Supplementary Materials for

Nonstationarity in the global terrestrial water cycle and its interlinkages in the Anthropocene

Wanshu Nie[1], Sujay V. Kumar[2], Augusto Getirana[2,3], Long Zhao[4], Melissa L. Wrzesien[2,5], Goutam Konapala[2,3], Shahryar Khalique Ahmad[2,3], Kim A. Locke[2,3], Thomas R. Holmes[2], Matthew Rodell[2], Bryant D. Loomis[6]

[1]Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD, USA
[2]Hydrological Sciences Lab, NASA Goddard Space Flight Center (GSFC), Greenbelt, MD, USA
[3]Science Applications International Corporation, McLean, VA, USA
[4]Department of Analytics and Operations, National University of Singapore, Queenstown, Singapore
[5]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA
[6]Geodesy and Geophysics Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA

This file includes:

Section 1: Model performance evaluation

In all the comparisons below, the differences in the metric of interest between the model OL and DA are presented. The color schemes are designed such that cool colors represent improvements from DA (e.g., increases in MI, reductions in RMSE) and warm colors represent degradation from DA. We focus on MI as the metric of evaluation when the comparisons involve reference products derived from other models (e.g., ALEXI, FLUXSAT, GLEAM, ERA5land) or involve variables that are related (e.g., soil moisture, SIF), but that cannot be compared directly. When comparing with site-level measurements, we employ metrics such as RMSE that allows direct comparison of magnitudes.

Extended Data Fig. 10 (part 1) presents the impact of multivariate DA on ET estimates with FLUXCOM, ALEXI, GLEAM, and ERA5land as the reference data. All four comparisons indicate an overall positive impact from DA on ET, with prominent patterns of improvements over Europe, Midwest U.S., parts of South America, Africa, India, and Australia. Some patterns of degradation (e.g., western U.S., northern Africa, and southern parts of Argentina in the evaluations against FLUXCOM, and parts of the Sahel in the comparison against GLEAM) are also observed. Nevertheless, the overall positive impact of DA on ET is confirmed in these comparisons, even though these datasets employ different sources of information to develop ET estimates.

Extended Data Fig. 10 (part 2) presents an evaluation of the impact on modeled runoff fields from DA, against the GRUN and ERA5land datasets. While there is a domain-wide increase in MI in the ERA5land comparison, the model skill generally reduces with DA relative to the GRUN data, with an overall decrease in MI. However, the model simulations still show a closer agreement with the GRUN data in the RMSE evaluations, with larger improvements over eastern Europe, parts of India, South America, and Africa. The RMSE change map relative to GRUN also indicates degradations over parts of the Amazon and southern Africa. Note that GRUN data is derived from in-situ gauge discharge measurements, which includes impacts of reservoir operations. Some of the

degradations in runoff from DA is likely because the DA integrations do not include direct observational inputs to represent such processes.

The overall impact of DA on soil moisture fields is marginal, as indicated in Extended Data Fig. 10 (part 3). Over most locations, there is a neutral impact on surface soil moisture from multivariate DA. Interestingly, there is a general trend toward more beneficial impacts from DA on root zone soil moisture, particularly over the Midwest and western US, Europe, and Australia.

On the other hand, Extended Data Fig. 10 (part 4) indicates that there are systematic and major improvements in the simulated groundwater storage from DA. The MI increases significantly with DA over Australia, northern India, the western and central US, southern Brazil, and Chile. Some degradations are also observed over parts of the eastern US and the Scandinavian region of Europe. Note that many of these areas (e.g., India, western US, Brazil) with major improvements in groundwater storage simulations are documented to have significant groundwater pumping activities.

Finally, Extended Data Fig. 10 (part 5) presents the impacts on GPP from multivariate DA by comparing it to FLUXSAT and GOME-2 datasets. Similar to the patterns in ET, assimilation leads to significant improvements in GPP in most parts of the world. The MI is significantly improved over in most parts of the world along with a reduction in RMSE. Similar patterns of major improvements over the Midwest US, Amazon, Africa, India, Europe, Australia, and parts of China in GPP are also observed in the GOME-2 comparisons. The significant improvements, particularly over regions such as the Midwest US, India, and eastern China observed in these comparisons are likely a result of DA capturing the impacts of more intense agricultural activities.

Section 2: A Case Study (Region 18)

By cross comparing the nonstationarity of different variables, we find that there is significant variability in the dominant factor that contributes to the overall nonstationarity. The relative contribution of the nonstationarity constituent factors across variables also

differs. These findings suggest that human interventions on the water cycle process impact the transformation of the type of dominant nonstationarity from driving variables such as precipitation to storage terms such as TWS. For instance, the hotspot in Northeastern China (region 18, Fig. 3e) is estimated to have an average TWS depletion of -115 mm/yr, with 64% of the land occupied by trend as the highest rank contributing to the nonstationarity index. Within the region, 62% of the area is identified as having abrupt changes in TWS, and the majority of them (~88%) occur at or after 2010. However, the extreme frequency ratio for this region does not rank as high as the trend in the global context. Only 11% of the area under abrupt change is dominated by the extreme frequency ratio contributing to nonstationarity, all of which implies an enhanced extreme occurrence after the change point. For this region, the large decline in TWS is the dominant signal for nonstationary water cycle changes, which result from groundwater pumping for both agricultural and domestic needs. Agricultural irrigation has led to widespread cropland greening for the region[1], with an averaged positive trend of 1.7 g/mo/yr in GPP. However, the fact that the depletion of TWS is associated with degraded water cycle variability after the change point indicates that groundwater recharge is negatively affected when the groundwater level declines to a certain point. Unsustainable water use has led to shifts in both mean and variability in TWS and runoff, which may result in a groundwater crisis in terms of both water quality and water supply and increase the vulnerability to sea-level-rise along the coastal areas[2]. The high nonstationarity in TWS also propagates into runoff by altering its variability across scales, leading to a different attribution distribution for runoff nonstationarity. Areas with seasonal variation weighting the highest occupy 49%, followed by 38% contributed by dampened extreme frequency, and only 13% by long-term trend. Northern Tunisia (region 9), Middle East (region 13), and the surrounding area of the Caspian and Aral Sea (region 12) experience similar impacts of human water footprint on the transformation of nonstationary hydrological responses. In these areas, although the region averaged GPP trend is insignificant, localized nonstationary water storage change (primarily due to depletion) is associated with increasing GPP trends as indicated in the pie charts (Fig. 3c, f), leading to strong seasonal and extreme shifts in runoff.

Section 3: Relationship between trends and changes in extremes

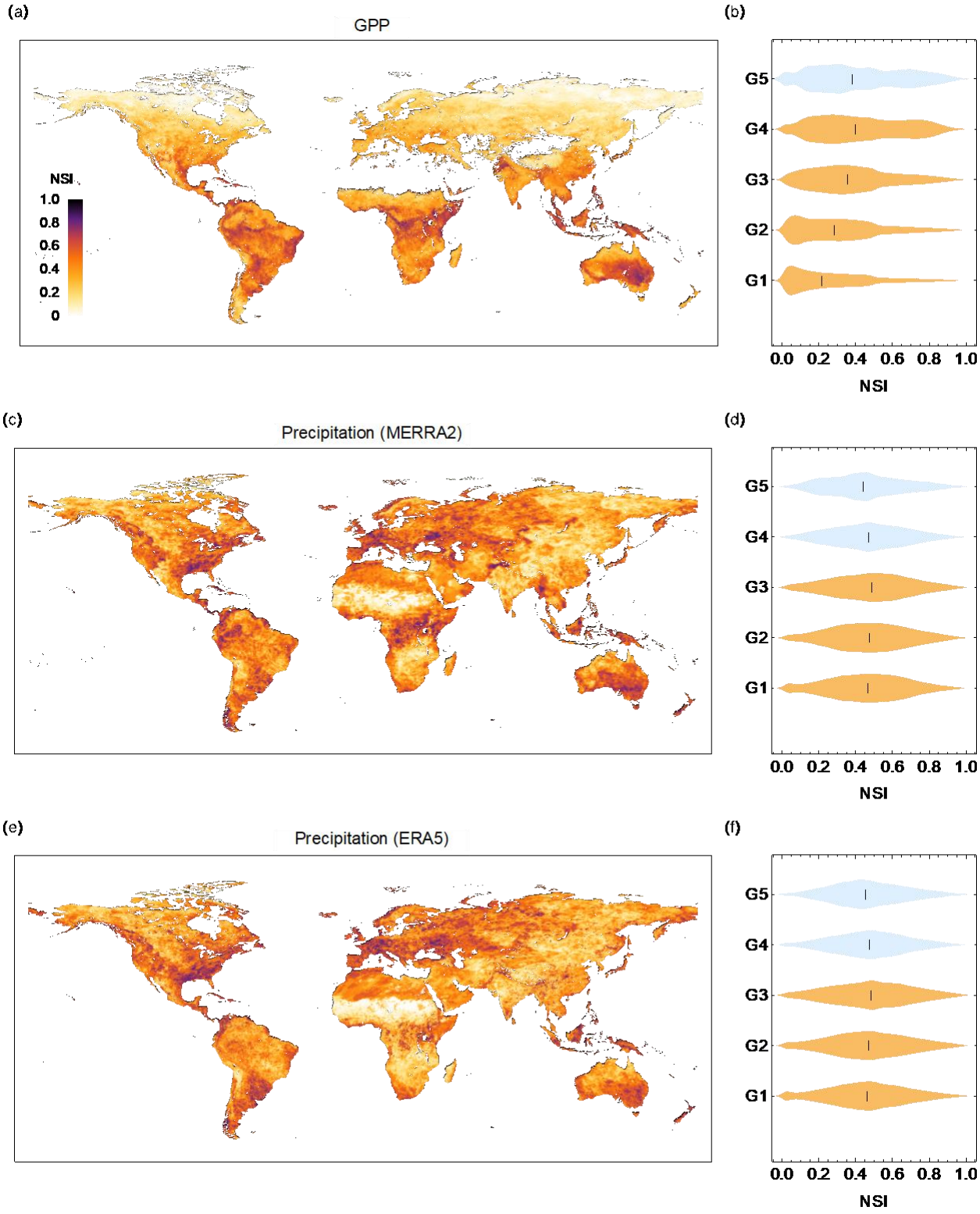
Prior studies suggest that for some types of extremes (e.g., heatwave and precipitation extremes), there are strong links between the increased number of extremes and the observed trend[3,4]. Here, we first explore the relationship between long-term TWS trends and the abundance of abrupt changes associated with extremes, followed by examining the spatial coherence between trends and the intensity of changes in extremes for all variables. Because the hotspot regions are identified based on high NSI values for TWS, investigating the relationship between the trend and extreme frequency ratio based on the hotspot regions could likely introduce an element of confirmation bias, as there is a tendency to see a positive relationship between the TWS trend percentile and EFR percentile if only considering the hotspot regions. We therefore bin the TWS trend percentiles over the globe into twenty equal-width bins and examine the mean and interquartile range of the EFR percentiles for each bin to see whether there is evidence linking a stronger trend with increased area under the change point and enhanced extreme frequencies. As illustrated in Fig. 4 (a-e), approximately 40% or more of the areas are detected with abrupt changes in extremes for TWS and runoff within each bin. In contrast, there is much less area experiencing abrupt changes in extremes for ET, GPP, and precipitation. Notably, for TWS and Runoff, we find a nonlinear increase in the percentage of areas identified with extreme related change points with the rank of the TWS trend, and these areas are more associated with negative trends as the trend percentile increases. The percentage of area with a change point in the last bin, equivalent to the area under the 5% of the greatest TWS trends, has increased to 61% and 77% for TWS and runoff, respectively (Fig. 4a, c).

Note that as the ranks of the negative TWS trends increase, the mean of the extreme frequency ratio percentiles increases nonlinearly (Fig. 4f-j). The increase in the percentile of TWS extremes is gradual until its trend exceeds the upper quartile (Fig. 4f). Interestingly, for places with strong positive TWS trends, the changes in TWS extreme frequency actually drop. However, for runoff, the strong increase in extreme changes starts even when the trend in TWS is not that large and holds true for places under both

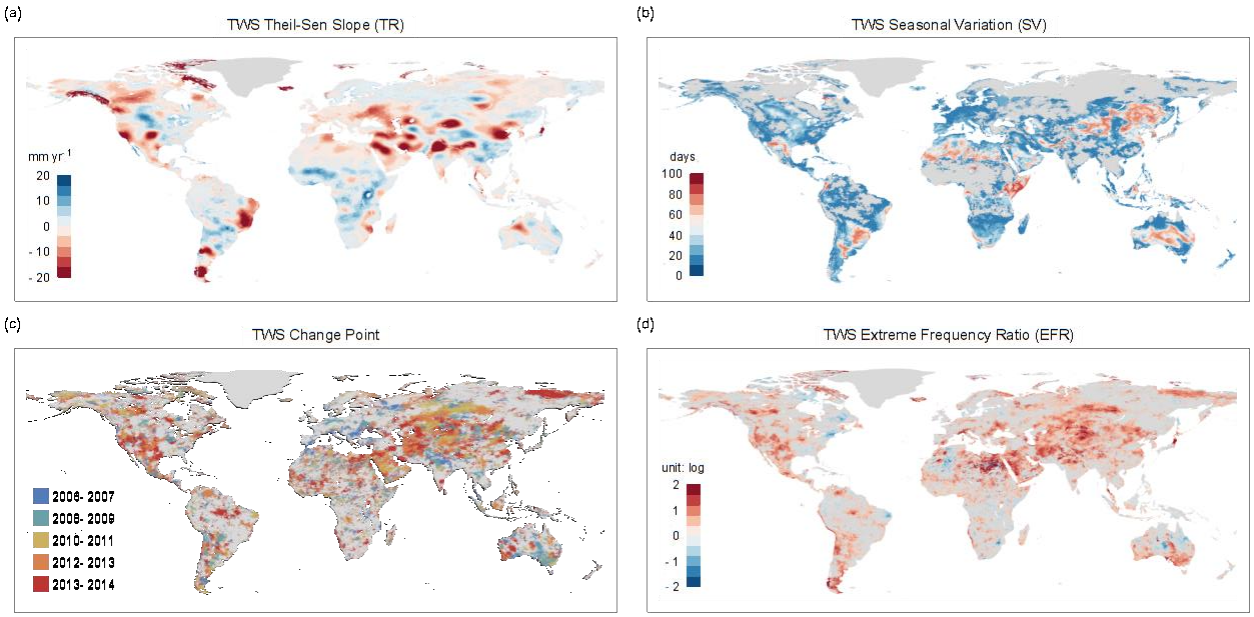
positive and negative TWS trends (Fig. 4h). This indicates that areas under greater terrestrial water storage depletion are experiencing greater changes in extreme events in both storage and streamflow. Although there are only limited regions where changes in extreme frequencies are observed for GPP, the EFR rank for GPP slightly increases as the long-term trend rank increases for TWS (Fig. 4i). However, such patterns are relatively minor compared to the corresponding magnitude of the interquartile range. This is understandable since excessive use of freshwater, such as groundwater pumping for irrigation, may stabilize crop production, similar to a favorable water storage condition in response to abundant rainfall. In both cases, they can lead to less systematic extreme frequency changes in carbon fluxes, which is also the case for many of the hotspot regions. There is almost no clear pattern between the TWS trend and the EFR ranks for ET, which again may be associated with the combined effect of depleting water storage limiting deep soil evaporation and enhanced transpiration due to irrigated agriculture. For precipitation extremes, the percentile rank does not change across regions with different levels of TWS depletion, which indirectly implies that the long-term negative trend in TWS for these regions is more due to human water abstractions than natural variability. However, the rank of changes in precipitation extremes reduces to 31% for places with the strongest positive TWS trends. Overall, for regions that have a long-term trend dominating the nonstationarity in water storage, the extremes of both flood potential and droughts are exacerbated under such a background trend, while such unsustainable water use has supported and stabilized the growth of vegetation. As a result, fewer areas are identified with change points in GPP, and even where they exist, the rank of the extreme frequency ratio for GPP is relatively low as compared to elsewhere over the globe.

Our results indicate that there is strong evidence linking the magnitude of the water storage trends to both the existence of change points in extremes and the changes in extreme frequencies for terrestrial water storage and runoff. The fact that small changes in mean state might produce large changes in extremes has been explored in terms of direct (e.g., record-breaking heat extremes in response to a warming trend[4]) and indirect (e.g., linking occurrences of large extreme floods to land use change and

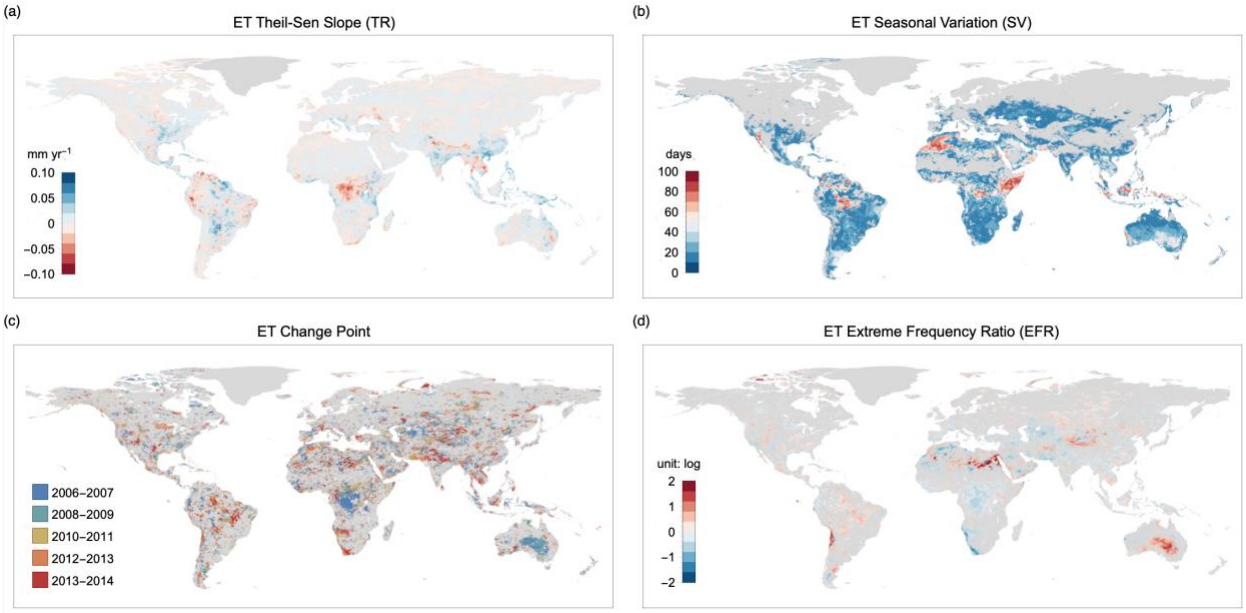
climate warming[5,6]) impact studies. For other variables, such as ET and precipitation, such associations are much weaker and less conclusive. However, based on the documented observed declining trends in groundwater and the associated vegetation greening, such patterns can be attributed to human influence on water use and associated vegetation responses (such as agriculture). Improving our understanding of the physical and dynamical processes behind the nonlinear relationships between long-term trends and changes in extremes is important and needs further investigation as it may provide a foundation with respect to how to better incorporate the nonstationarity concept into conventional practices.



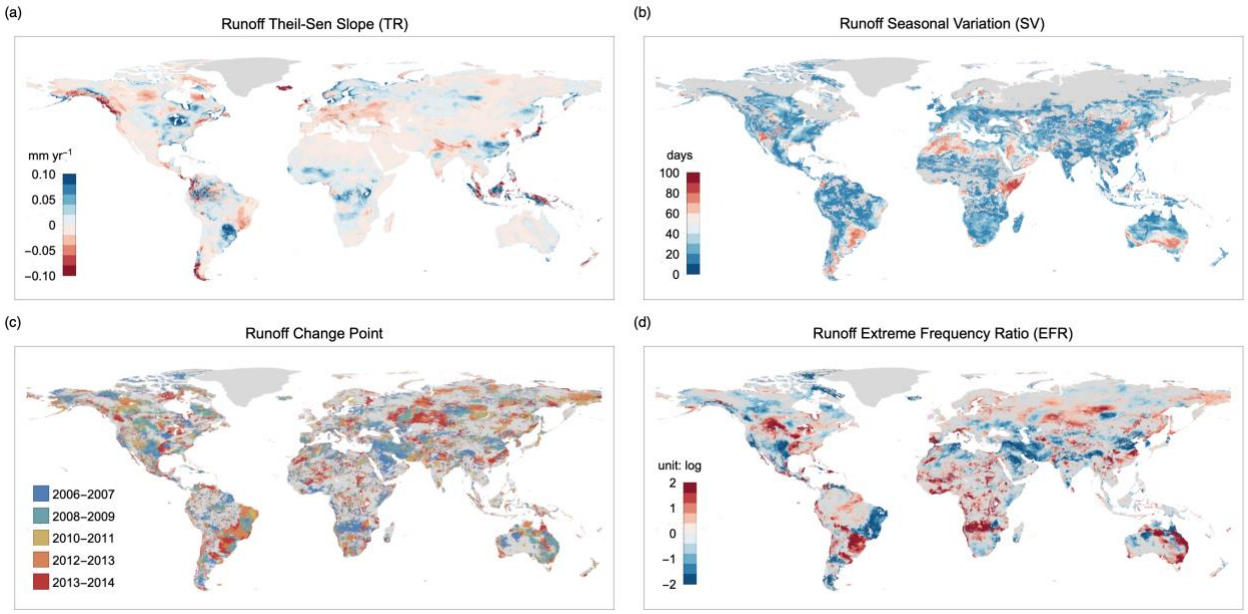
Extended Data Fig. 1 | Global nonstationarity index (NSI) and the corresponding categorical (G1 -G5) distribution for GPP, MERRA-2, and ERA5 precipitation.



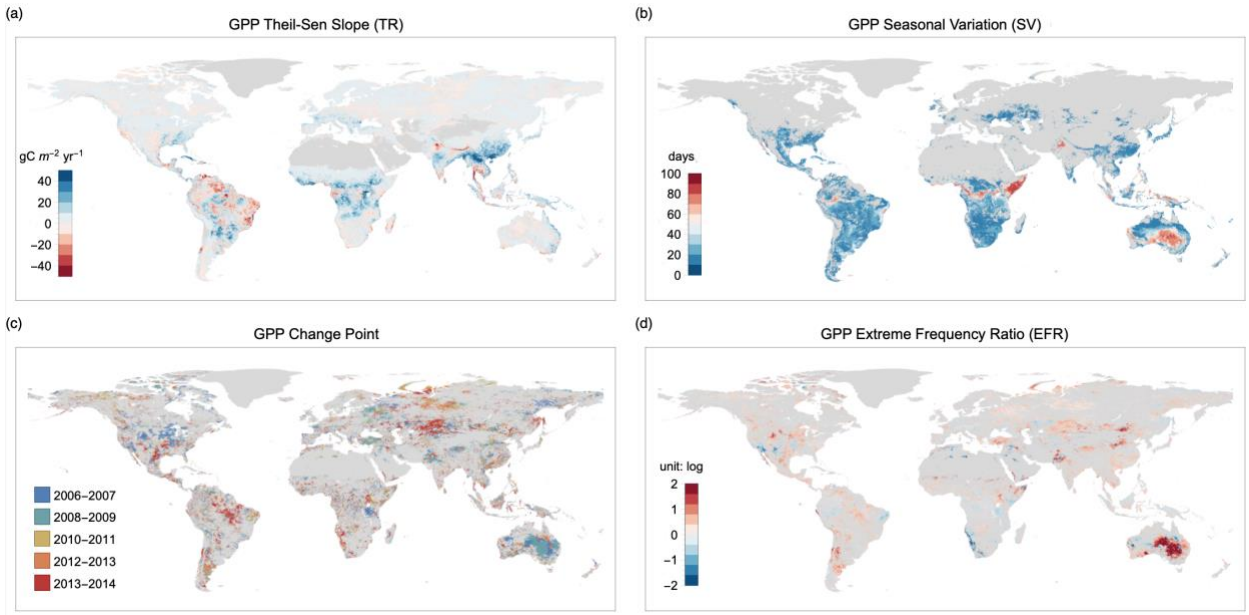
Extended Data Fig. 2 (part 1) | (a) Theil-Sen slope (TR), (b) seasonal variation (SV), (c) change point, and (d) extreme frequency ratio (EFR) with respect to change point for TWS.



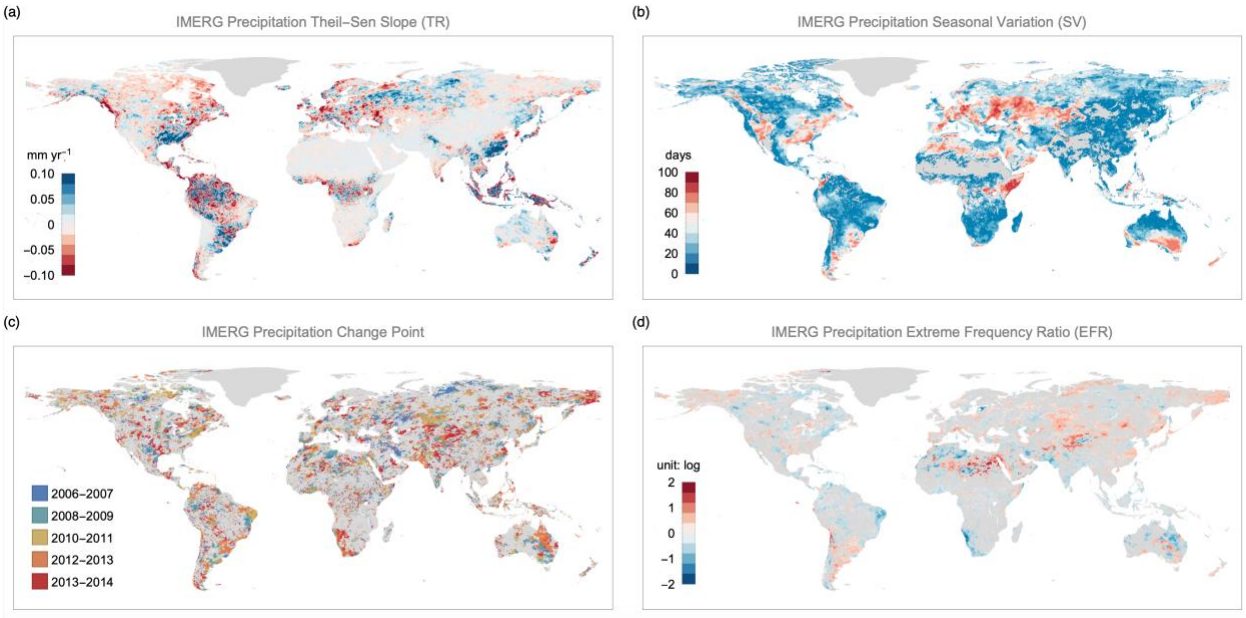
Extended Data Fig. 2 (part 2) | The same as part 1, but for ET.



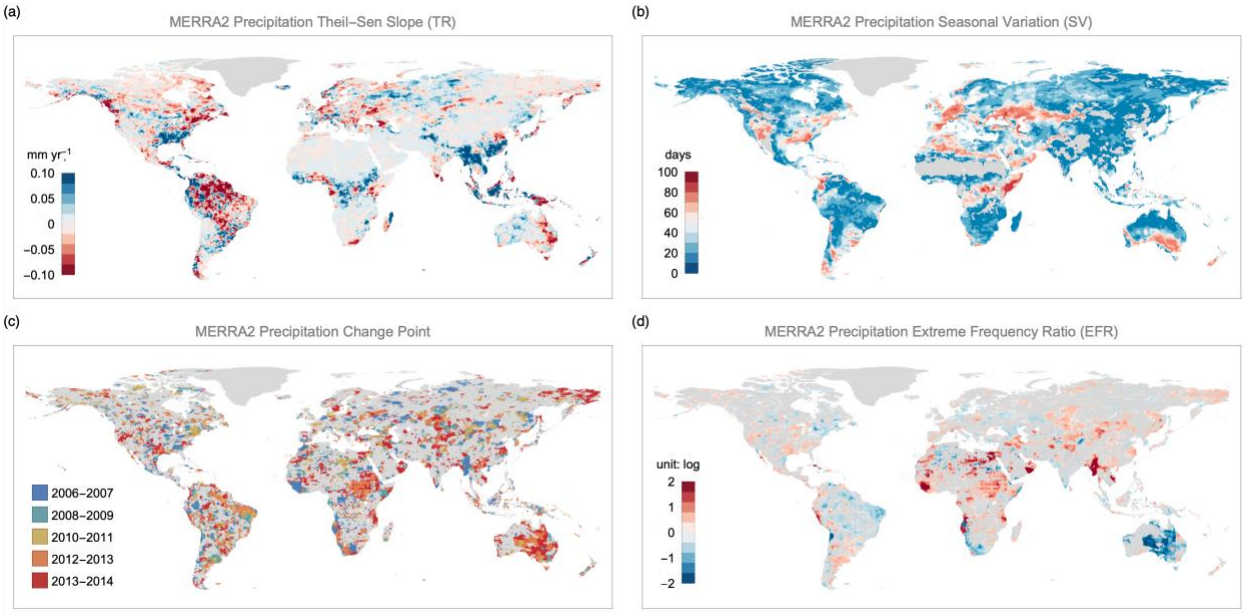
Extended Data Fig. 2 (part 3) | The same as part 1, but for runoff.



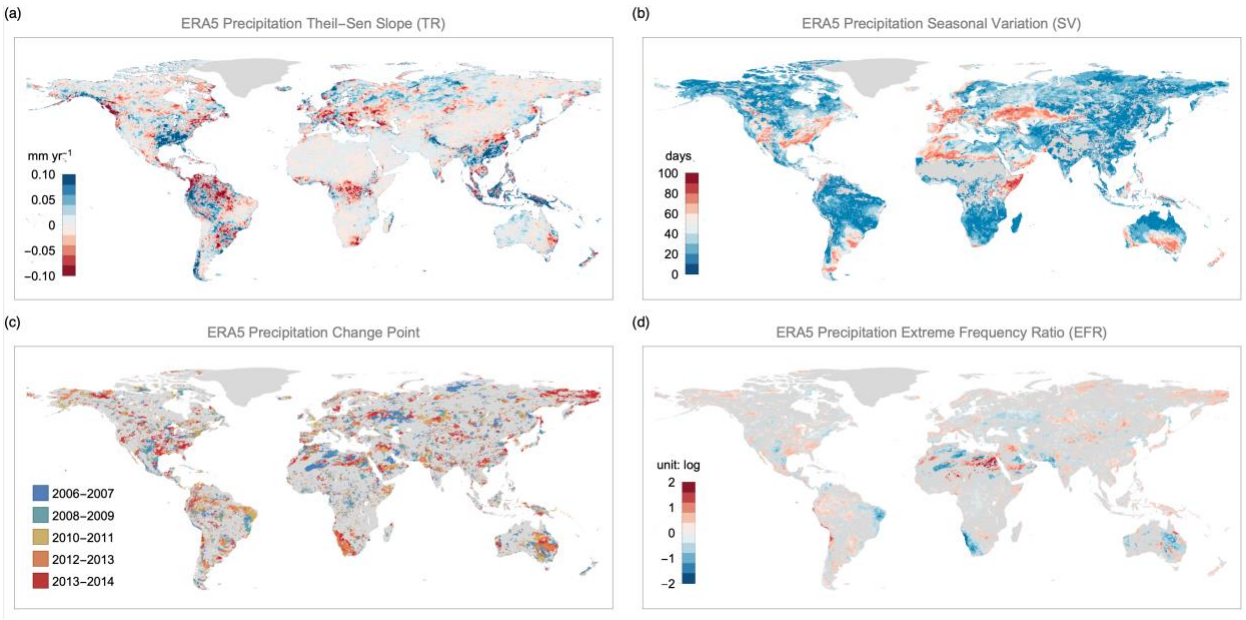
Extended Data Fig. 3 | The same as Extended Data Fig. 2, but for GPP.



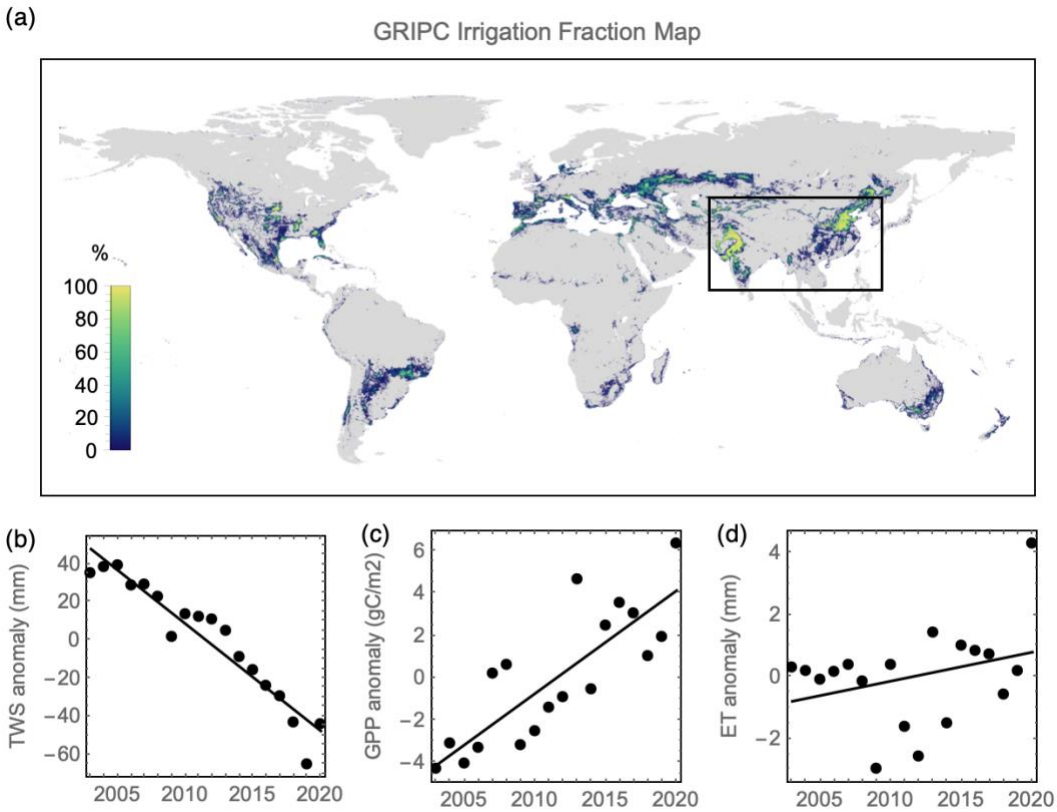
Extended Data Fig. 4 (part 1) | The same as Extended Data Fig. 2, but for IMERG precipitation.



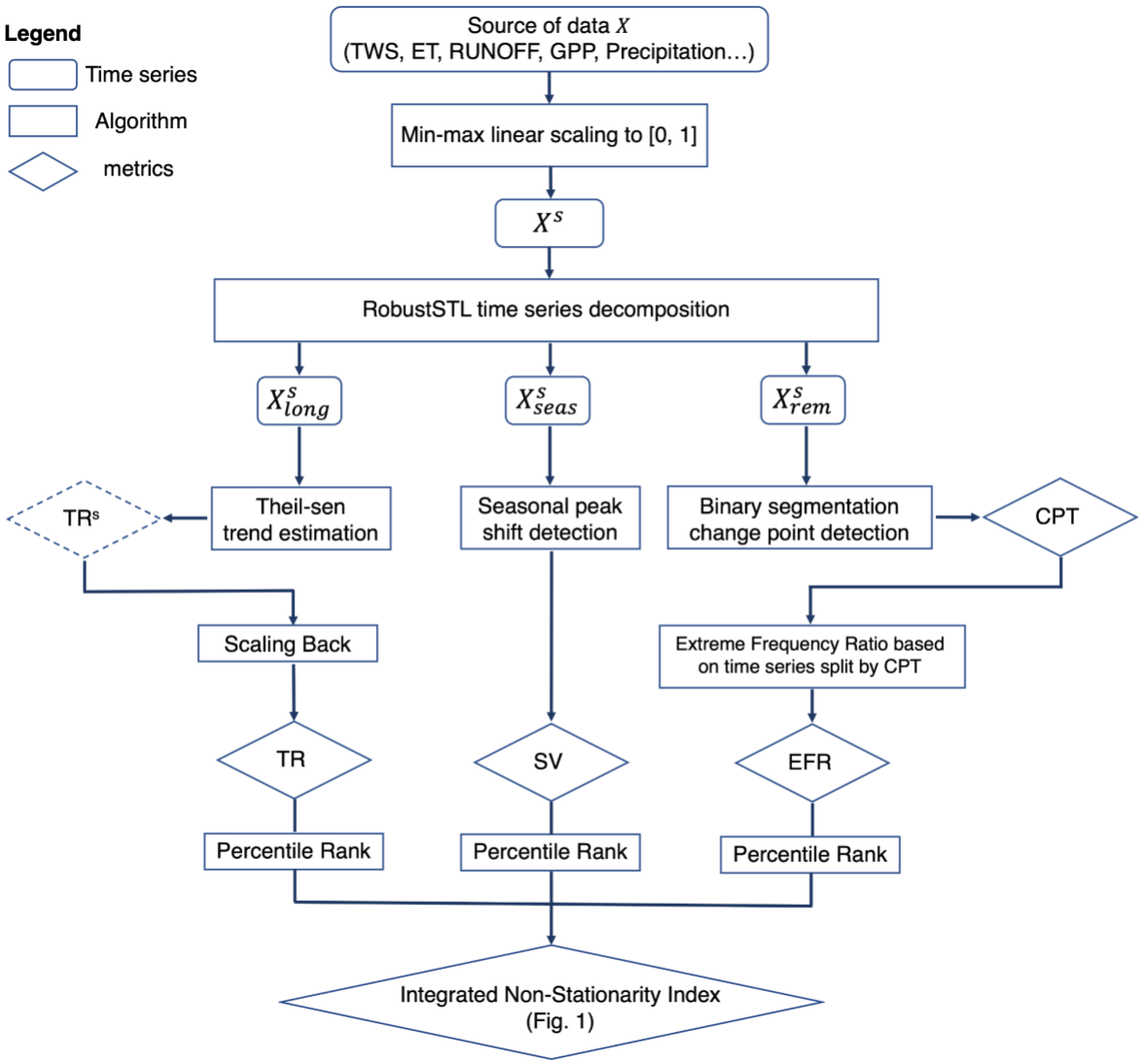
Extended Fig. 4 (part 2) | The same as part 1, but for MERRA-2 precipitation.



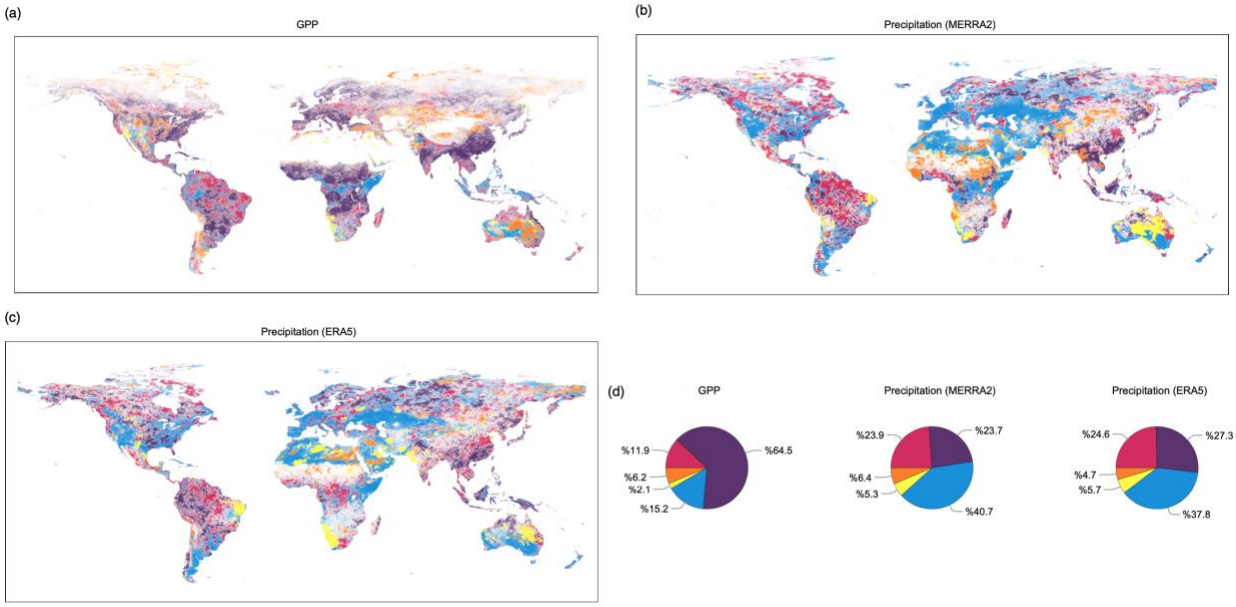
Extended Data Fig. 4 (part 3) | The same as part 1, but for ERA5 precipitation.



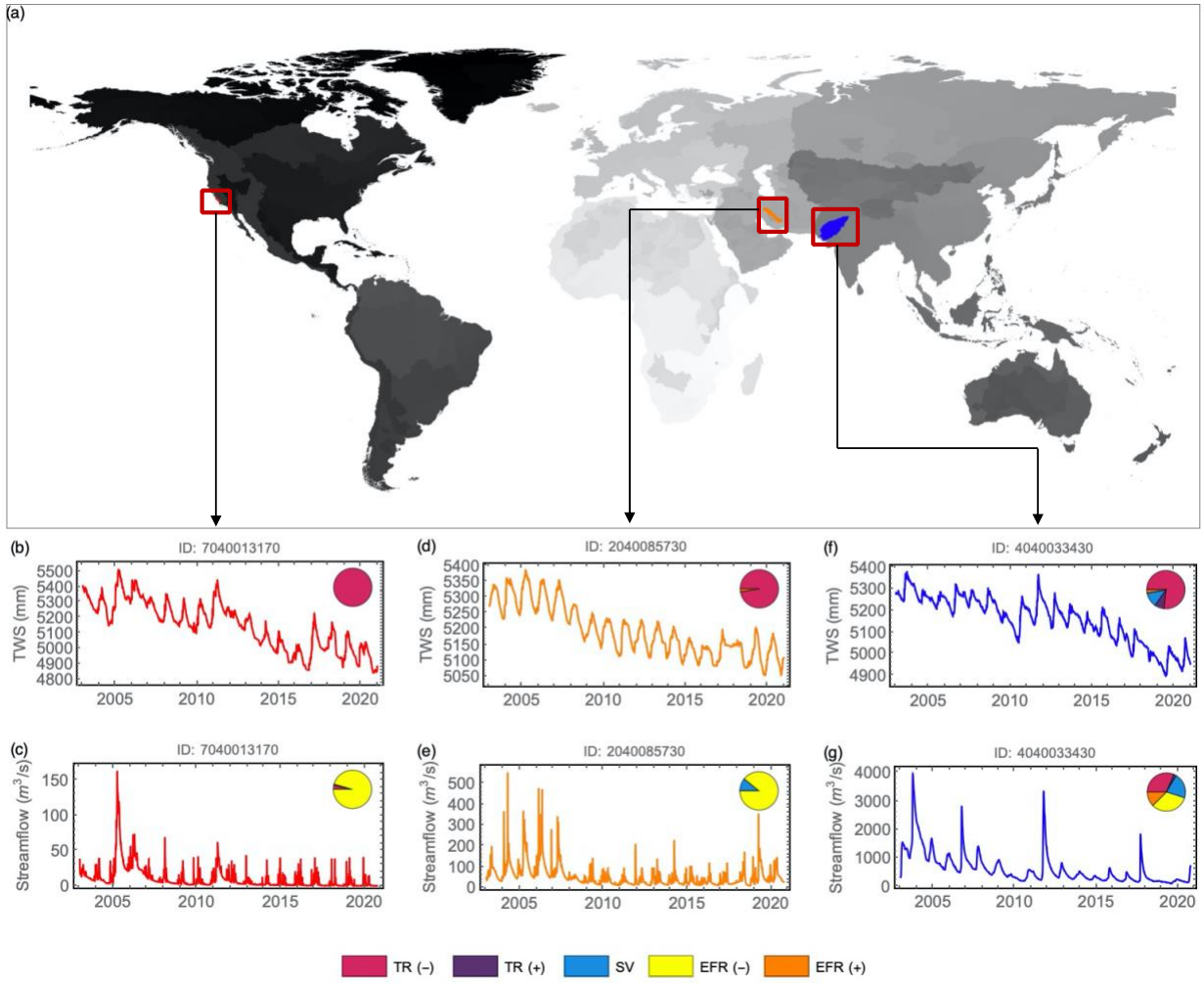
Extended Data Fig. 5 | (a) global irrigation fraction map based on the Global Rainfed, Irrigated, and Paddy Croplands dataset (GRIPC)[7] and (b-d) the annual averaged TWS, GPP, and ET time series along with the linear fitted line of irrigated South Asia domain (outlined in black box on the irrigation map). This example demonstrates that for irrigated South Asia, the depletion of TWS due to groundwater pumping for irrigation has supported the crop production and vegetation growth. Such opposite effects that decreased belowground water storage and increased gross primary production lead to smaller overall long-term patterns in ET.



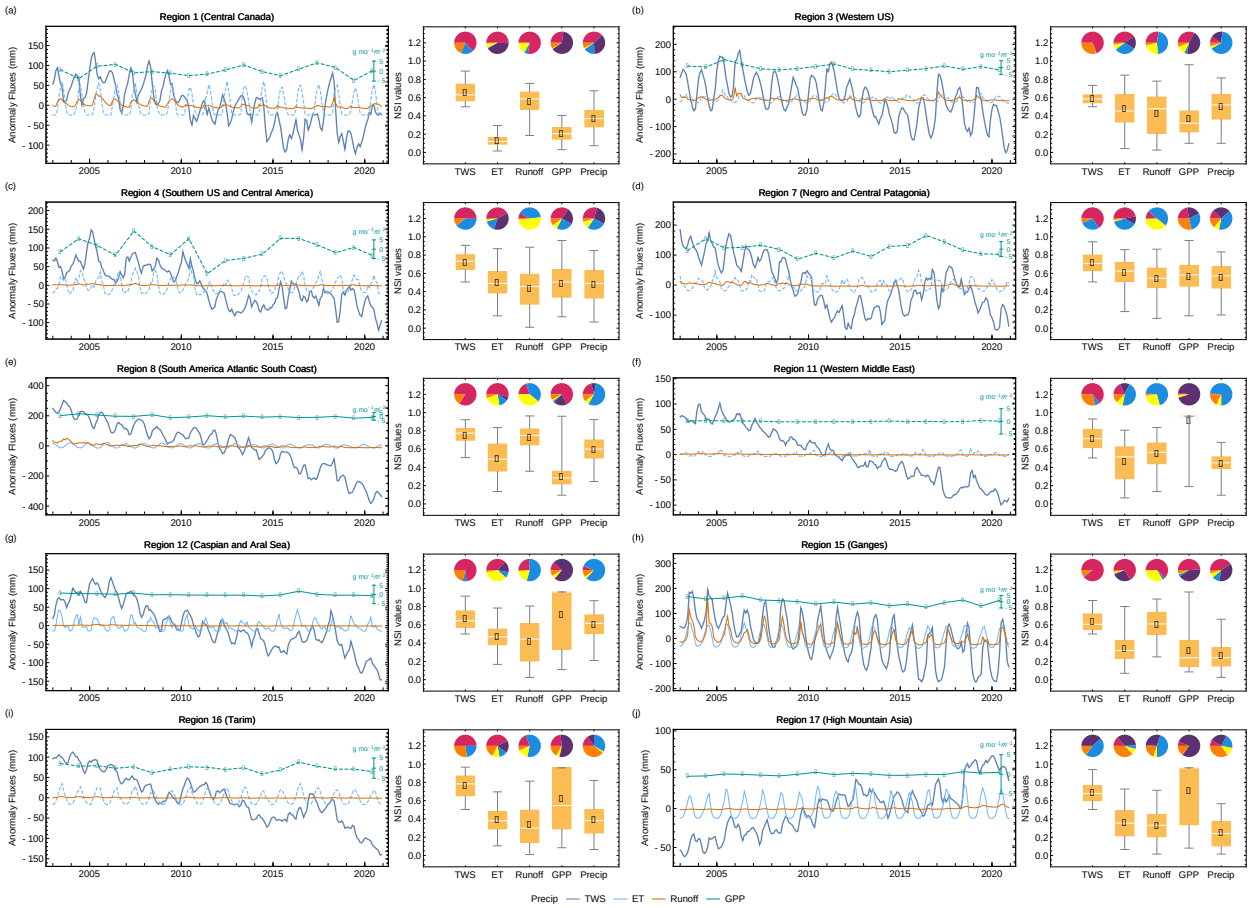
Extended Data Fig. 6 | Flowchart of the derivation of the integrated nonstationarity index (NSI).



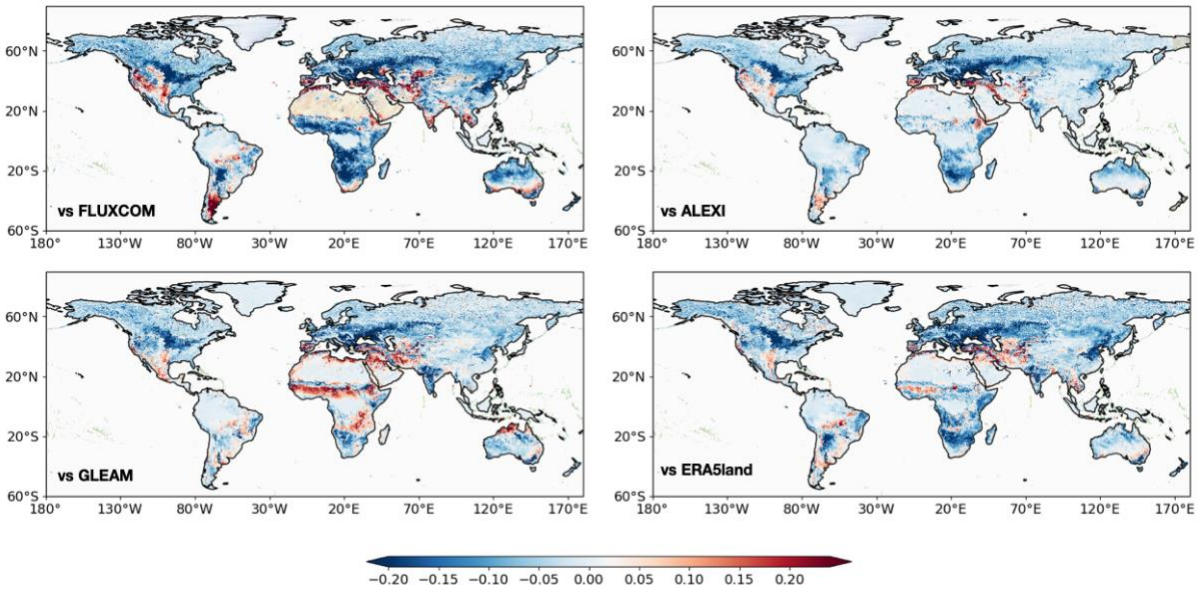
Extended Data Fig. 7 | Spatial distribution of the relative percentile rank amongst trend, seasonal variation, and extreme frequency ratio to nonstationarity for GPP, MERRA-2, and ERA5 precipitation.



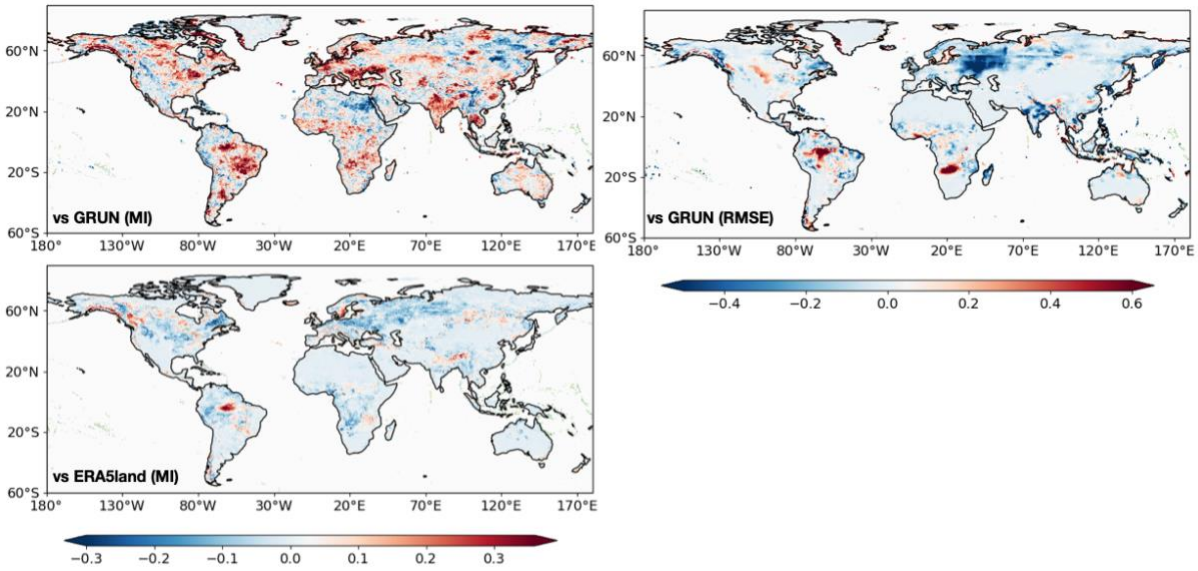
Extended Data Fig. 8 | Three highlighted basins based on WMO HydroSHEDS Level 4 database (a), along with the basin averaged daily TWS and basin outlet daily streamflow for part of Western US (b,c), Iran (d, e) and South Asia region (f, g). Pie charts show the distribution of factors that contribute to nonstationarity in TWS and runoff within each basin respectively.



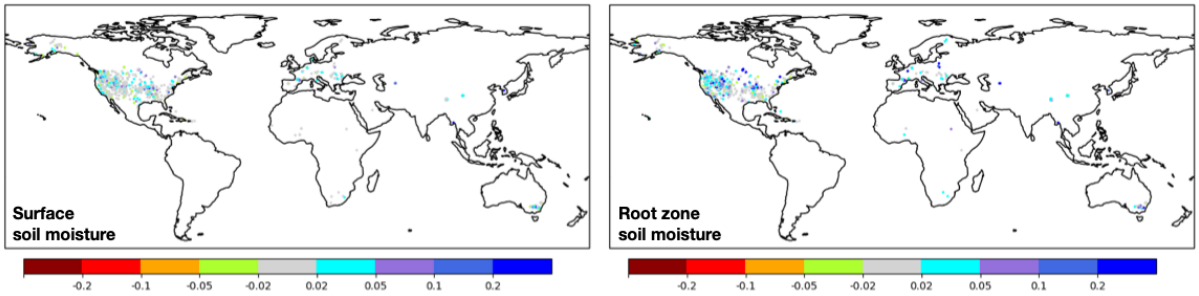
Extended Data Fig. 9 | Changes in fluxes for remaining nonstationary hot spot regions and its relative contributor complementary to Fig. 3.



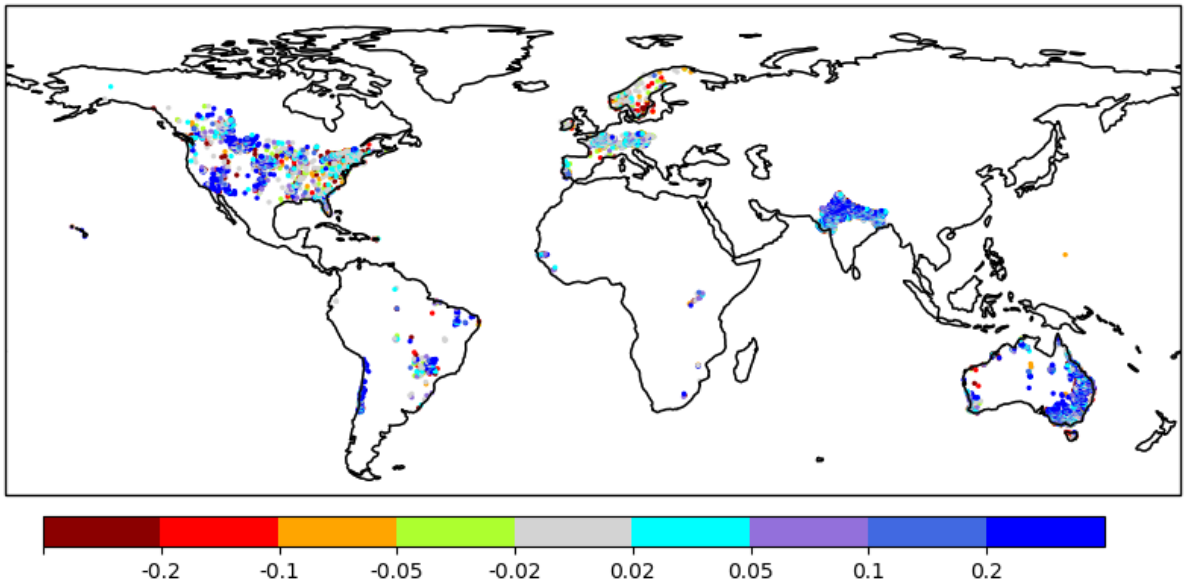
Extended Data Fig. 10 (part 1) | Changes in mutual information of ET (expressed as MI (OL) minus MI (DA) using FLUXCOM, ALEXI, GLEAM, and ERA5land as reference datasets. The warm colors represents degradations from DA and cool colors indicate improvements from DA.



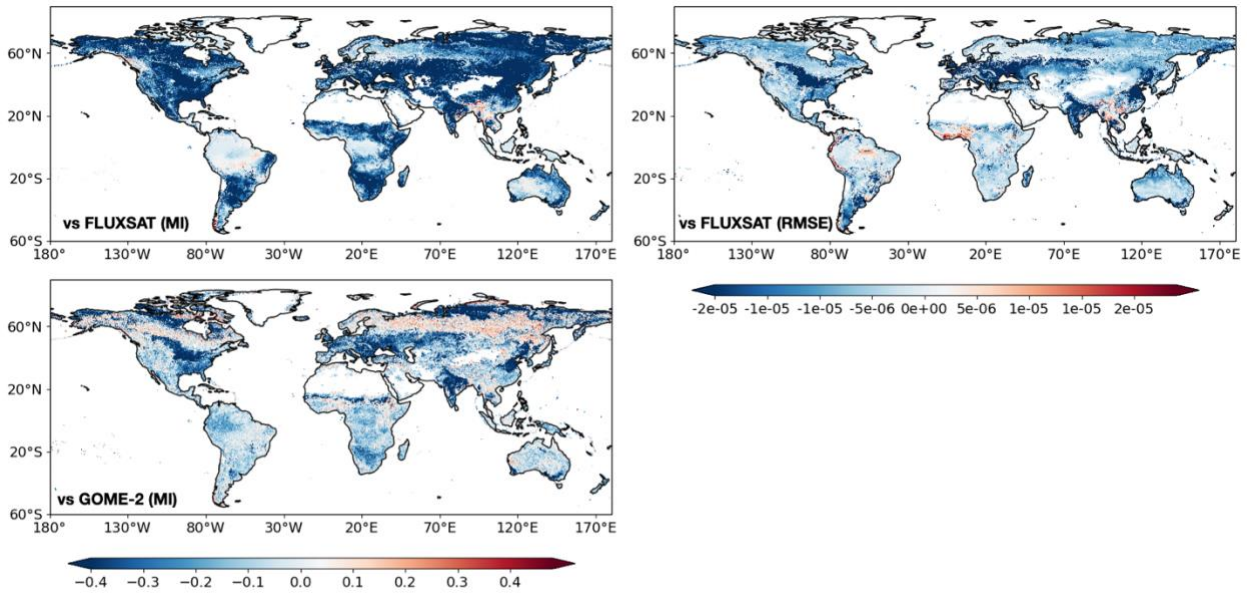
Extended Data Fig. 10 (part 2) | Left column represents the changes in mutual information of runoff (expressed as MI (OL) minus MI (DA) using GRUN and ERA5land as reference datasets. The right column represents the changes in RMSE (mm/day) (computed as RMSE (DA) minus RMSE (OL)) compared to the GRUN data as the reference. Similar to part 1, warm colors represent degradations from DA and cool colors indicate improvements from DA.



Extended Data Fig. 10 (part 3) | Changes in mutual information (expressed as MI (OL) minus MI (DA) for surface soil moisture (left column) and root zone soil moisture (right column) using in-situ soil moisture measurements as the reference.



Extended Data Fig. 10 (part 4) | Changes in mutual information (expressed as MI (OL) minus MI (DA) for groundwater storage with the in-situ groundwater well data as the reference.



Extended Data Fig. 10 (part 5) | Left column represents the changes in mutual information of GPP (expressed as MI (OL) minus MI (DA) using FLUXSAT and GOME-2 as reference datasets. The right column represents the changes in RMSE (g/m2s) (computed as RMSE (DA) minus RMSE (OL)) compared to the FLUXSAT data as the reference.

References

1. Chen, C. *et al.* China and India lead in greening of the world through land-use management. *Nature sustainability* 2, 122–129 (2019).
2. Getirana, A. *et al.* Avert Bangladesh's looming water crisis through open science and better data. *Nature* 610, 626–629 (2022).
3. Coumou, D. & Rahmstorf, S. A decade of weather extremes. *Nature Climate Change* 2, 491–496 (2012).
4. Rahmstorf, S. & Coumou, D. Increase of extreme events in a warming world. *Proc Natl Acad Sci USA* 108, 17905–17909 (2011).
5. Pall, P. *et al.* Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature* 470, 382–385 (2011).
6. O'Connell, P. E., Ewen, J., O'Donnell, G. & Quinn, P. Is there a link between agricultural land-use management and flooding? *Hydrol. Earth Syst. Sci.* 11, 96–107 (2007).
7. Salmon, J. M., Friedl, M. A., Frolking, S., Wisser, D. & Douglas, E. M. Global rain-fed, irrigated, and paddy croplands: A new high resolution map derived from remote sensing, crop inventories and climate data. *International Journal of Applied Earth Observations and Geoinformation* 38, 321–334 (2015).