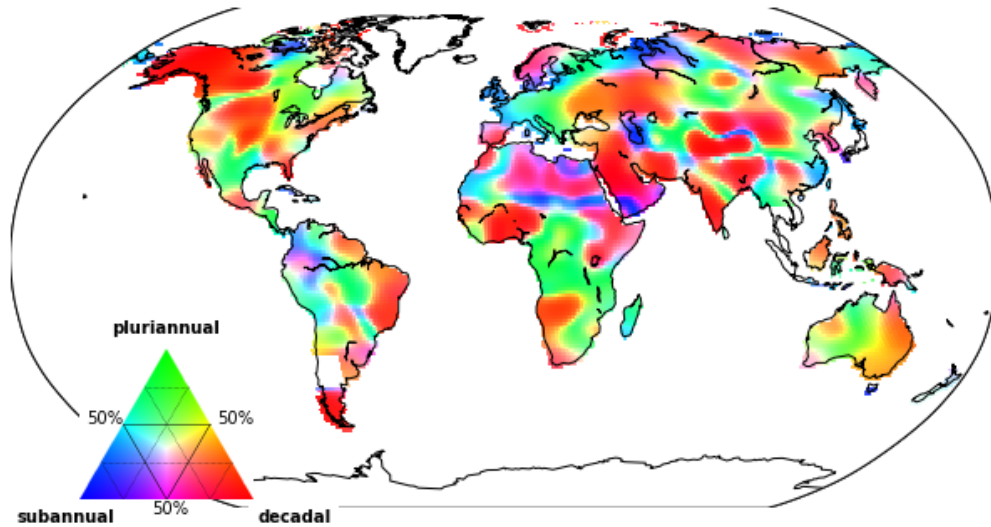




**a) Contribution of subannual, pluri-annual and decadal signals in residual TWS anomalies calculated as the difference between mascons and ISBA**

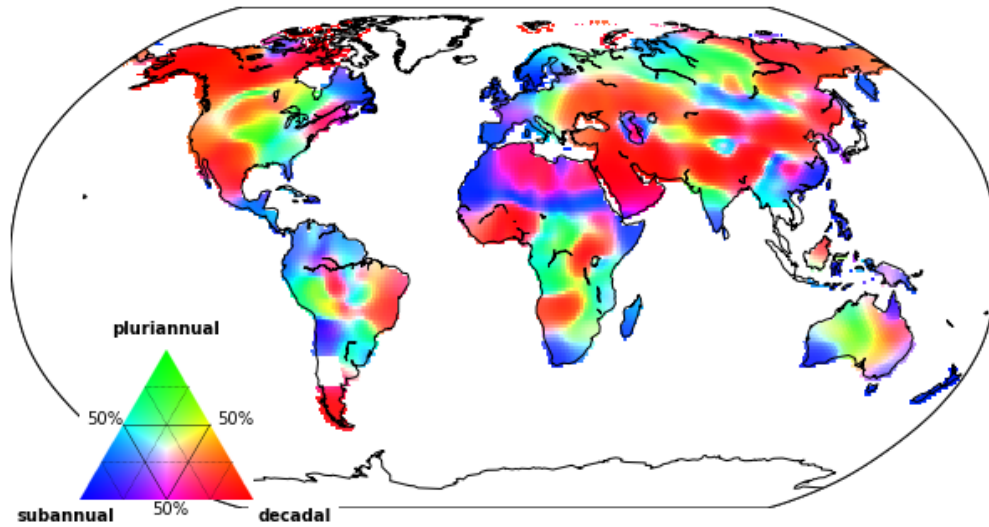


**b) Contribution of subannual, pluri-annual and decadal signals in residual TWS anomalies calculated as the difference between mascons and WGHM**

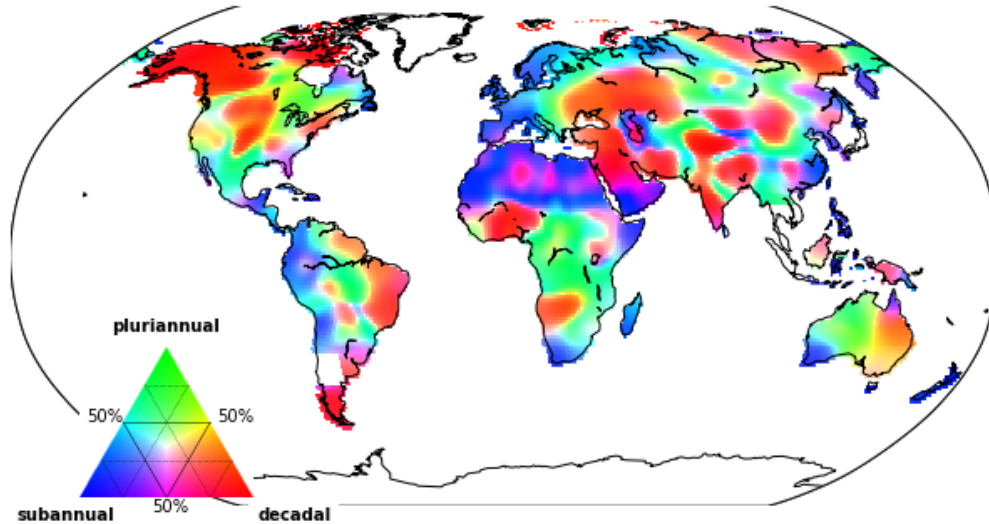


**Figure S2.5: Characteristic time scales in residual TWS anomalies calculated as the differences between the average of three GRACE mascon solutions and ISBA-CTRIP (a) or WGHM (b). Subannual, pluriannual and decadal contributions have been computed with high-pass (cut-off period at 1.5 years), band-pass (cut-off periods at 1.5 and 10 years) and low-pass (cut-off period at 10 years) filters respectively. The percentage of variance explained by one contribution has been calculated as the coefficient of determination with respect to the full residual signal.**

**a) Contribution of subannual, pluri-annual and decadal signals in residual TWS anomalies calculated as the difference between sh and ISBA**



**b) Contribution of subannual, pluri-annual and decadal signals in residual TWS anomalies calculated as the difference between sh and WGHM**



**Figure S2.6: Characteristic time scales in residual TWS anomalies calculated as the differences between the average of six GRACE spherical harmonic solutions and ISBA-CTRIP (a) or WGHM (b). Subannual, pluriannual and decadal contributions have been computed with high-pass (cut-off period at 1.5 years), band-pass (cut-off periods at 1.5 and 10 years) and low-pass (cut-off period at 10 years) filters respectively. The percentage of variance explained by one contribution has been calculated as the coefficient of determination with respect to the full residual signal.**

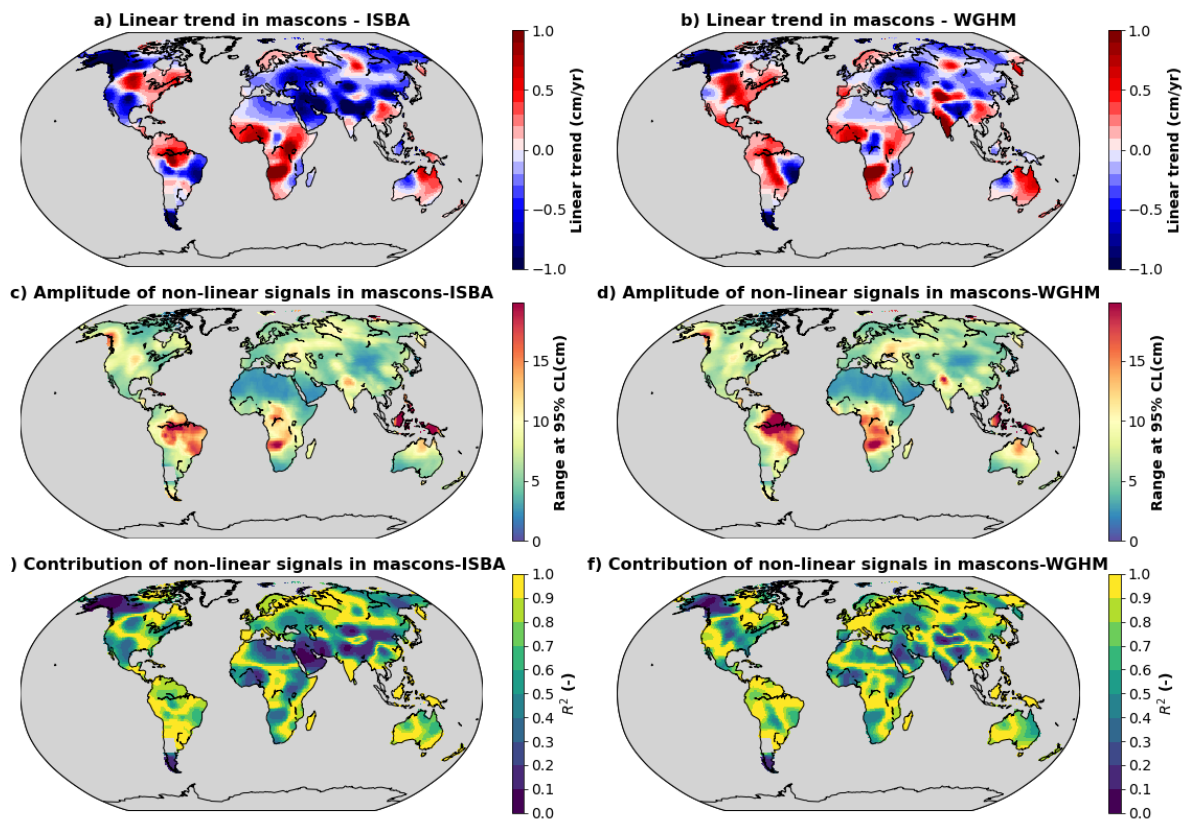
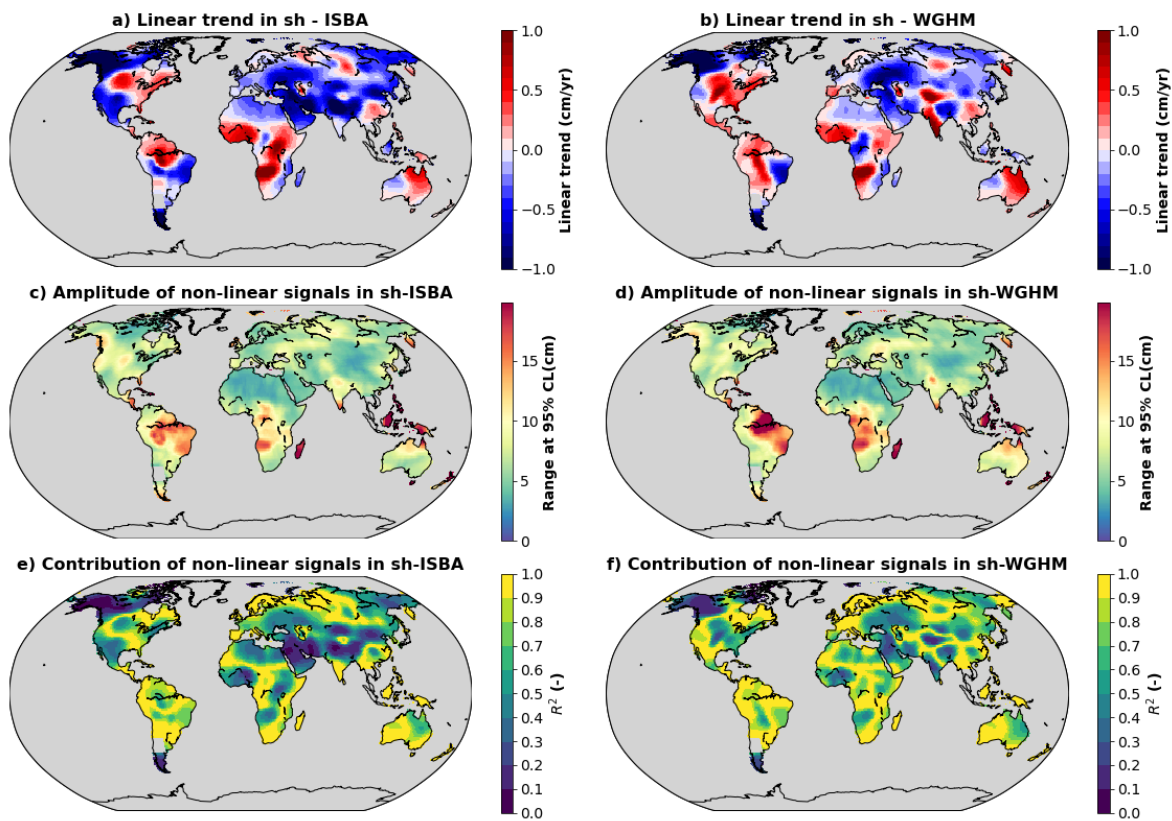


Figure S2.7: a) Linear trends in residual TWS anomalies calculated as the difference between the average of three GRACE mascon solutions and ISBA-CTRIP. b) Same as (a) with WGHM. c) Amplitude of non-linear signals in residual TWS anomalies calculated as the difference between the average of three GRACE mascon solutions and ISBA-CTRIP. The amplitude is calculated as the difference between the 97.5 and 2.5 percentiles. d) Same as (c) with WGHM. e) Coefficient of determination calculated for non-linear signals with respect to TWS anomalies calculated as the difference between the average GRACE mascon solution and ISBA-CTRIP. f) Same as (e) with WGHM.



**Figure S2.8:** a) Linear trends in residual TWS anomalies calculated as the difference between the average of six GRACE spherical harmonic solutions and ISBA-CTRIP. b) Same as (a) with WGHM. c) Amplitude of non-linear signals in residual TWS anomalies calculated as the difference between the average of six GRACE spherical harmonic solutions and ISBA-CTRIP. The amplitude is calculated as the difference between the 97.5 and 2.5 percentiles. d) Same as (c) with WGHM. e) Coefficient of determination calculated for non-linear signals with respect to TWS anomalies calculated as the difference between the average GRACE mascon solution and ISBA-CTRIP. f) Same as (e) with WGHM.

### **Supplementary S3 Comparison of TWS anomalies from GRACE and global hydrological models over large river basins**

Non-seasonal precipitation, TWS and residual TWS anomalies have been calculated and plotted for the 40 largest river basins of the world (Fig C1) according to the Global Runoff Data Centre (GRDC) Major River Basins (MRB) database (GRDC, 2020). The main conclusions drawn from global (section 3, main text) and regional (section 4, main text) analyses remain valid at basin scale. In particular, large residual TWS anomalies are observed at pluri-annual and decadal timescales, due to an underestimation of slow TWS anomalies by the two global hydrological models considered in this study (ISBA-CTRIP and WGHM) when compared to GRACE. The amplitude of ISBA-CTRIP TWS predictions is closer to GRACE in remote river basins such as the Amazon, Lake Eyre, Murray Darling, Nelson, Okavango, Orinoco, Orange and Zambezi basins. WGHM better predicts TWS anomalies observed by GRACE in anthropized basins such as the Aral Sea, Colorado, Columbia, Ganges, Indus, Rio Grande or Yellow River basins. The difference of behaviour between both hydrological models is however not systematic. For example, the TWS predictions from ISBA-CTRIP are closer to GRACE than WGHM across the Mississippi, Parana, Saint Lawrence or Yangtze basins, which are significantly affected by human interventions. Adversely, WGHM predictions fit better GRACE-based TWS anomalies than ISBA-CTRIP across the remote Yenisei and Kolyma river basins. However, it must be noted that large discrepancies are observed for both models when compared to GRACE for the Yenisei and Kolyma basins. Indeed, for a majority of basins (Dnieper, Danube, Amur, Brahmaputra, Congo, Chad, Jubba, Lena, Mackenzie, Mekong, Niger, Nile, Ob, Sao Francisco, Shatt Al Arab, Tarim He, Tocantins, Volga, Yukon), both models struggle to reproduce non-seasonal TWS anomalies at pluri-annual and decadal time-scales.

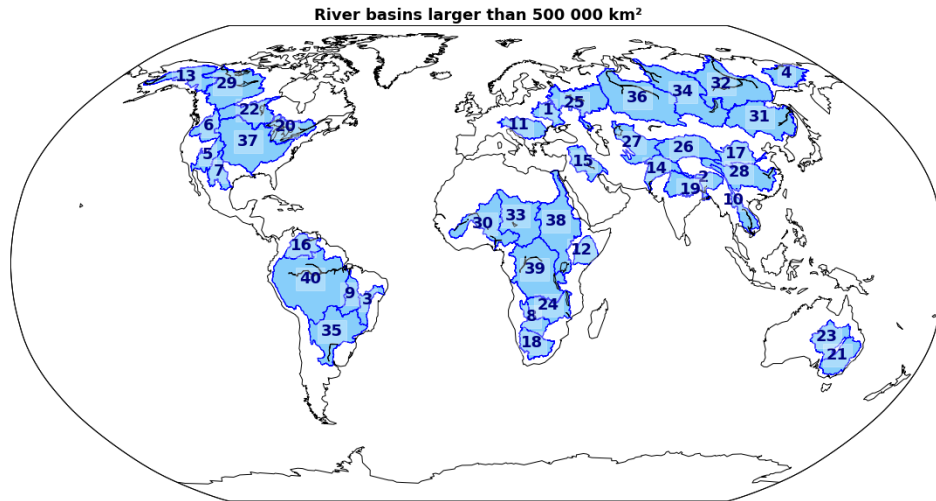


Figure S3.1: Map of the 40 largest river basins considered in this study: 1) Dnieper, 2) Brahmaputra, 3) Sao Francisco, 4) Kolyma, 5) Colorado, 6) Columbia, 7) Rio Grande, 8) Okavango, 9) Tocantins, 10) Mekong, 11) Danube, 12) Jubba, 13) Yukon, 14) Indus, 15) Shatt Al Arab, 16) Orinoco, 17) Yellow River, 18) Orange, 19) Ganges, 20) Saint Lawrence, 21) Murray, 22) Nelson, 23) Lake Eyre, 24) Zambezi, 25) Volga, 26) Tarim He, 27) Aral Sea, 28) Yangtze, 29) Mackenzie, 30) Niger, 31) Amur, 32) Lena, 33) Chad, 34) Yenisei, 35) Parana, 36) Ob, 37) Mississippi, 38) Nile, 39) Congo, 40) Amazon.

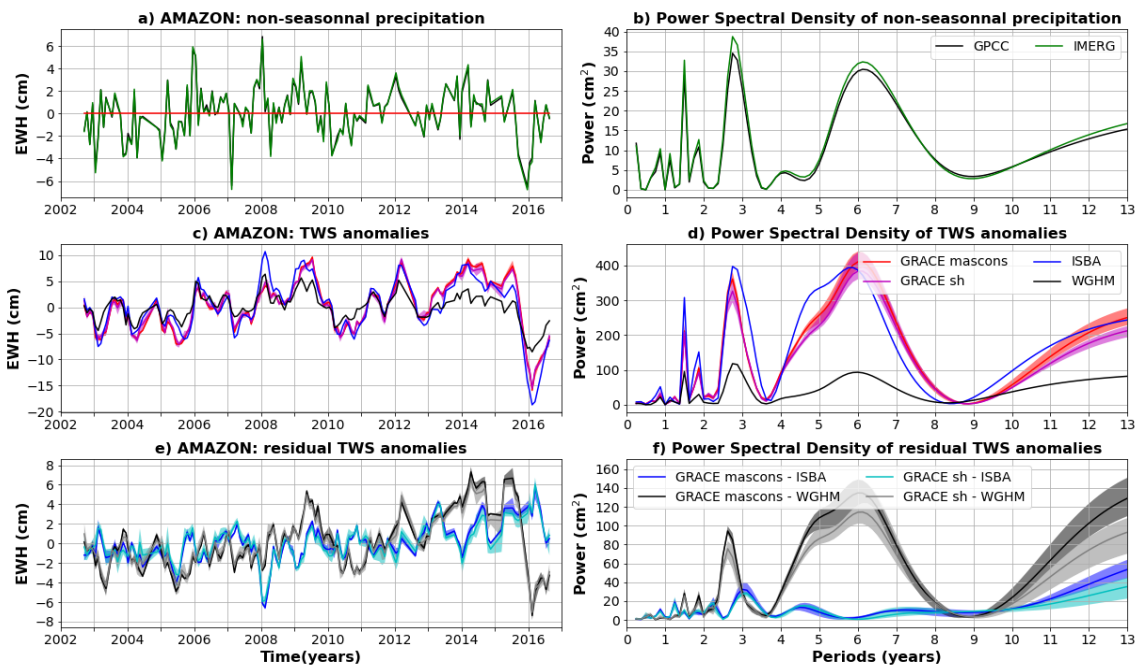


Figure S3.2: Comparison of TWS and precipitation anomalies averaged over Amazon basin. a) Average precipitation anomalies for the GPCC (gauge-based) and IMERG (satellite-based) products. b) Power Spectral Density (PSD) of average precipitation anomalies. c) TWS anomalies average over the central Amazon for two global hydrological models (ISBA-CTrip in blue and WGHM in black) and 9 GRACE solutions (mascons in red, spherical harmonic in magenta). The solid line corresponds to the average of the sub-ensemble, the shaded area to the minimum to maximum envelope. d) PSD of the averaged TWS anomalies shown in (c). e) Residual TWS anomalies averaged over the central Amazon corridor and calculated as the difference between GRACE and ISBA-CTrip (blue when the difference is calculated with mascons, cyan with spherical harmonics) or WGHM (black when the difference is calculated with mascons, grey with spherical harmonics).

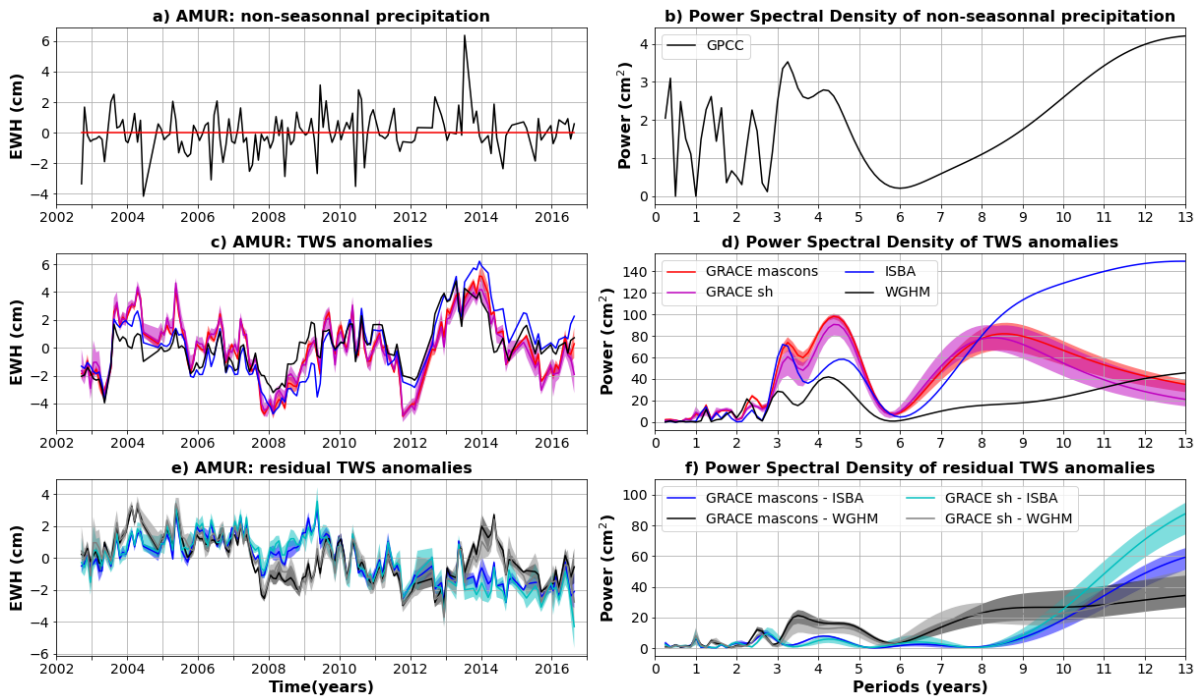


Figure S3.3: Same as S3.2 for the Amur Basin. Non-seasonal precipitation anomalies are only estimated with GPCC, as a significant part of the basin is not covered by IMERG satellites due to the high latitude of the Amur basin.

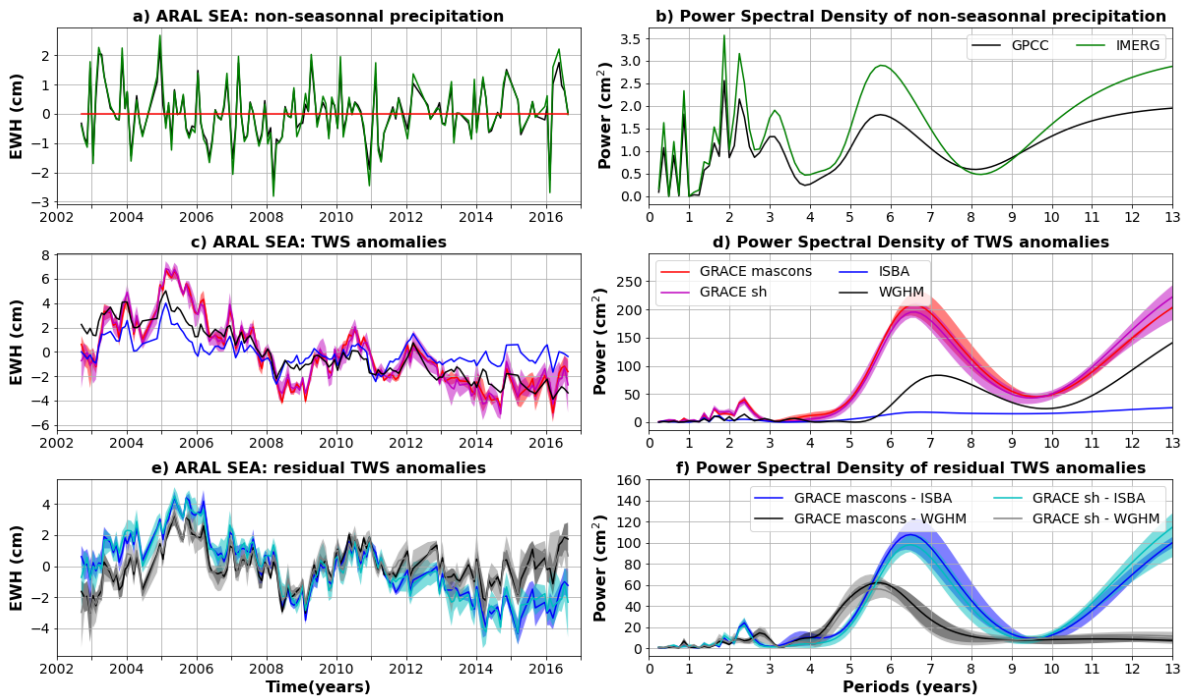


Figure S3.4: Same as S3.2 for the Aral Sea basin.



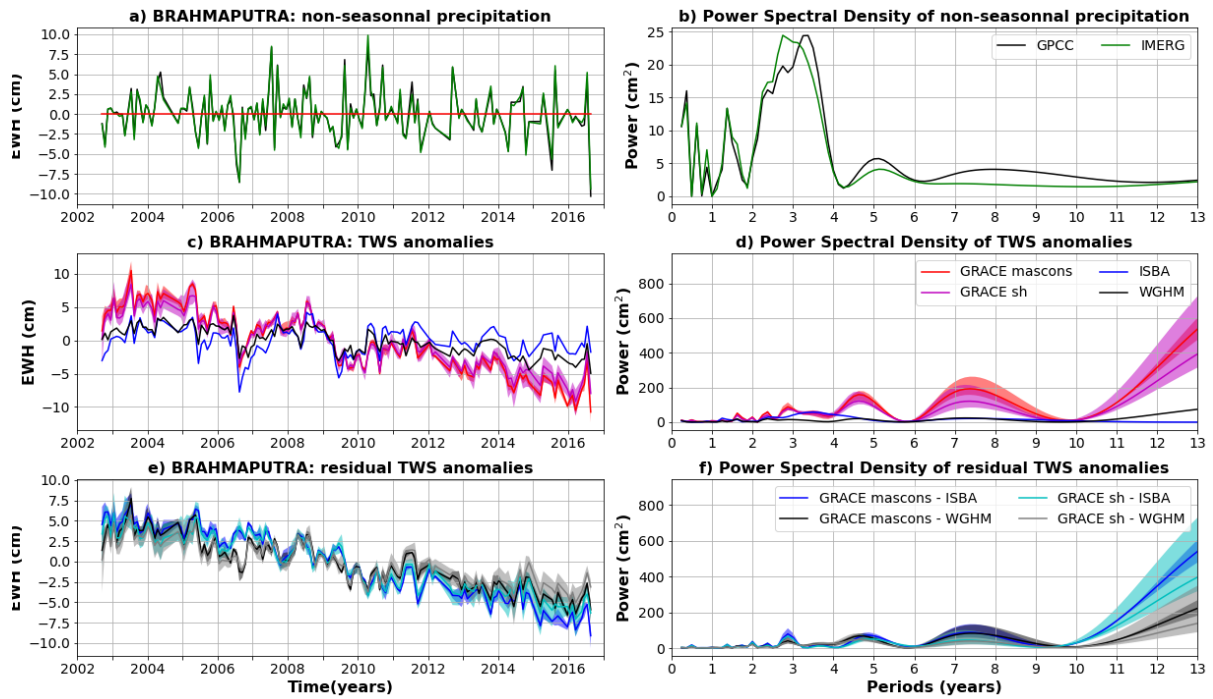


Figure S3.5: Same as S3.2 for the Brahmaputra basin.

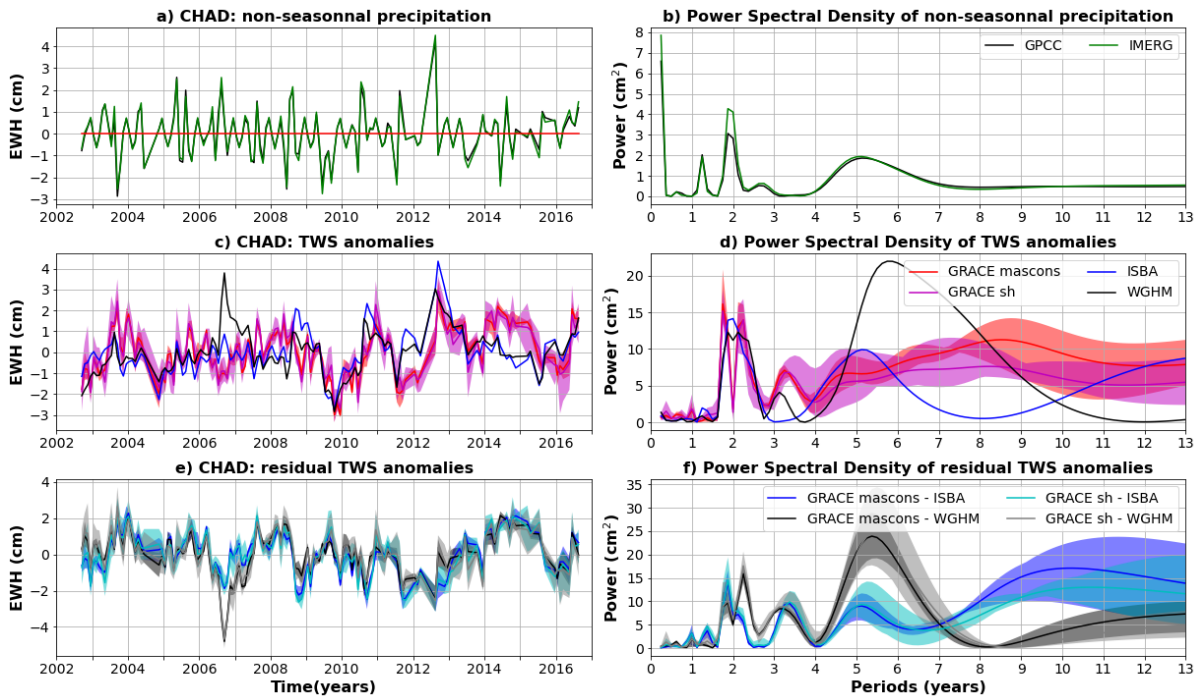


Figure S3.6: Same as S3.2 for the Chad basin.

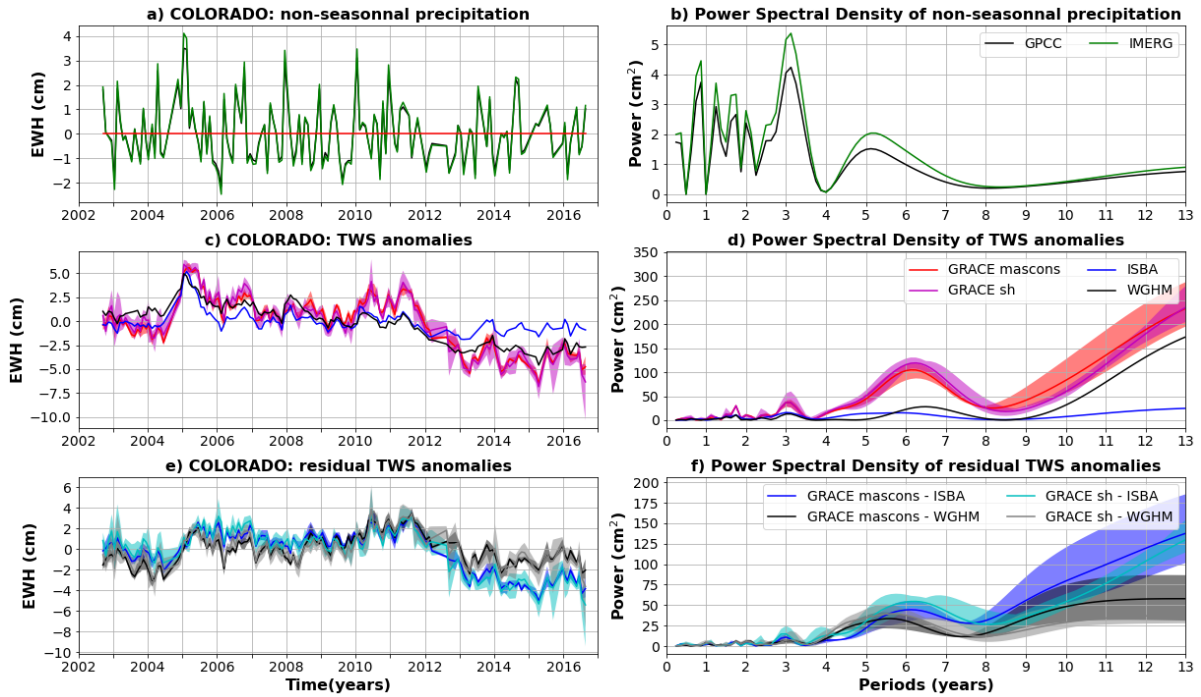


Figure S3.7: Same as S3.2 for the Colorado basin.

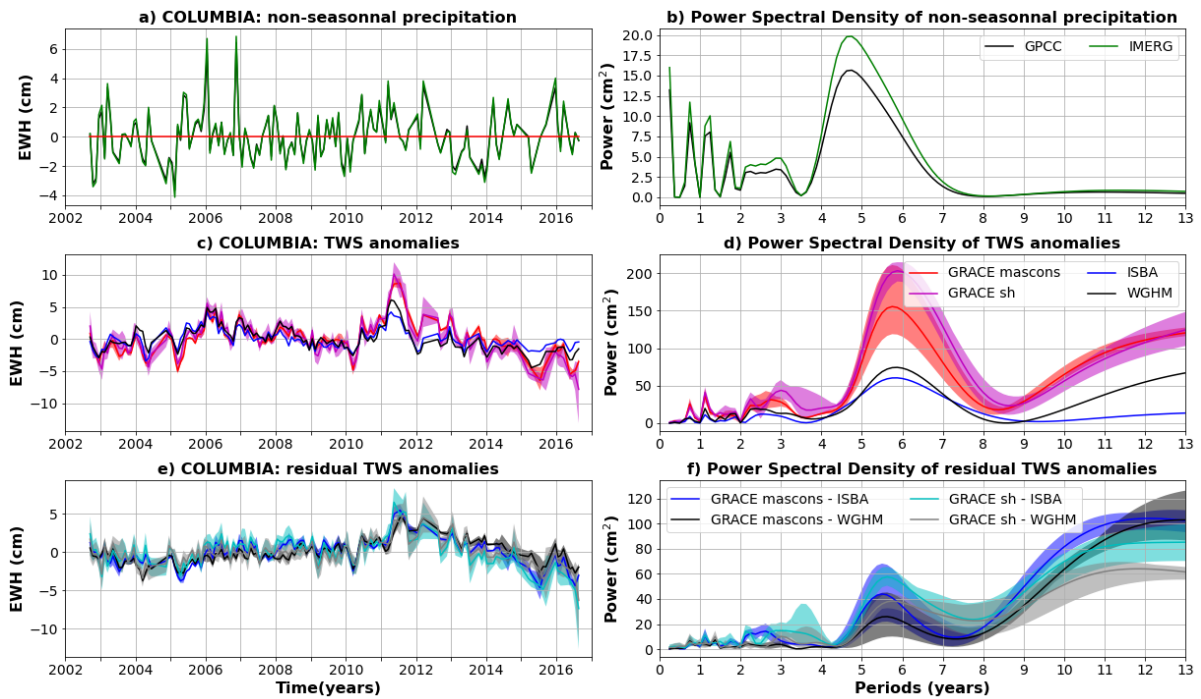


Figure S3.8: Same as S3.2 for the Columbia basin.

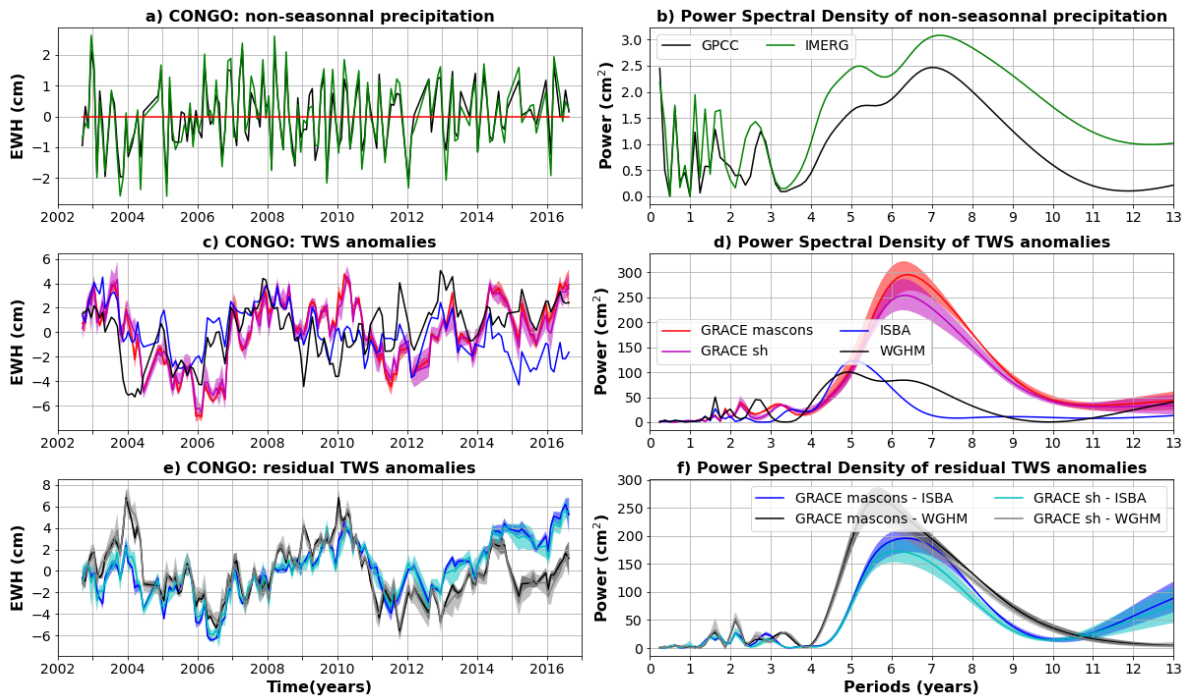


Figure S3.9: Same as S3.2 for the Congo basin.

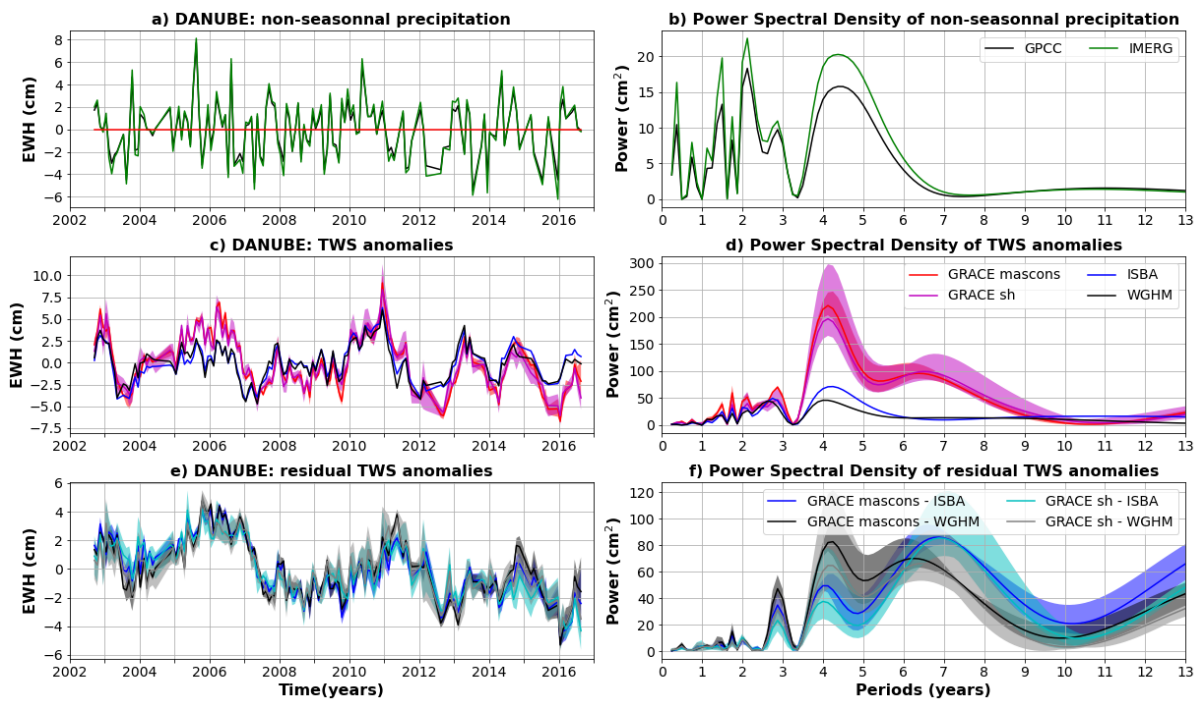


Figure S3.10: Same as S3.2 for the Danube basin.

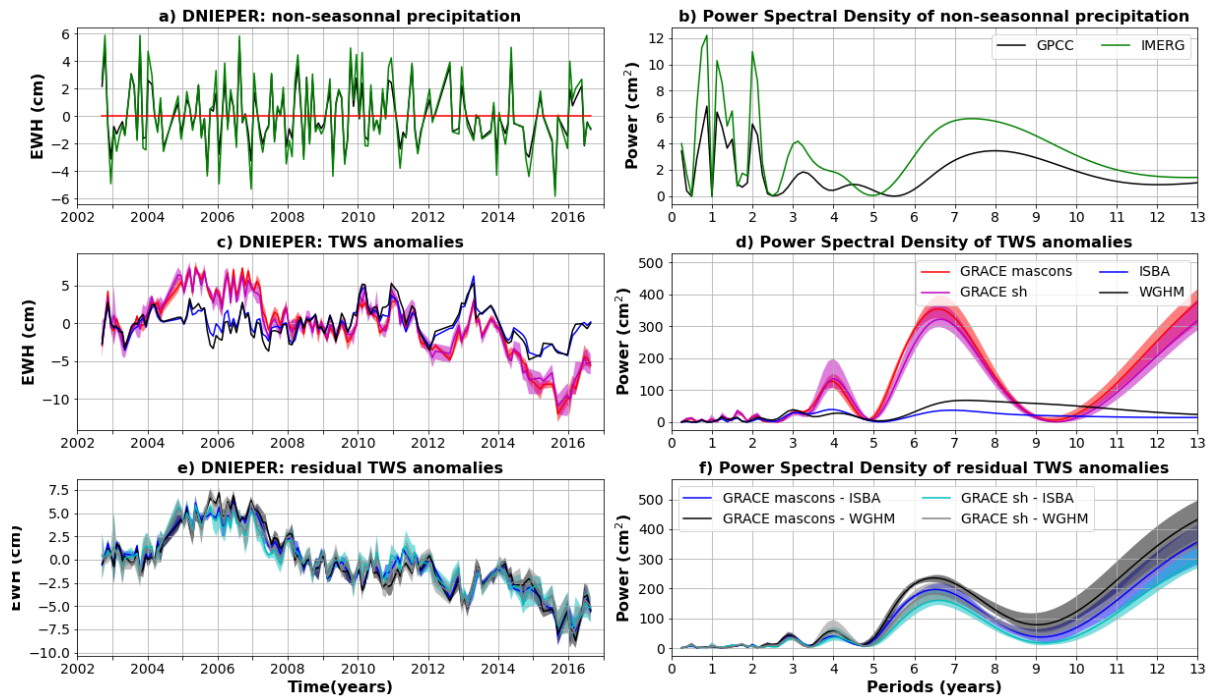


Figure S3.11: Same as S3.2 for the Dnieper basin.

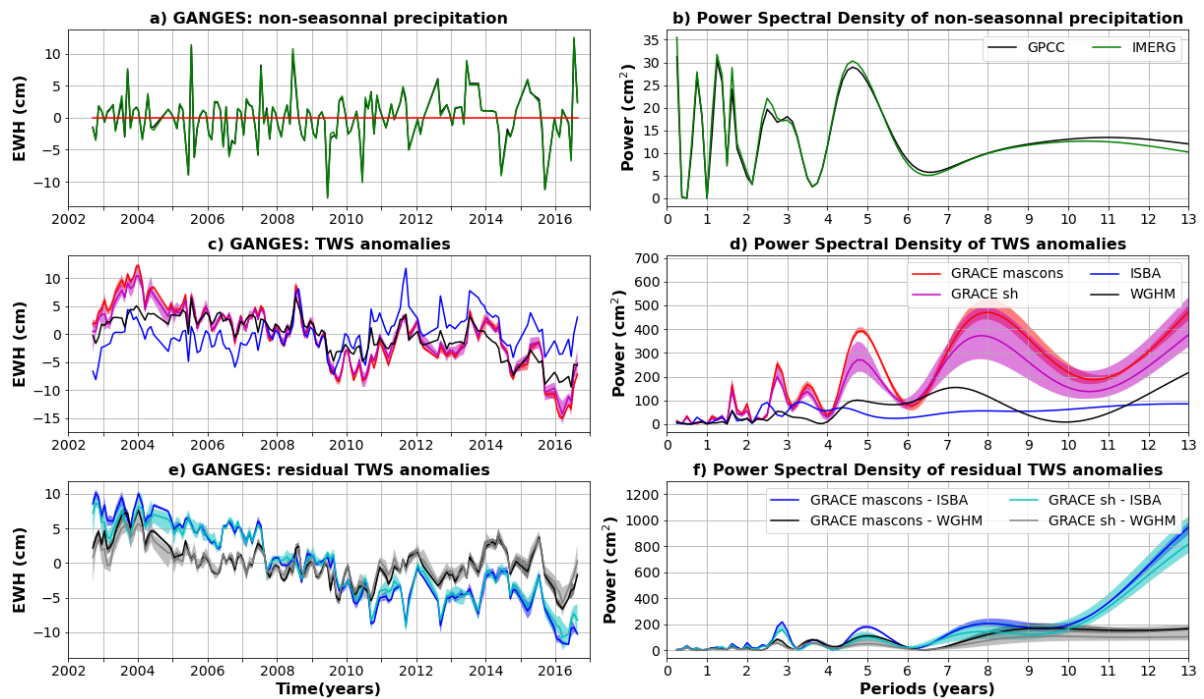


Figure S3.12: Same as S3.2 for the Ganges basin.

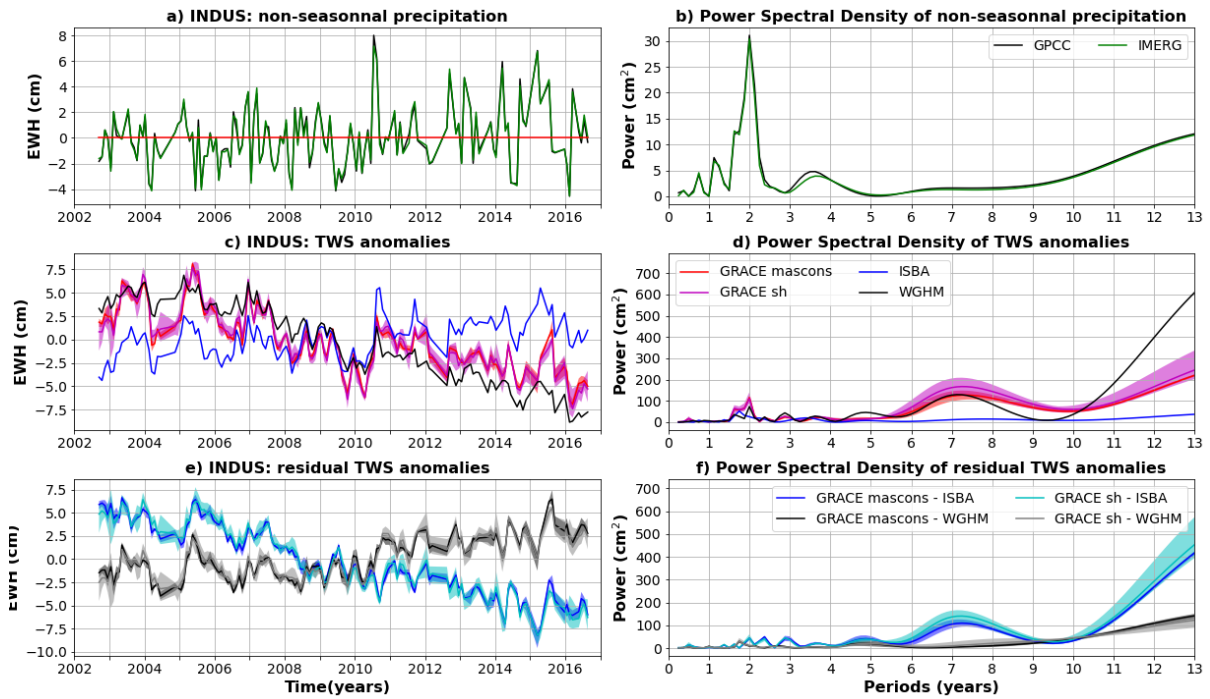


Figure S3.13: Same as S3.2 for the Indus basin.

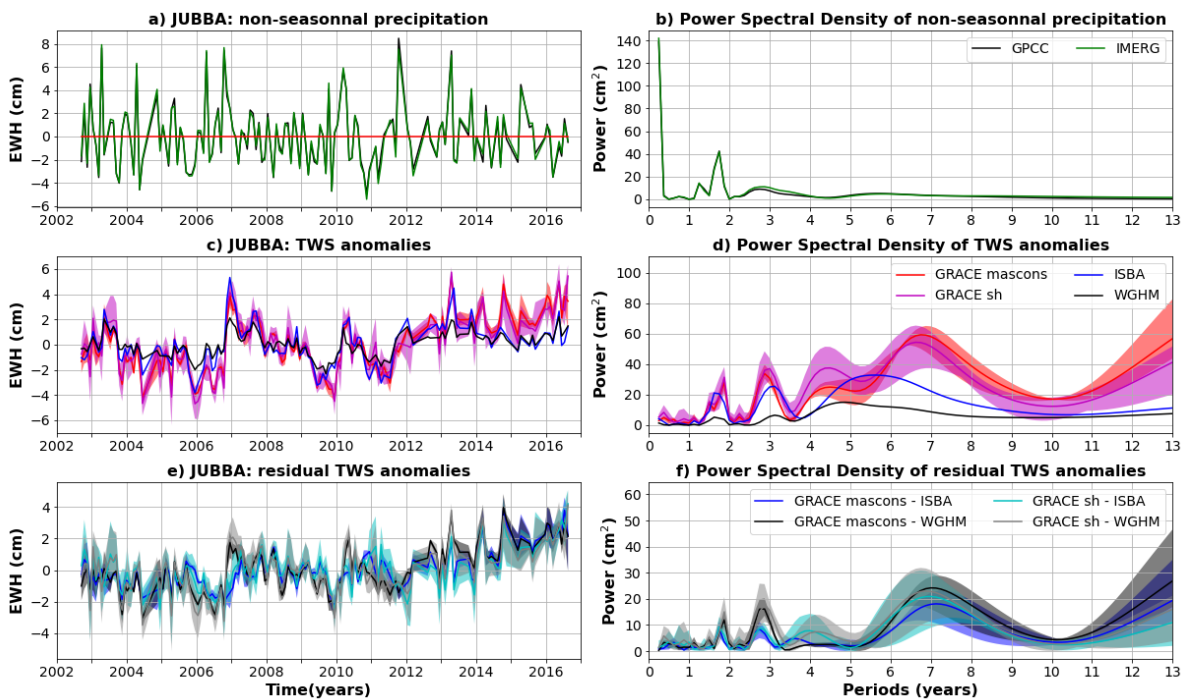


Figure S3.14: Same as S3.2 for the Jubba basin.

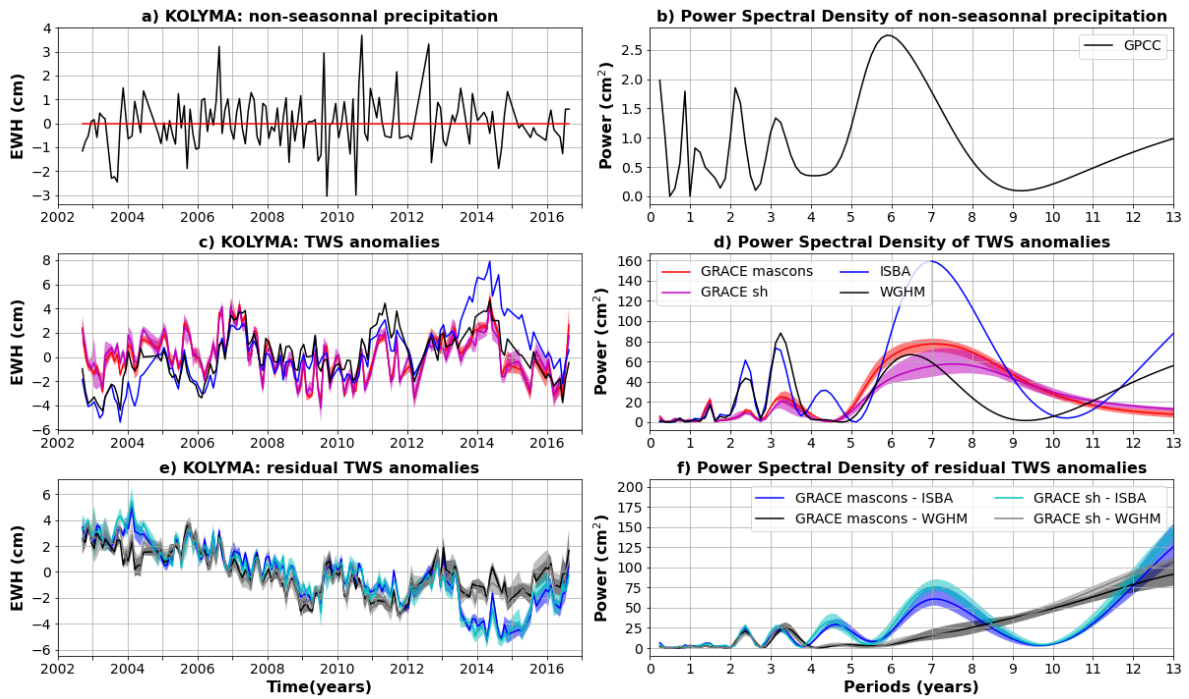


Figure S3.15: Same as S3.2 for the Kolyma basin. Non-seasonal precipitation anomalies are only estimated with GPCCC, as a significant part of the river basin is not covered by IMERG satellites due to its high latitude.

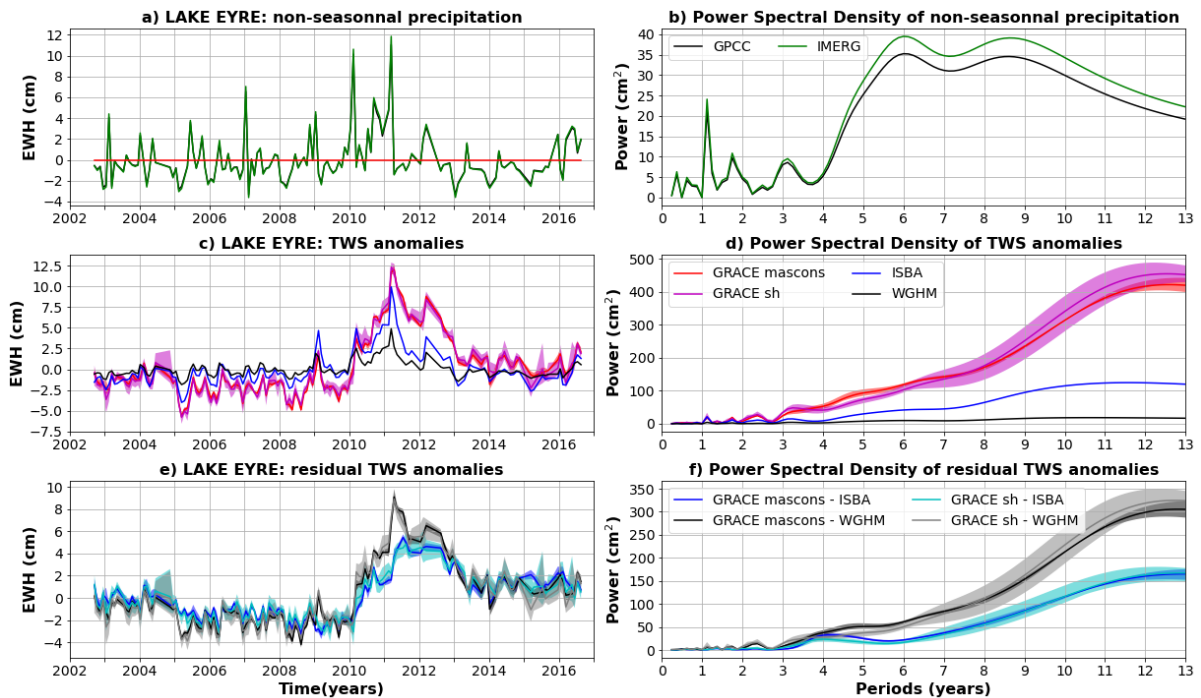


Figure S3.16: Same as S3.2 for the Lake Eyre basin.

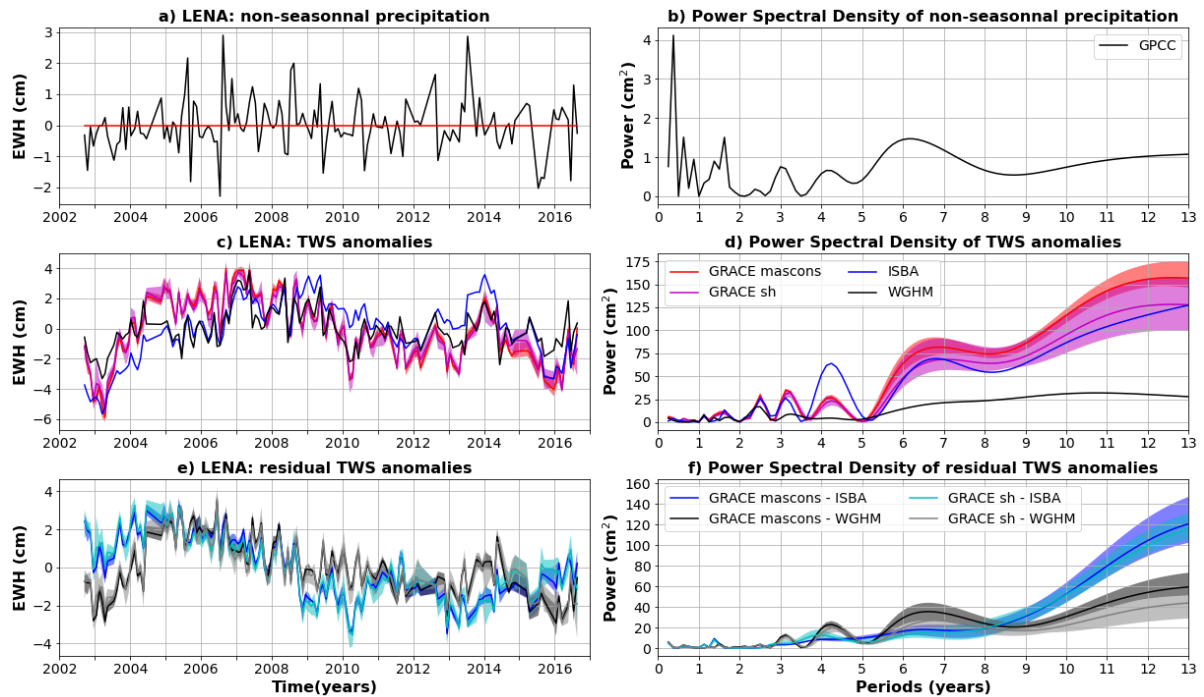


Figure S3.17: Same as S3.2 for the Lena basin. Non-seasonal precipitation anomalies are only estimated with GPCC, as a significant part of the river basin is not covered by IMERG satellites due to its high latitude.

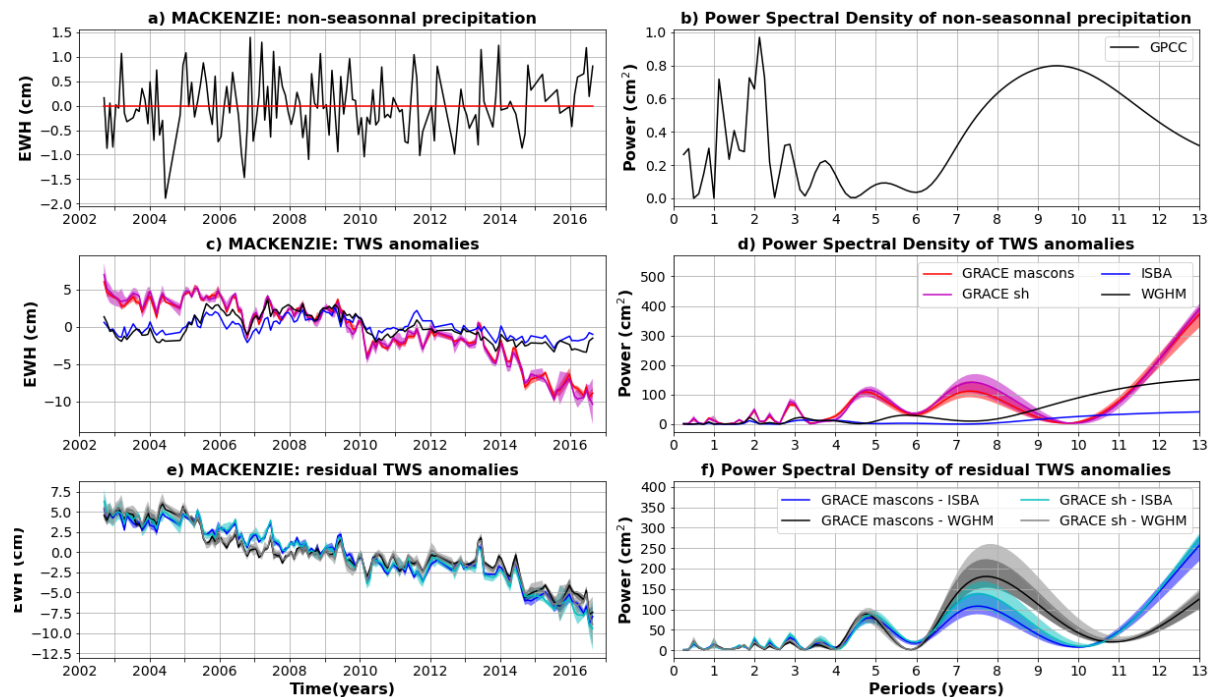


Figure S3.18: Same as S3.2 for the Mackenzie basin. Non-seasonal precipitation anomalies are only estimated with GPCC, as a significant part of the river basin is not covered by IMERG satellites due to its high latitude.

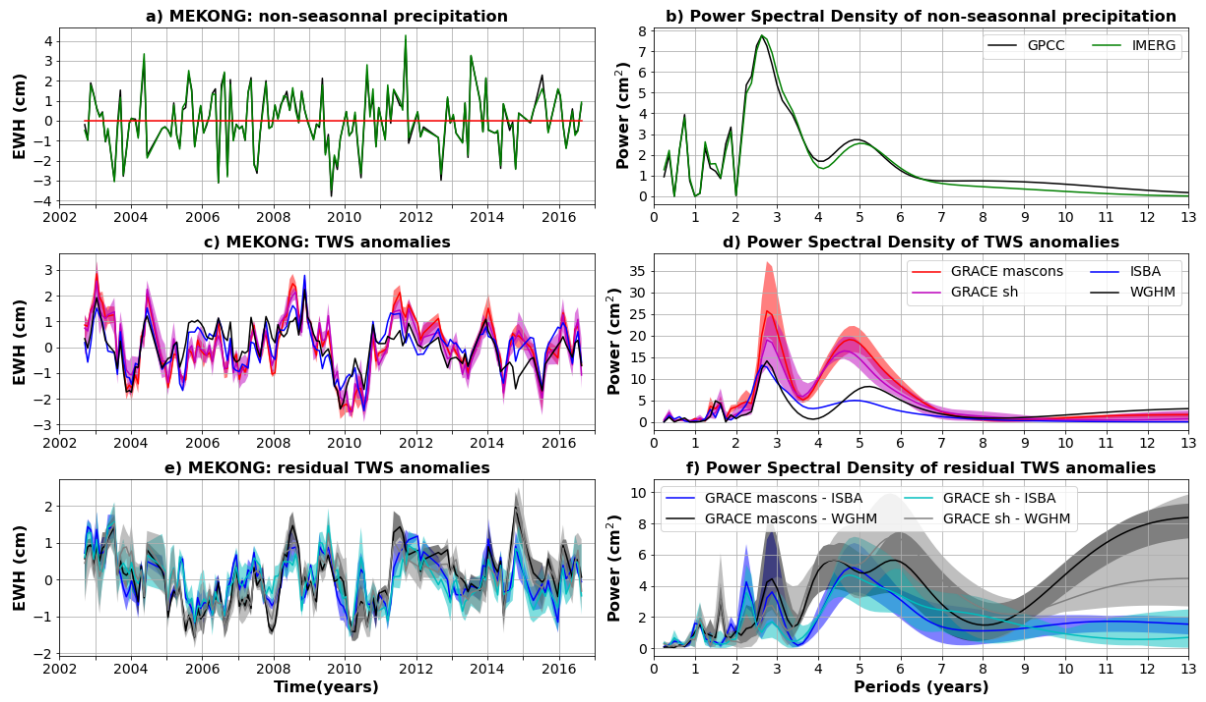


Figure S3.19: Same as S3.2 for the Mekong basin.



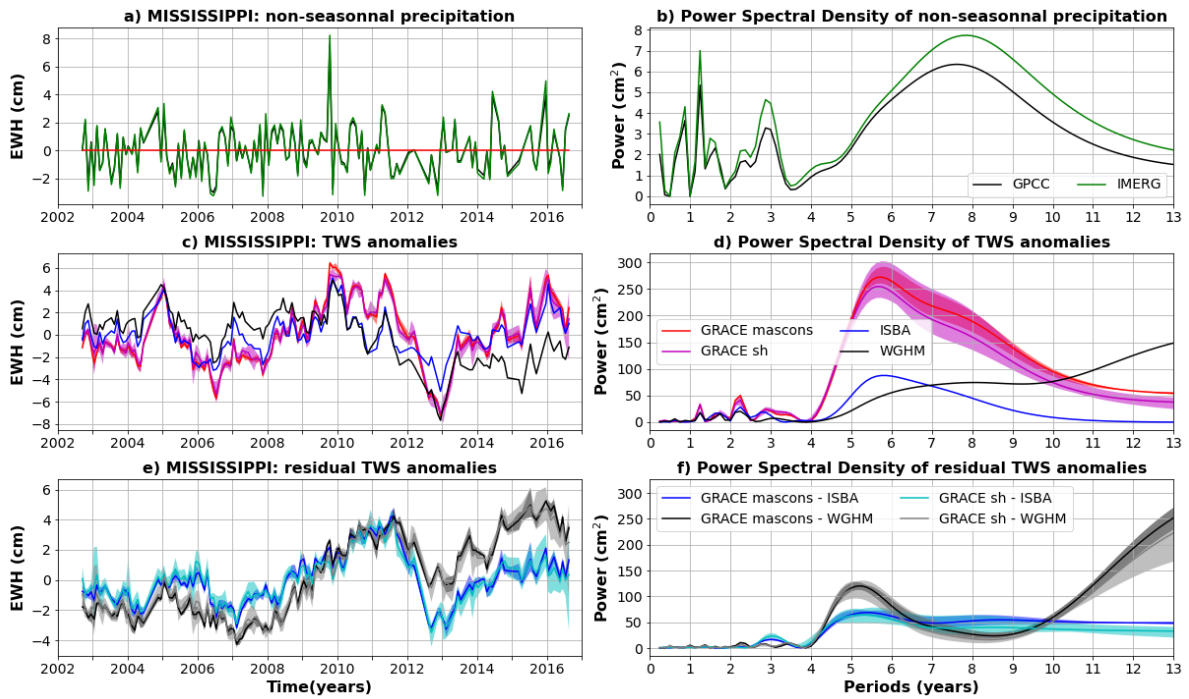


Figure S3.20: Same as S3.2 for the Mississippi basin.

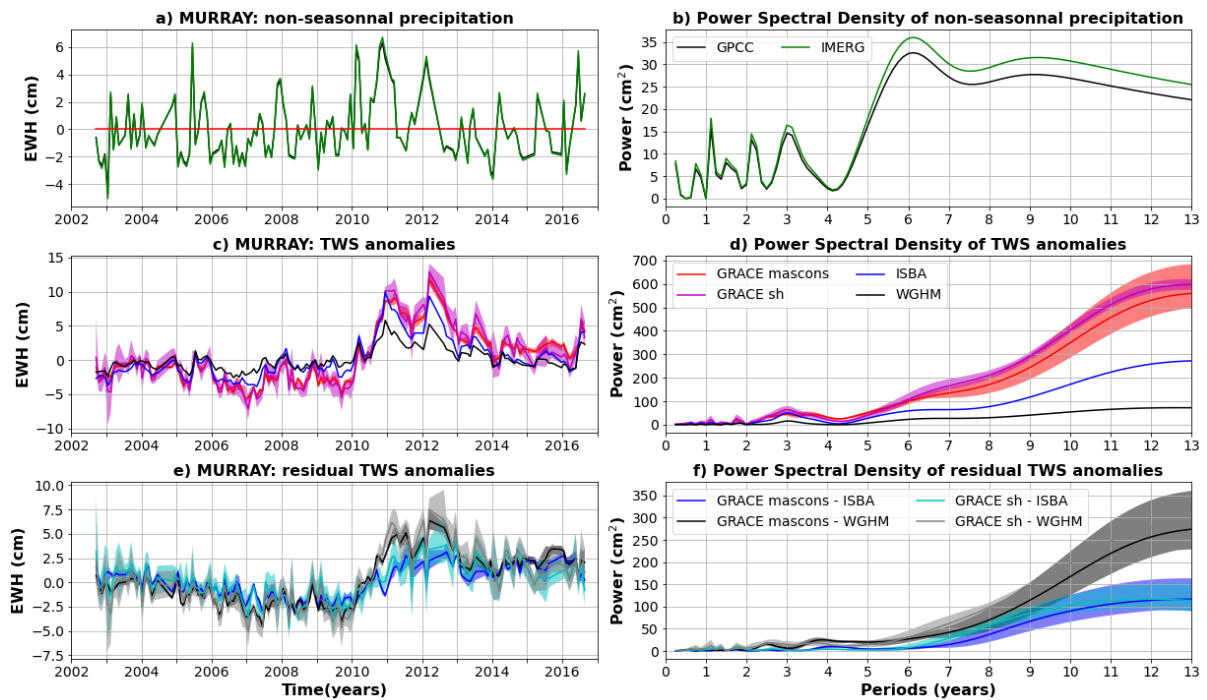


Figure S3.21: Same as S3.2 for the Murray basin.

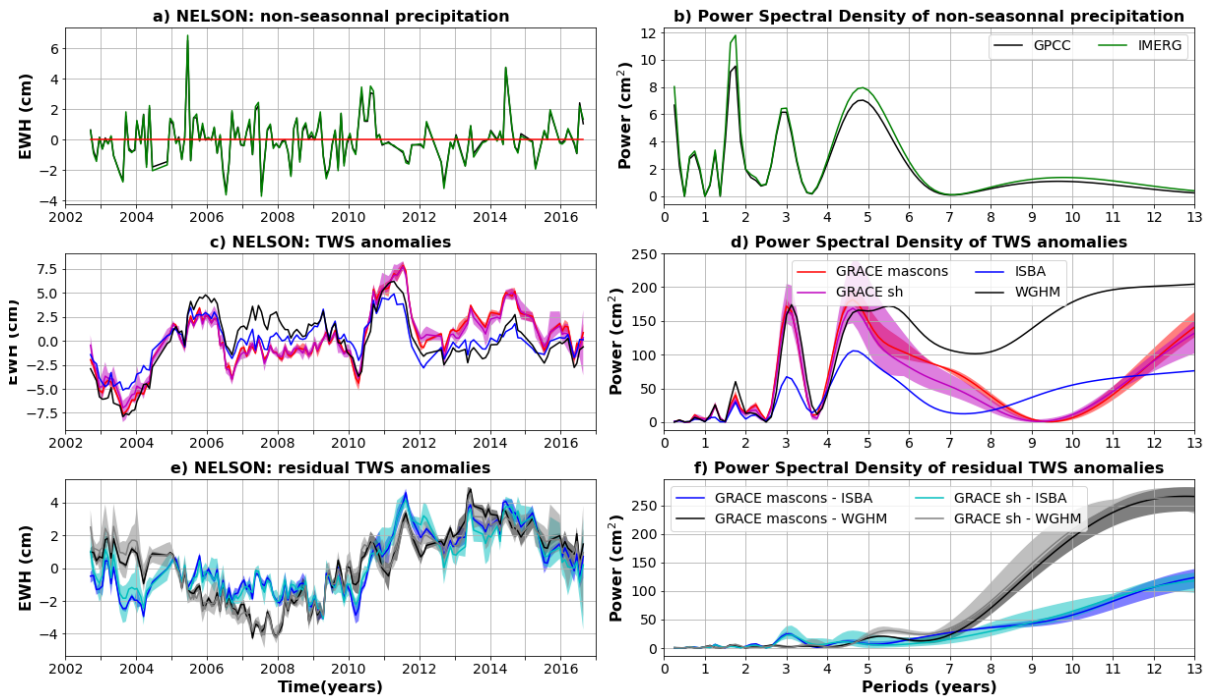


Figure S3.22: Same as S3.2 for the Nelson basin.

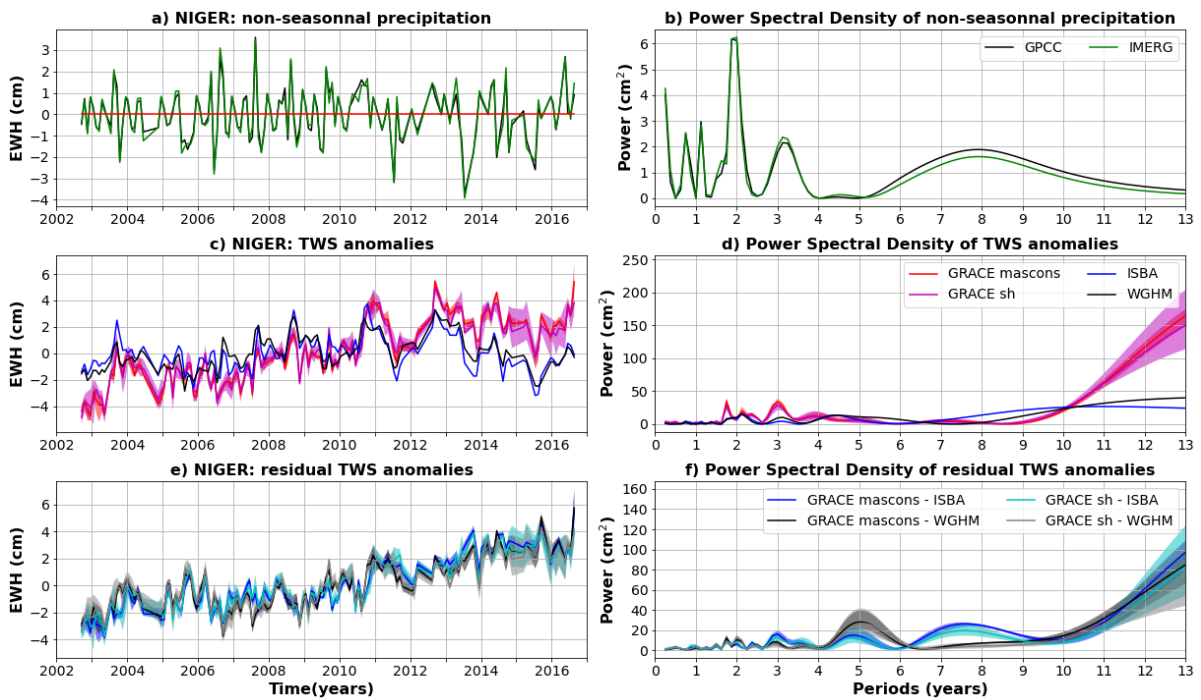


Figure S3.23: Same as S3.2 for the Niger basin.

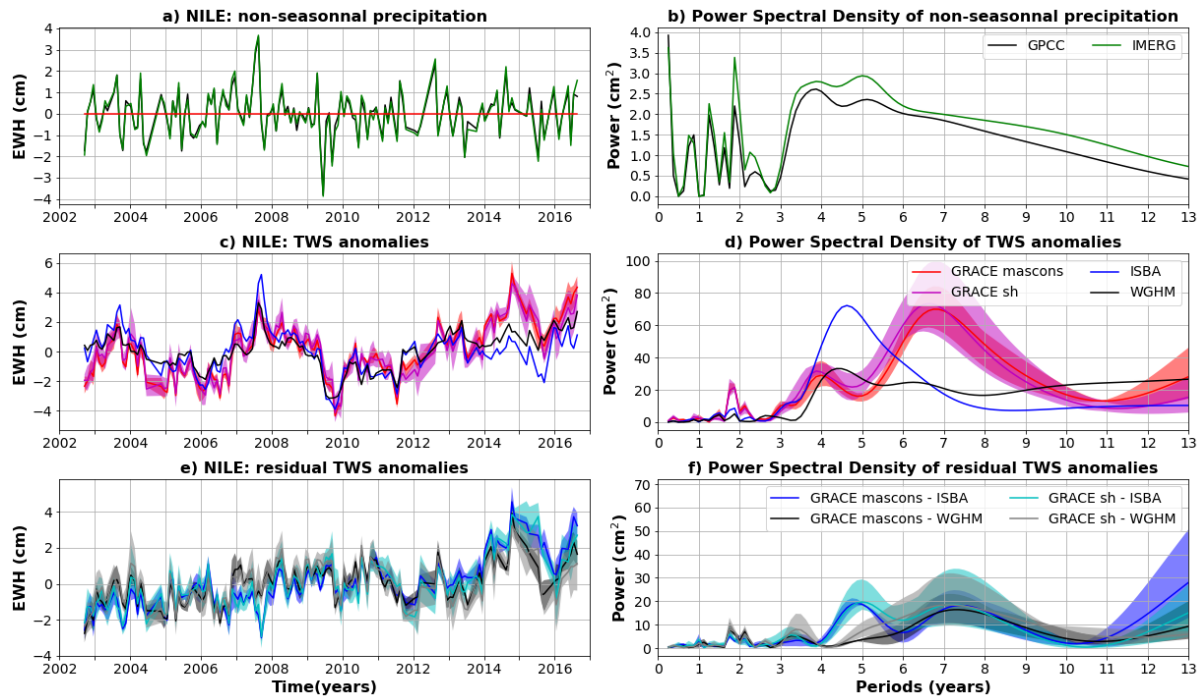


Figure S3.24: Same as S3.2 for the Nile basin.

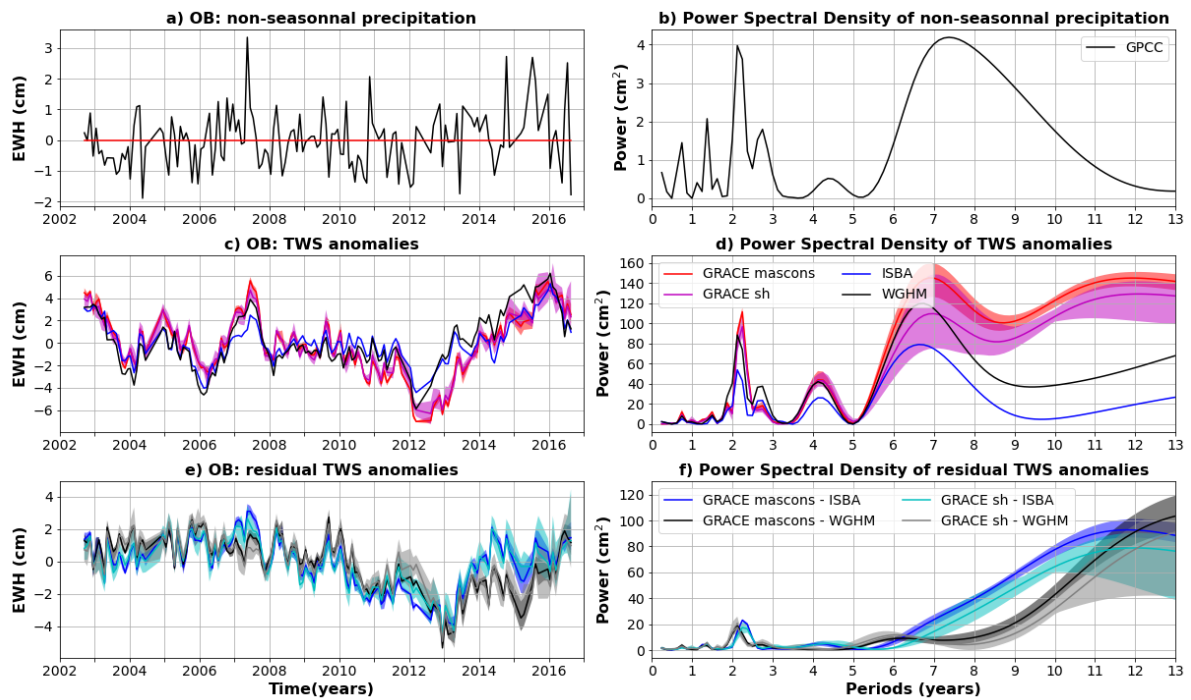


Figure S3.25: Same as S3.2 for the Ob basin. Non-seasonal precipitation anomalies are only estimated with GPCC, as a significant part of the river basin is not covered by IMERG satellites due to its high latitude.

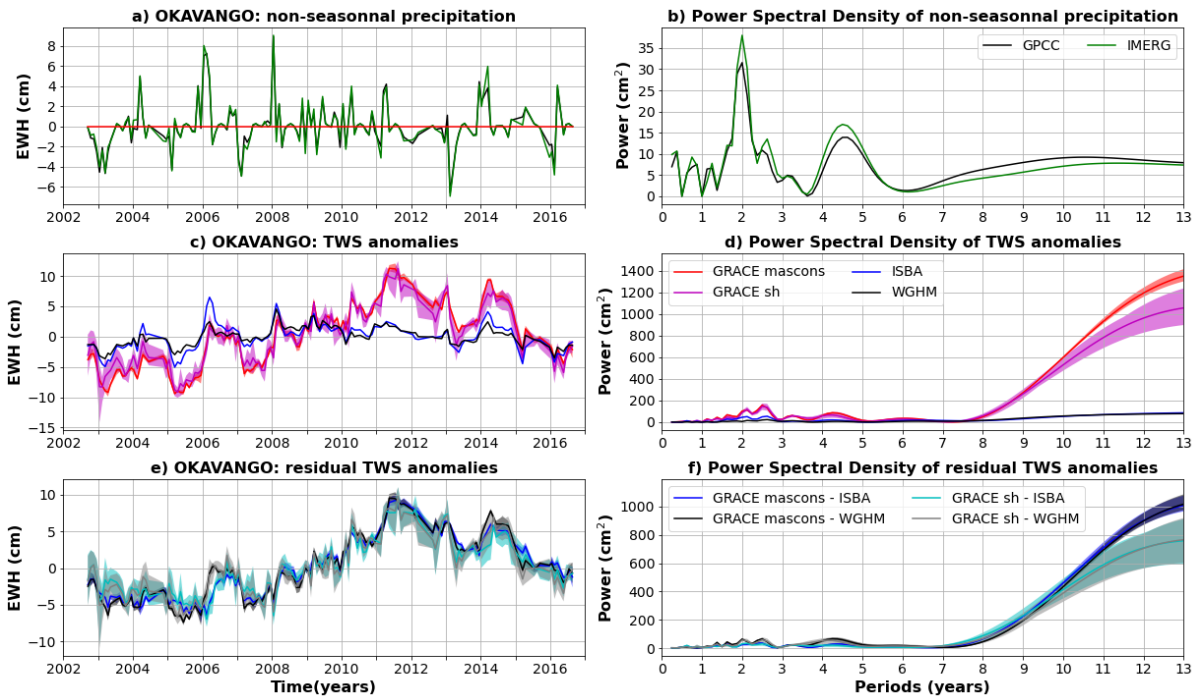


Figure S3.26: Same as S3.2 for the Okavango basin.

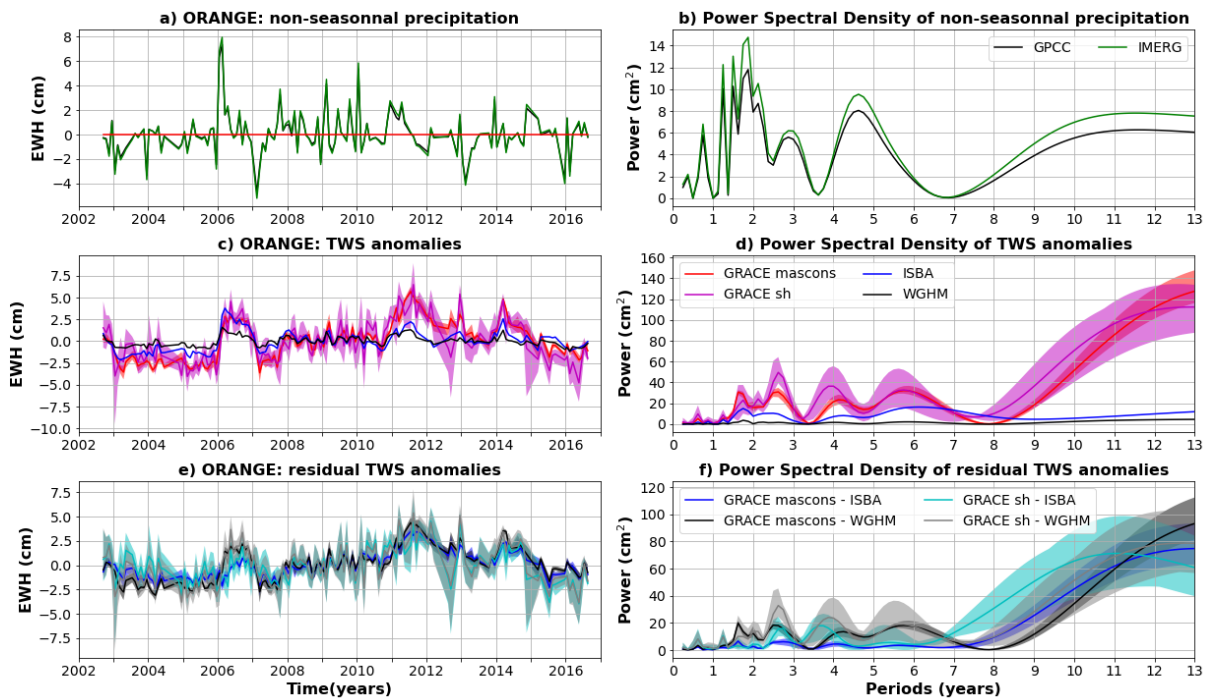


Figure S3.27: Same as S3.2 for the Orange basin.

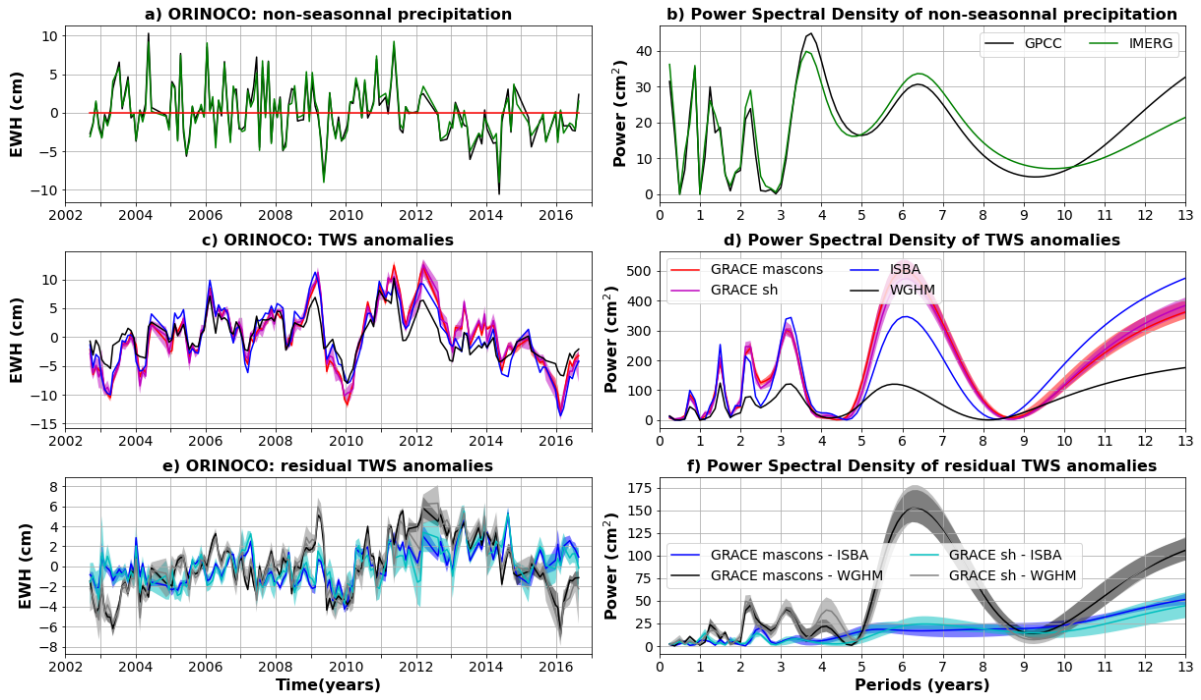


Figure S3.28: Same as S3.2 for the Orinoco basin.

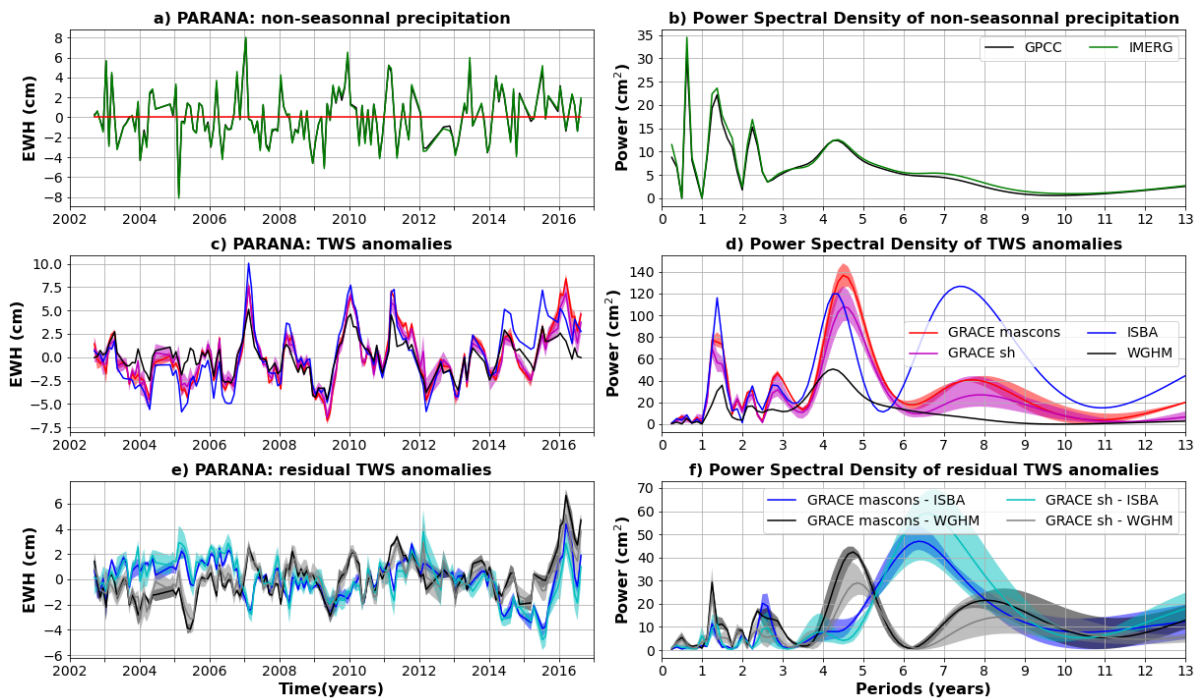


Figure S3.29: Same as S3.2 for the Parana basin.

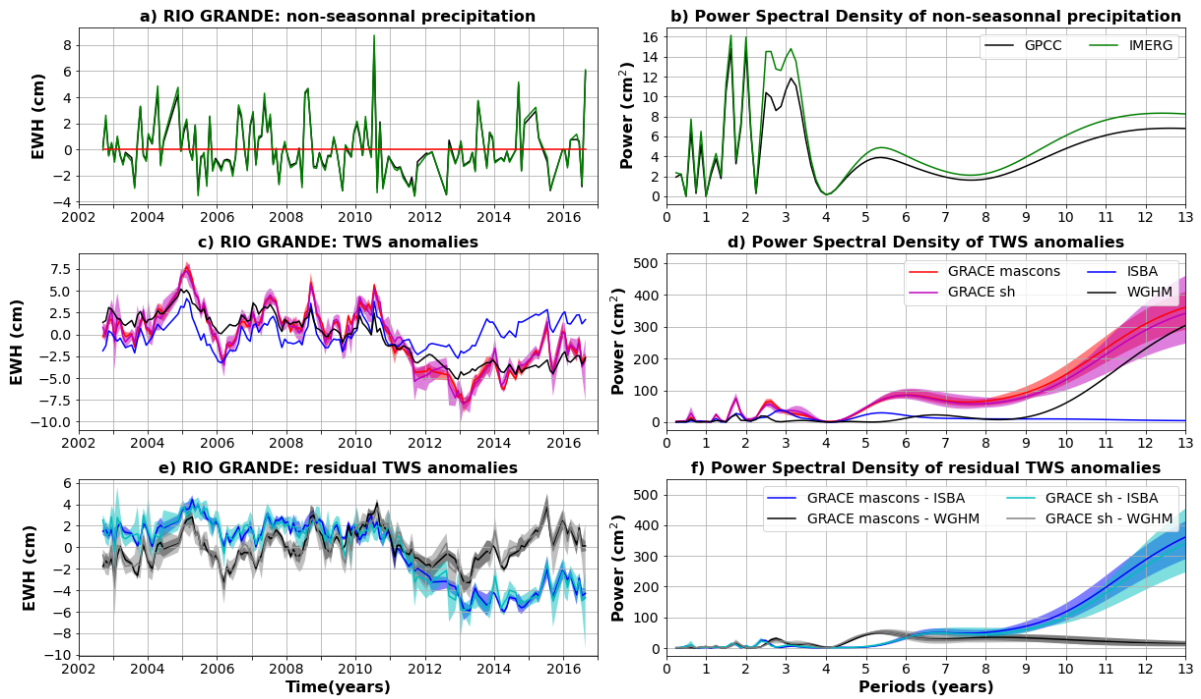


Figure S3.30: Same as S3.2 for the Rio Grande basin.

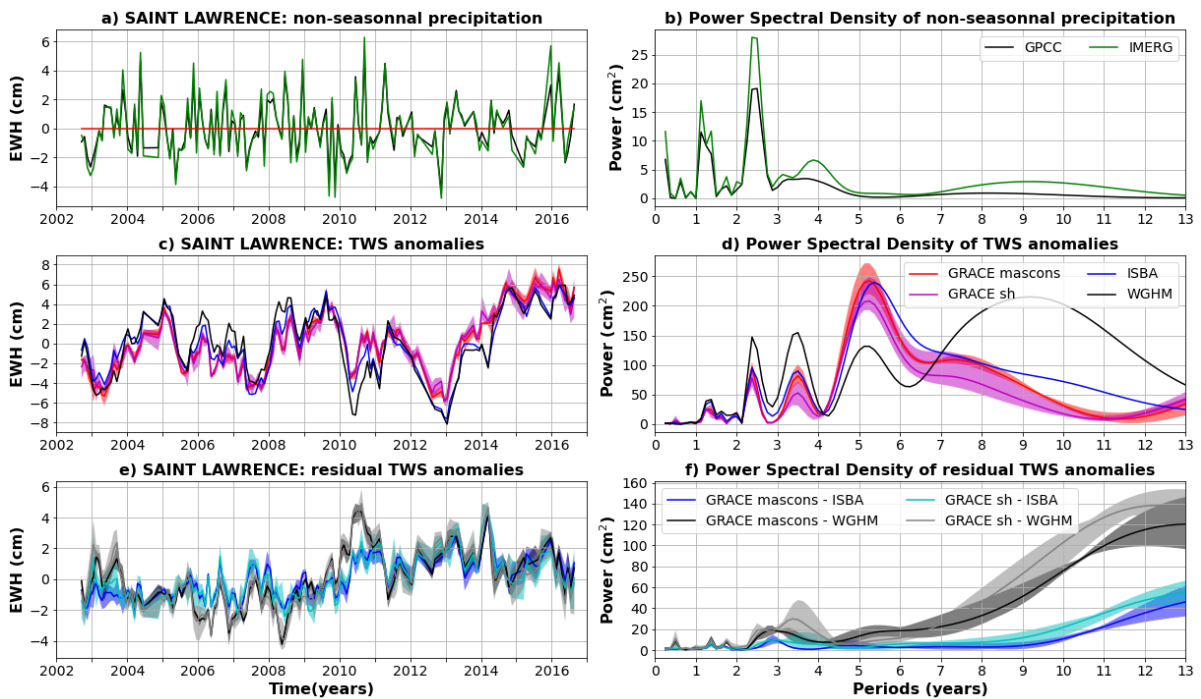


Figure S3.31: Same as S3.2 for the Saint Lawrence basin.

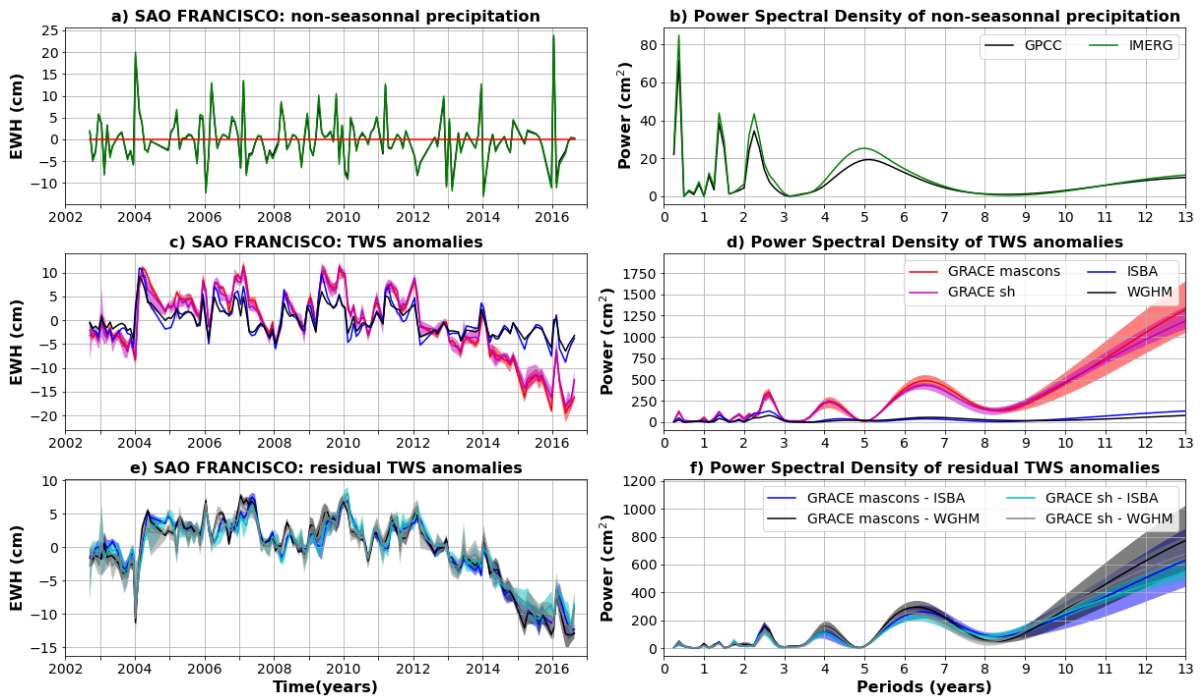


Figure S3.32: Same as S3.2 for the Sao Francisco basin.

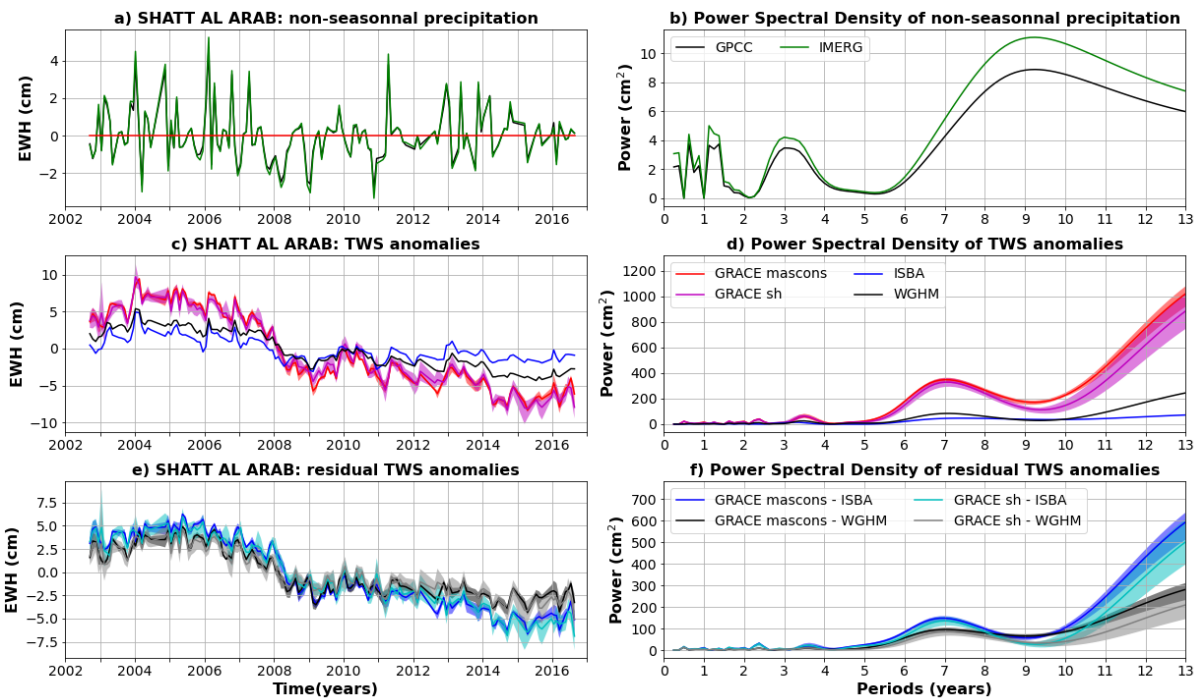


Figure S3.33: Same as S3.2 for the Shatt al Arab basin.

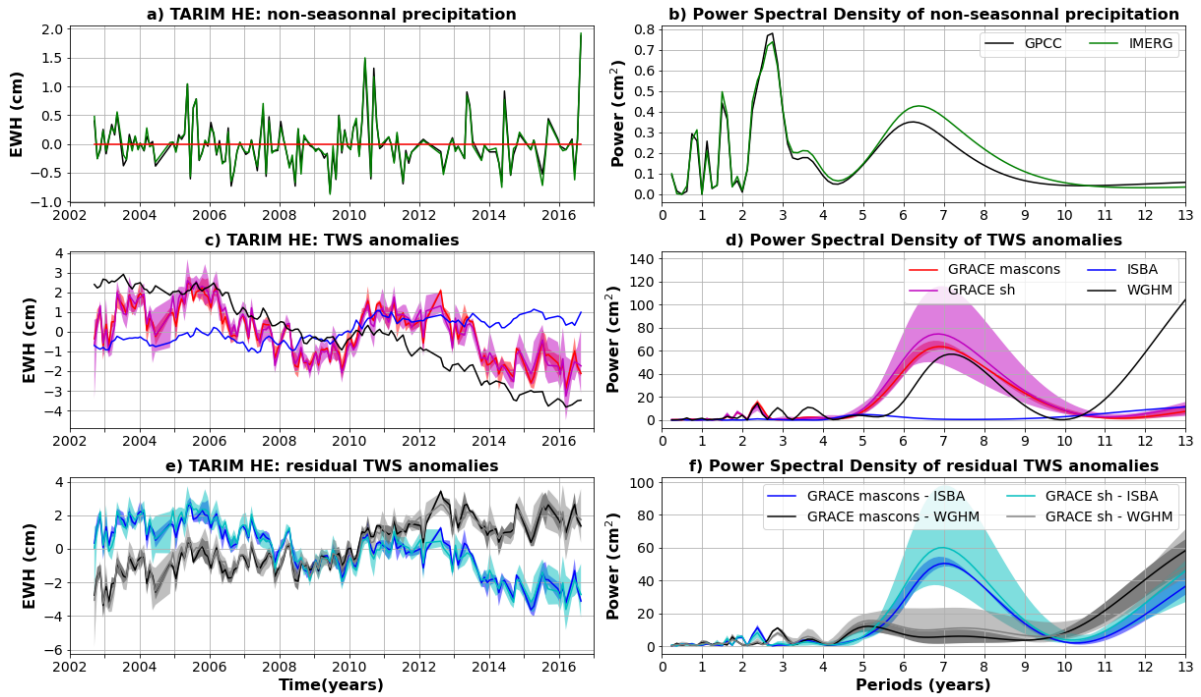


Figure S3.34: Same as S3.2 for the Tarim He basin.

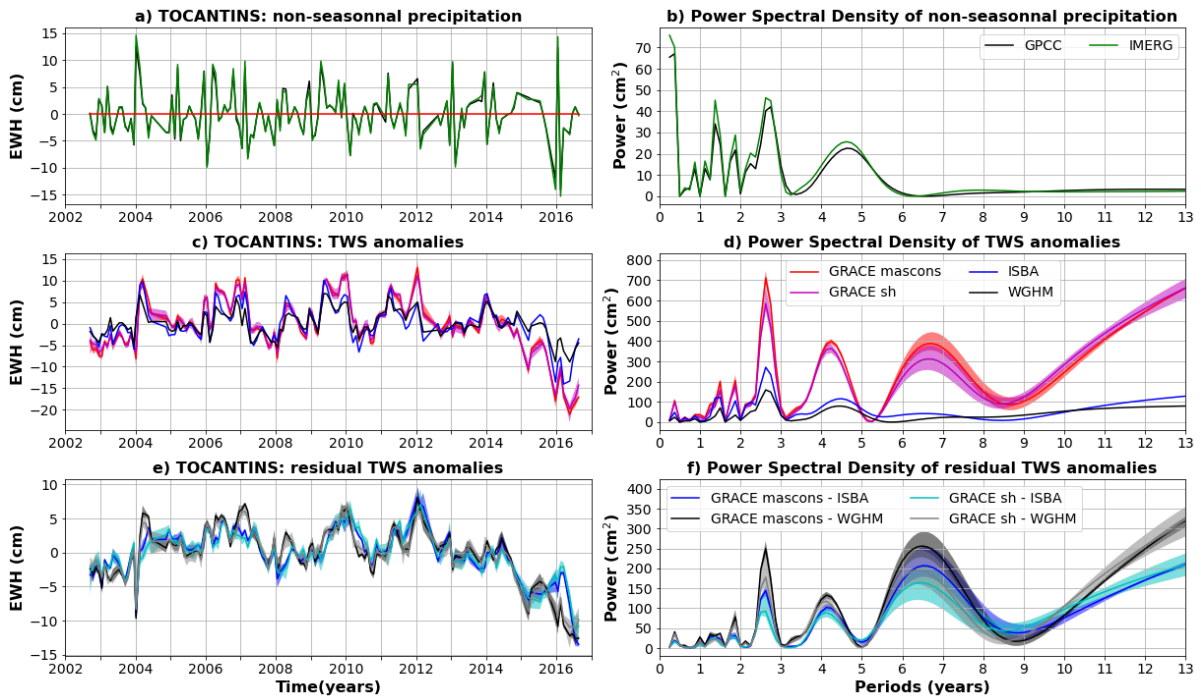


Figure S3.35: Same as S3.2 for the Tocantins basin.



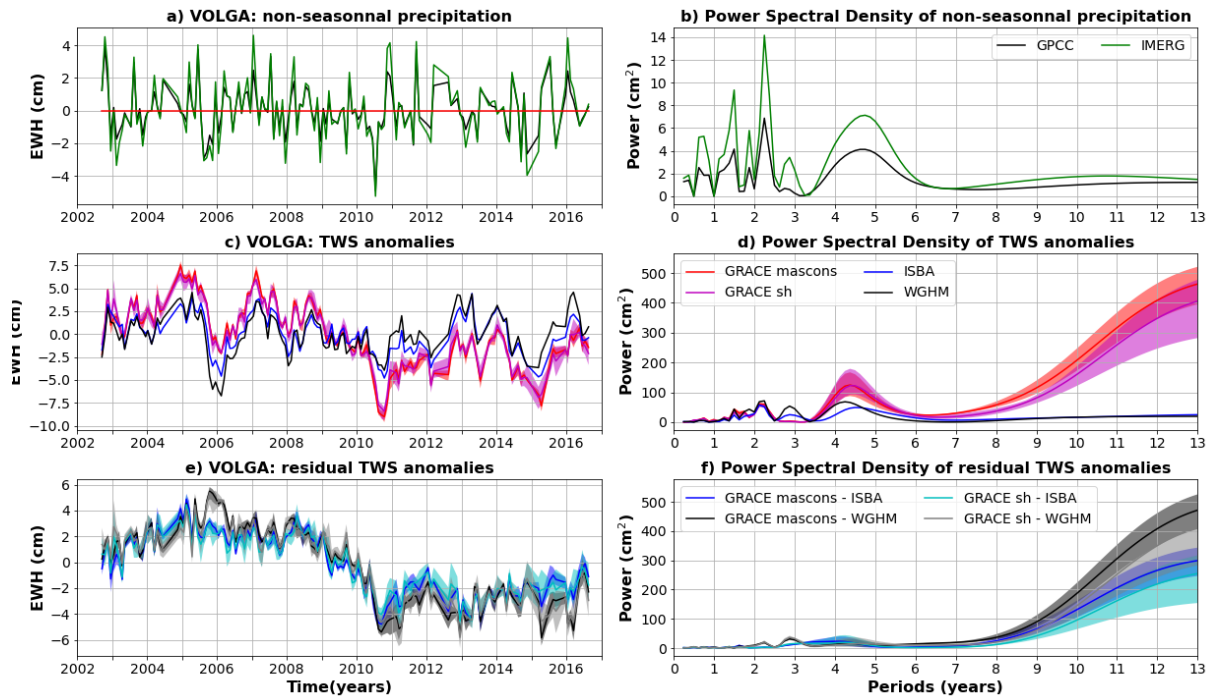


Figure S3.36: Same as S3.2 for the Volga basin.

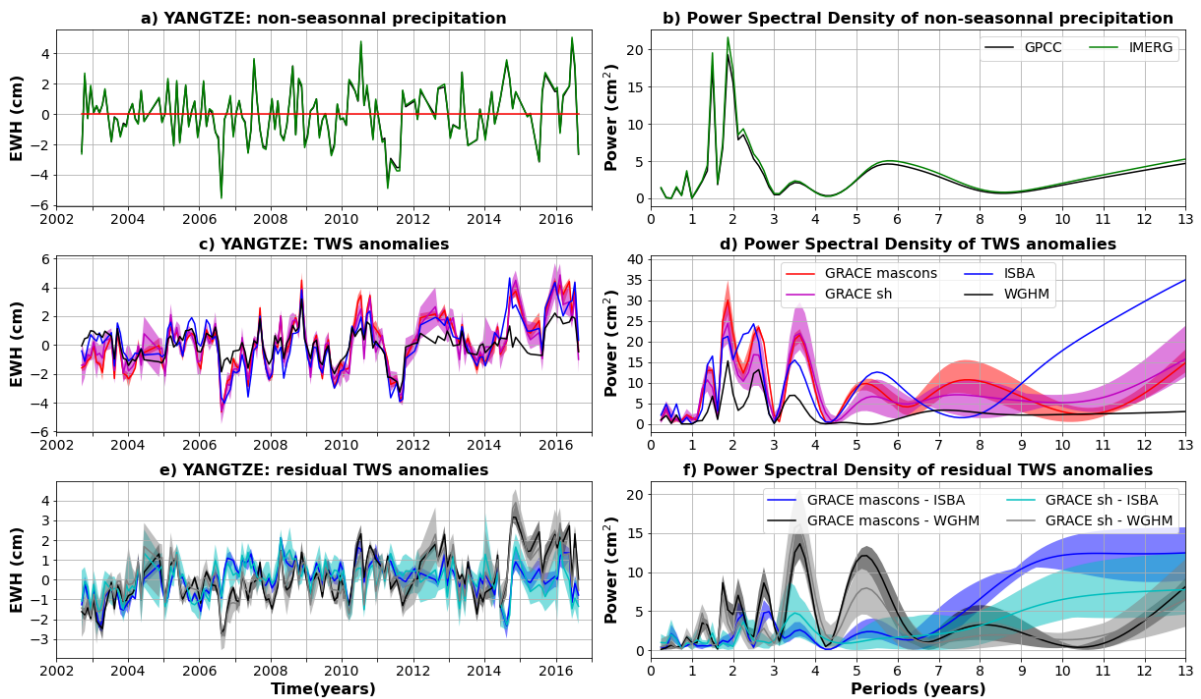


Figure S3.37: Same as S3.2 for the Yangtze basin.

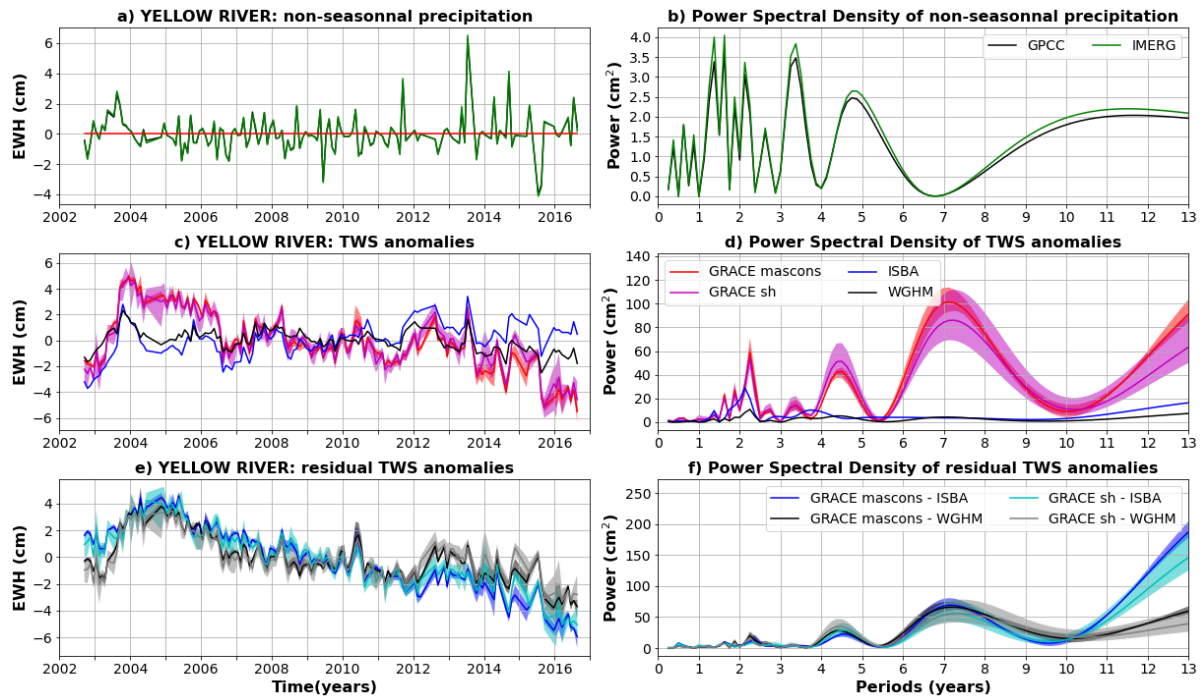


Figure S3.38: Same as S3.2 for the Yellow River basin.

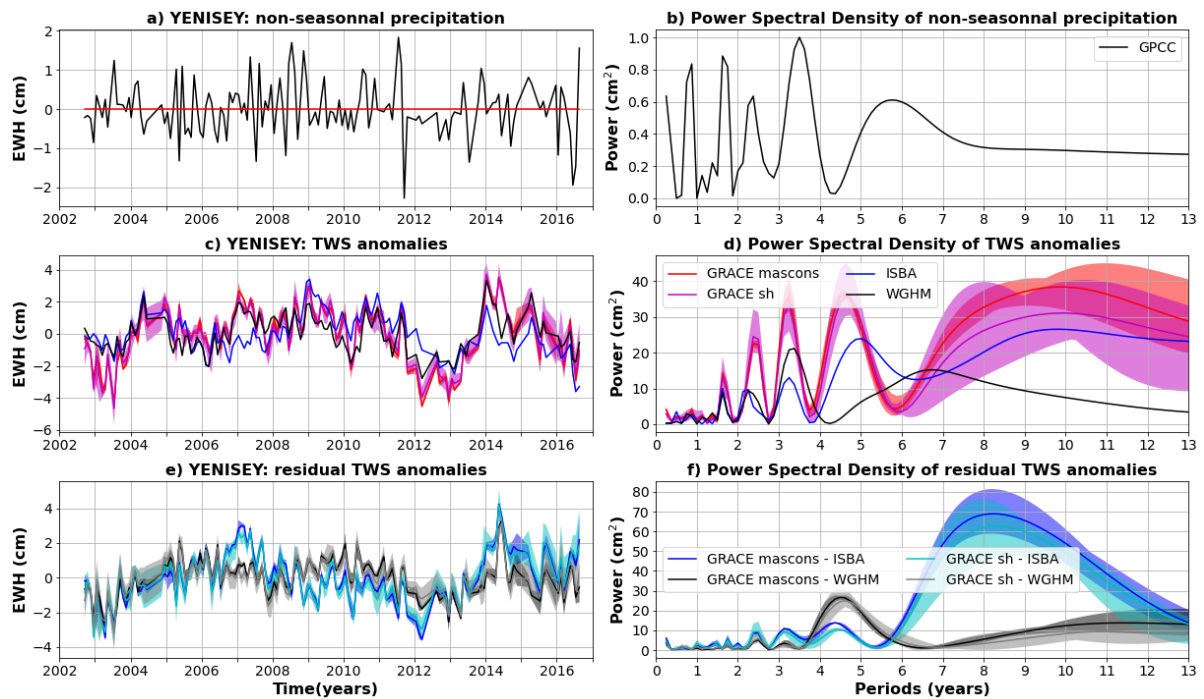


Figure S3.39: Same as S3.2 for the Yenisei basin. Non-seasonal precipitation anomalies are only estimated with GPC, as a significant part of the river basin is not covered by IMERG satellites due to its high latitude.

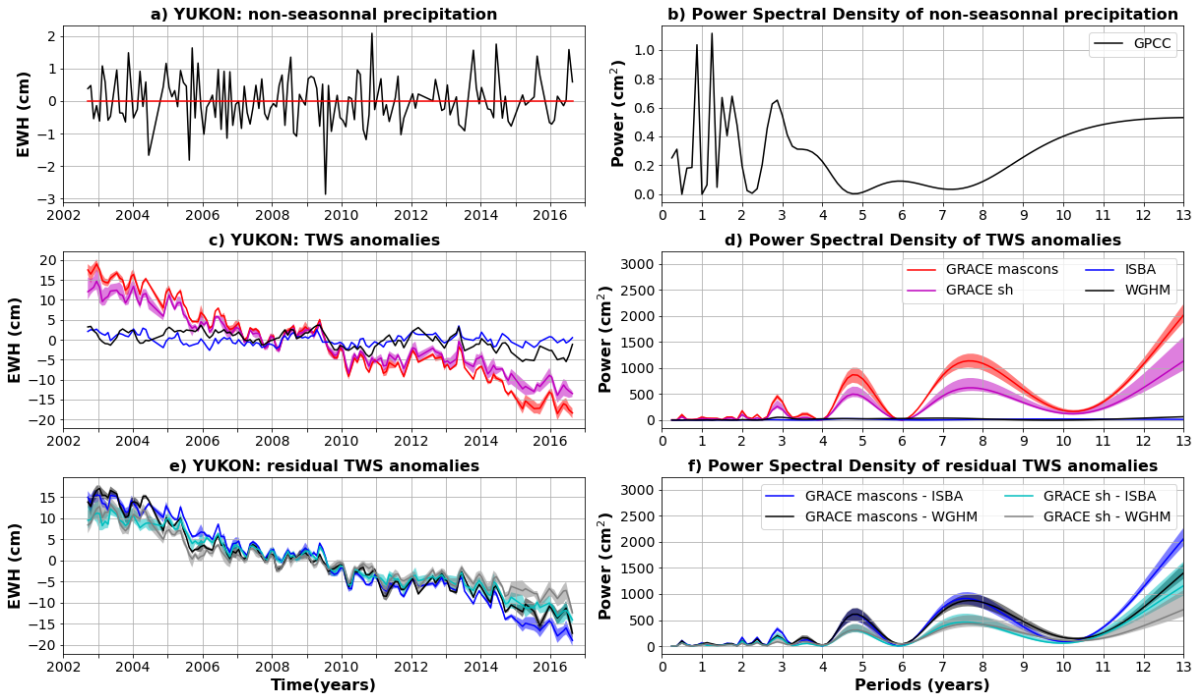


Figure S3.40: Same as S3.2 for the Yukon basin. Non-seasonal precipitation anomalies are only estimated with GPCCC, as a significant part of the river basin is not covered by IMERG satellites due to its high latitude.

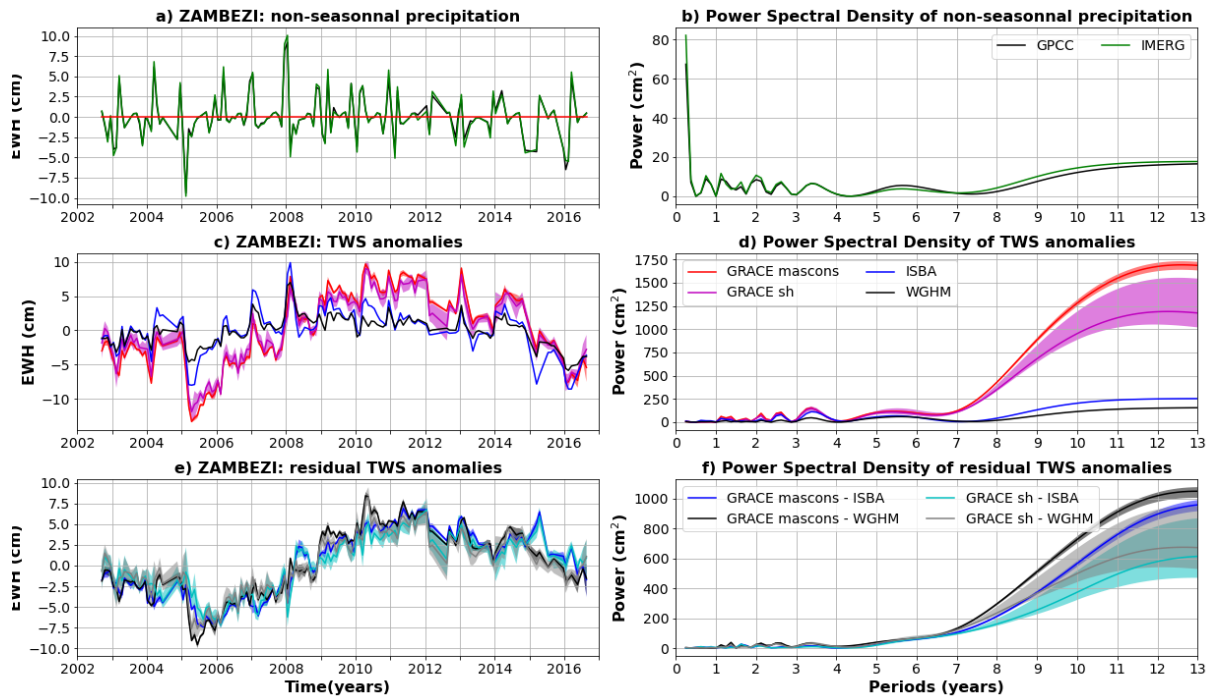
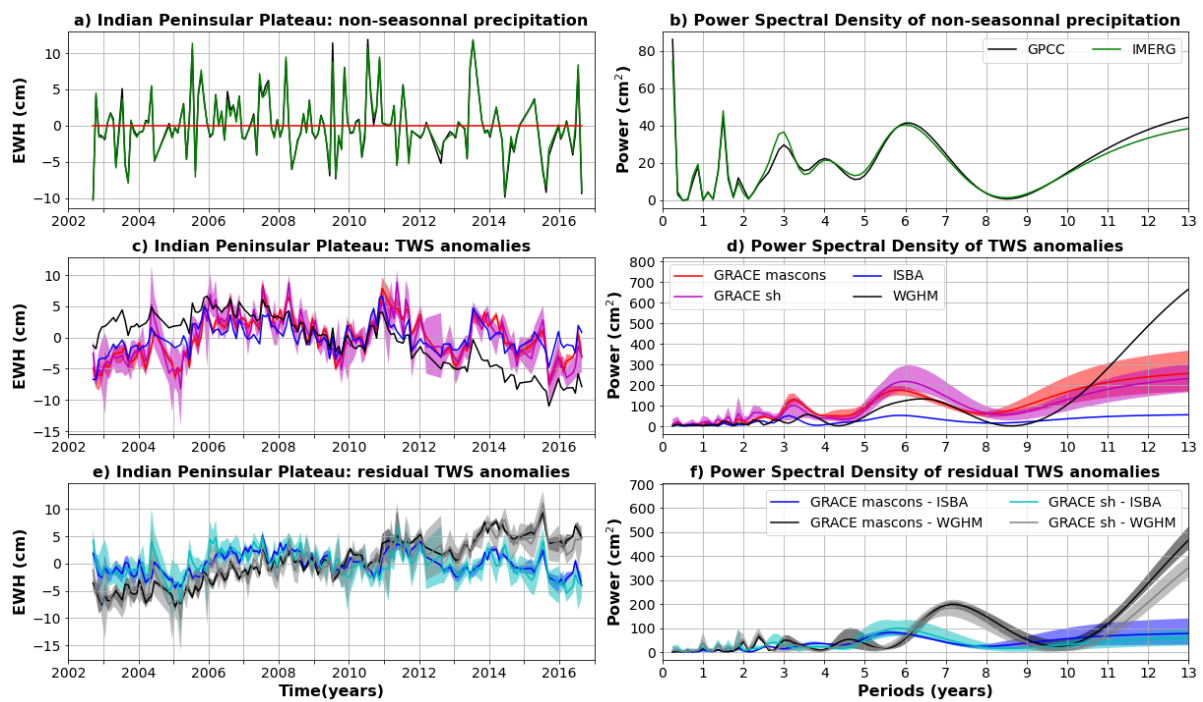


Figure S3.41: Same as S3.2 for the Zambezi basin

## Supplementary S4 Comparison of TWS anomalies from GRACE and global hydrological models over Southern India

TWS anomalies estimated from GRACE and global hydrological models have been averaged over Southern India and compared to in-situ and satellite precipitation (Fig. D1). The TWS anomalies captured with GRACE are well correlated with ISBA-CTRP ( $R=0.77$ ) and mildly correlated ( $R=0.47$ ) with WGHM predictions and precipitation ( $R=0.41$  with a lag of 1 month). A spurious negative trend is observed in WGHM prediction over 2006-2016 (Fig. D1c), likely due to overestimated groundwater abstractions. Better performances are reached with ISBA-CTRP, although anthropogenic contributions are neglected (Decharme et al., 2019).



**Figure S4.1 Comparison of TWS and precipitation anomalies averaged across the Indian Peninsular Plateau (latitudes 7 -23°N; longitudes 70-80°E).** a) Average precipitation anomalies for the GPCP (gauge-based) and IMERG (satellite-based) products. b) Power Spectral Density (PSD) of average precipitation anomalies. c) TWS anomalies average over the central Amazon for two global hydrological models (ISBA-CTRP in blue and WGHM in black) and 9 GRACE solutions (mascons in red, spherical harmonic in magenta). The solid line corresponds to the average of the sub-ensemble, the shaded area to the minimum to maximum envelope. d) PSD of the averaged TWS anomalies shown in (c). e) Residual TWS anomalies averaged over the central Amazon corridor and calculated as the difference between GRACE and ISBA-CTRP (blue when the difference is calculated with mascons, cyan with spherical harmonics) or WGHM (black when the difference is calculated with mascons, grey with spherical harmonics).