Reviewer 1

Comment: In the southeast, the PRISM LT dataset uses mostly COOP and WBAN stations so should be very close to what the author is looking for. I respectfully ask that the LT dataset be used as the PRISM dataset in this study.

Response: PRISM LT has been added to the analyses of monthly datasets. Since the study also includes daily datasets, the PRISM AN product remains in the study. In the Discussion section of the manuscript, it is now suggested that PRISM LT could be used for long-term analyses after the correction of a small discontinuity.

Comment: There is also a fairly recent paper from 2021 which covers some important issues about PRISM precipitation time series modeling and can be referenced if desired: https://journals.ametsoc.org/view/journals/atot/38/11/JTECH-D-21-0054.1.xml

Response: That article has been referenced in the following sentence: "PRISM LT data are similar to PRISM AN but are available only at monthly time steps and incorporate substantially fewer gauge networks than the AN product, a design choice intended to improve temporal stability for long-term analyses (Daly et al., 2021)."

Comment: 110: with a minimum of eight years required in each group. Why was eight years chosen?

Response: A minimum of eight years per group was required to ensure adequate statistical power and to reduce sensitivity to individual anomalous years.

Comment: 131: I am not sure what is meant by weather-bureau gauges. Do you mean ASOS (Automated Surface Observing System), WBAN (Weather Bureau Army-Navy), or something else? Based on the rest of the paper, I think you mean ASOS.

Response: The term "weather bureau" has been removed from the manuscript. Instead, terms such as "first-order airport surface observing stations" and "USW/WBAN" are used.

Comment: 134: PRISM, which used gauges from 15 networks, showed a similar pattern, with increasing CoCoRaHS and decreasing COOP coverage.

This suggests that the author used the PRISM AN (All Networks) dataset, rather than the LT (Long Term) dataset. AN (All Networks) is not the best PRISM dataset to use for this analysis. In contrast, the LT dataset was developed precisely for this purpose. See general comments for more information.

Response: PRISM AN was the product used in the first version of the manuscript. Both PRISM LT and PRISM AN are now analyzed. In the Discussion section of the manuscript, it is now suggested that PRISM LT could be used for long-term analyses after the correction of a small discontinuity.

Comment: 137: TerraClimate had much less gauge coverage overall, with a maximum of 25% from cooperative gauges and an abrupt decline from 22% to <1% between 2010 and 2011.

According to the text, TerraClimate used WorldClim climatologies and anomalies from CRU time series and JRA reanalyses (but I think JRA is not used in the CONUS). Is the use of COOP data until 2010 shown in the figure derived from what used in CRU? And why did it suddenly stop?

Response: The data shown in the figure were obtained from CRU. I do not know why the use of COOP data for the TerraClimate product decreased dramatically in 2010.

Comment: 138-139: Information on gauge coverage for gridMET was unavailable.

And yet Figure 3 shows "PRISM & gridMET" station usage. I believe gridMET uses PRISM grids so perhaps Figure 3 is mostly correct

Response: That sentence ("Information on gauge coverage for gridMET was unavailable.") has been removed and gridMET was added to the following sentence. "PRISM AN and gridMET, which used gauges from 15 networks, showed a similar pattern, with increasing CoCoRaHS and decreasing COOP coverage."

Comment: Figure 4. Percent coverage of the southeastern United States over time by gauge networks used in the five precipitation products.

I do not quite understand this figure. The text refers to Figure 4 when describing a comparison of daily and monthly versions of the datasets, but that is not what the caption refers to. Could it be that the caption should read something like: "Difference between monthly and daily precipitation totals (monthly minus daily?) for the five precipitation products." But even so, I would not think that gridMET and TerraClimate would have the same results since I don't think those products are related. Please clarify the figure and the text.

Response: The wrong caption was applied to the figure. The correct caption is as follows: "**Figure 4.** Percent coverage of the southeastern United States over time by gauge networks used in the precipitation products. COOP is the U.S. Cooperative Observer Program. USW/WBAN gauges are first-order airport surface observing stations. CoCoRaHS is the Community Collaborative Rain, Hail and Snow network. RAWS is the Remote Automated Weather Stations network. With respect to the PRISM and gridMET panel, only COOP, RAWS, and first-order airport (WBAN) gauges are used to develop the PRISM LT product."

TerraClimate, which is a monthly product, was the initial product used in earlier versions of the study. When I realized I should include daily versions of products, I included gridMET as its daily counterpart, because both datasets are produced by Climatology Lab at the University of California, Merced.

Comment: 149-151: The overall best products were combinations of products, and those products were nClimGrid-PRISM, Daymet-nClimGrid-TerraClimate, Daymet-nClimGrid-PRISM, Daymet-gridMET-nClimgrid, and Daymet-nClimGrid.

Taking the mean of these datasets at the daily or monthly time step could produce some pretty strange results, even if the combination results in an overall unbiased dataset. They also could be a result of happenstance. For example, does the combination of PRISM and nClimGrid happen to cancel out PRISM's wetting trend from CoCoRaHS with a drying of nClimGrid by increasing its proportion of ASOS stations? I suggest using the term "most unbiased" rather than "best" since it implies that the combined dataset is superior in every way, rather than for this study's narrowly defined purpose.

Response: Thank you for this thoughtful comment. I agree that the term "best" was too broad and could be interpreted as implying overall superiority across temporal scales and applications. My evaluation pertains only to annual precipitation totals and does not extend to performance at daily or monthly time steps. I also acknowledge that apparent bias cancellation among products may reflect differences in station composition and input data rather than an inherent advantage of combining datasets. To address these concerns and to incorporate the newly added PRISM LT product, the subsection has been substantially revised to focus on "most unbiased" performance within the narrowly defined criteria of this study. The paragraph is now two paragraphs, which are provided below.

> "Multiple products are suitable for multi-decadal analyses of annual precipitation totals for the southeastern United States. An optimal product should have three characteristics: (1) a relatively small cumulative residual total in the residual-mass analysis; (2) no statistically significant discontinuities; and (3) a precipitation trend within 10% of the reference

trend. Two product combinations—Daymet–nClimGrid and Daymet–nClimGrid–PRISM LT—meet these criteria. In these combinations, the wetting bias in Daymet is offset by the drying biases in nClimGrid and, to a lesser extent, PRISM LT. The result is a temporally stable series that exhibits no detectable discontinuities and produces a trend closely aligned with the reference time series. A limitation of these combined datasets is the reduction in spatial resolution resulting from use of nClimGrid, which has ~4-km grid cells that are at least 16 times larger than the ~1-km cells of Daymet and PRISM LT. For applications requiring finer spatial detail, this loss of resolution may be undesirable."

"PRISM LT also merits consideration as a viable individual product for multi-decadal precipitation analyses. Although the product has a statistically significant discontinuity in 1993, it produced the smallest cumulative residual total among the individual products, and its decadal trend (27 mm dec^{-1}) is within 10% of the reference trend. The 1993–2024 period can be homogenized by applying a multiplicative adjustment factor of 1.0095 derived using the mean-ratio approach outlined by Peterson et al. (1998). However, a potential long-term limitation of PRISM LT is the continued decline in COOP gauge coverage, as the product is dominated by COOP observations and further reductions in that network could decrease spatial representativeness and potentially reduce accuracy over time."

Comment: Figure 5: I find it interesting that the PRISM AN dataset showed a discontinuity in 2002. This is the year when weather radar-aided interpolation was introduced into AN. Radar was not used in the LT dataset. See general comments for more on this.

Response: I see the comments and I also do not know why the anomalous values in 2002 exist in the PRISM AN dataset.

Comment: 201-202: CoCoRaHS gauges generally record slightly higher precipitation totals than COOP gauges, with increases of about 1–5% (CoCoRaHS, 2019; Goble et al., 2019).

I don't see that information in either of the two references cited, but you are correct. There is a conference paper by Nolan Doesken (2005) worth citing that reports on results from a 10-year comparison of the 8" SRG with the 4" gauge used by CoCoRaHS. He found that overall, the 4" gauge caught 3% more precipitation than the 8" gauge.

Doesken, N., 2005. A ten-year comparison of daily precipitation from the 4” diameter clear plastic rain gauge versus the 8” diameter metal standard rain gauge. Preprints, 13th Symp. on Meteorological Observations and Instrumentation, Savannah, GA, Amer. Meteor. Soc., https://media.cocorahs.org/docs/AMS_NJD_GaugeComparison_AppldClimate_2-2.pdf

Response: The values in those sources were converted to percentages. The 114 mm vs. 112 mm comparison in Goble et al. (2019) corresponds to a 1.8% increase. The statement in the CoCoRaHS document that CoCoRaHS gauges have a collection efficiency of 101–105% compared to the standard NWS gauge correspond to a 1 to 5% increase. Thank you for the Doesken (2005) reference; it has been added to the manuscript.

Comment: Is this 3% increase sufficient to explain the increasing precipitation over the study period? There is also the possibility that COOP observers show low biases compared to CoCoRaHS because of the difficulty in measuring light precipitation amounts with a measuring stick and an opaque gauge. See this paper for more on that:

Daly, C., W.P. Gibson, G.H. Taylor, M.K. Doggett, and J.I. Smith. 2007. Observer bias in daily precipitation measurements at United States Cooperative Network stations. Bulletin of the American Meteorological Society 88(6): 899-912. https://journals.ametsoc.org/view/journals/bams/88/6/bams-88-6-899.xml

One other (maybe silly!) possibility is that CoCoRaHS observers live in areas that are wetter than the regional average, such as Florida, and are skewing the results in that manner. One would have to control for climatological precipitation conditions to see if wetter areas are now being oversampled compared to dry areas.

Response: I appreciate this comment. The potential influence of COOP observer-related biases was carefully considered. However, because Daly et al. (2007) evaluated biases specific to COOP and comparable assessments for CoCoRaHS are limited, we chose not to speculate on differential biases between the two networks without supporting evidence. The Daly et al. (2007) study is now cited in the Data section, and the following sentence has been added: “Although COOP provides the longest regional record, documented observer-related biases indicate that non-climatic variability may persist (Daly et al., 2007).”

A 3% difference in precipitation totals between CoCoRaHS and COOP is enough to cause a substantial change in a trend, so I do believe the introduction and expansion of the CoCoRaHS stations is the main cause of the wetting biases of the Daymet and PRISM AN products. Let’s use Daymet as an example. It has a discontinuity in 2012, after which the precipitation totals are higher than the totals during 1980-2011. If the totals during 2012-2024 are reduced by just 1.8%, then the discontinuity is removed.

Comment: 228-230: Although using this dataset reduces the spatial resolution inherent to Daymet, the resulting gain in temporal homogeneity makes the Daymet–nClimGrid product the most robust dataset for regional, multi-decadal precipitation assessments.

Again, robust does not seem like the proper term. I would again suggest 'to have the most unbiased trends" instead of robust. Also, the author would be remiss if they didn't make some qualifying statements, here. First, the assessment was made using annual precipitation only and did not dig into the monthly or daily data, for example. The second is that the conclusion does not say anything about the overall quality or accuracy of the dataset, only that temporal trends matched up well. Lastly, the evaluation was made at one spatial scale, that of the entire SEUS, with no subregions within it. It is important that your conclusions support the methods and results of the study. For example, I believe that the PRISM LT dataset has stable temporal characteristics and is likely ideal for your purposes, but it is likely not as accurate as PRISM AN on any given day, month, or year. Each dataset has been developed with specific goals in mind.

Response: Thank you for this thoughtful comment. I agree that the term robust was imprecise and that additional qualification is necessary to ensure that the conclusions are fully supported by the scope of the analyses conducted. I have revised the text to (1) replace "robust" with language that more accurately reflects agreement in temporal trends, and (2) explicitly clarify that the evaluation pertains only to annual precipitation trends at the regional scale and does not address daily or monthly performance or overall accuracy. The revised text is provided below.

> "Multiple products are suitable for multi-decadal analyses of annual precipitation totals for the southeastern United States. An optimal product should have three characteristics: (1) a relatively small cumulative residual total in the residual-mass analysis; (2) no statistically significant discontinuities; and (3) a precipitation trend within 10% of the reference trend. Two product combinations—Daymet–nClimGrid and Daymet–nClimGrid–PRISM LT—meet these criteria. In these combinations, the wetting bias in Daymet is offset by the drying biases in nClimGrid and, to a lesser extent, PRISM LT. The result is a temporally stable series that exhibits no detectable discontinuities and produces a trend closely aligned with the reference time series. A limitation of these combined datasets is the reduction in spatial resolution resulting from use of nClimGrid, which has ~4-km grid cells that are at least 16 times larger than the ~1-km cells of Daymet and PRISM LT. For applications requiring finer spatial detail, this loss of resolution may be undesirable."

> "PRISM LT also merits consideration as a viable individual product for multi-decadal precipitation analyses. Although the product has a statistically significant discontinuity in 1993, it produced the smallest cumulative residual total among the individual products, and its decadal trend (27 mm dec^{-1}) is within 10% of the reference trend. The 1993–2024

period can be homogenized by applying a multiplicative adjustment factor of 1.0095 derived using the mean-ratio approach outlined by Peterson et al. (1998). However, a potential long-term limitation of PRISM LT is the continued decline in COOP gauge coverage, as the product is dominated by COOP observations and further reductions in that network could decrease spatial representativeness and potentially reduce accuracy over time."

Reviewer 2

Comment: The author evaluates the Daymet, gridMET, nClimGrid, PRISM, and TerraClimate datasets. The author describes the dataset, showing that most products are largely based on station data. What were the reason for selecting exactly these products?

Response: These products were selected because they represent the primary high-resolution (≤4 km) gridded precipitation datasets available for the conterminous United States that extend back to 1980, which enabled the inhomogeneity and trend analyses in this study. Also included in the analysis is PRISM LT (Long Term), which meets the requirements listed above and was suggested to be added by another reviewer. The PRISM product that in the first version of the manuscript is now named PRISM AN (All Networks).

Comment: The authors mention different kinds of products as well (satellite based, reanalysis). Why did the author not test other potentially useful datasets that do not incorporate precipitation from weather stations? The evaluation of more datasets that do not or only marginally rely on station data would be interesting particularly for research in remote areas.

Response: We appreciate this suggestion. The scope of this study was limited to high-resolution (≤4 km) gridded precipitation products that extend back to at least 1980 in order to ensure a consistent multi-decadal comparison period and fine spatial representation across the southeastern United States. To our knowledge, no satellite-based or reanalysis precipitation datasets currently meet both of these criteria simultaneously. Many satellite and reanalysis products either begin later (e.g., post-1990s) or are available at substantially coarser spatial resolutions.

We agree that evaluating such datasets would be valuable, particularly for remote regions with sparse gauge coverage, but that objective was beyond the scope of the present study.

Comment: Some Figures, such as Figure 2, are hard to interpret given the continuous linear color scale. To cleary show differences between the datasets, difference rasters of precipitation totals of the products would be important, including a discrete color scale of the differences.

Response: I am very appreciative of the suggestion to create difference rasters. Those maps have been created, and that 15-panel figure is the new Figure 3. A new methods section (3.1 Evaluating Spatial Agreement among Precipitation Products) has been added and it as follows:

> “To evaluate spatial agreement and quantify inter-product differences in annual precipitation totals, several complementary comparative analyses were conducted. All six precipitation products were resampled to a common 1-km spatial resolution to ensure direct cell-by-cell comparability across the southeastern United States. Cell-specific percent differences in mean annual precipitation totals over the 1980–2024 period (45-year mean) were calculated for all pairwise combinations of products (15 total comparisons). These calculations produced spatially explicit surfaces representing the magnitude and direction of inter-product differences. To quantify the overall level of agreement, the percentage of the 879,861 grid cells exhibiting absolute differences within ±5% was computed for each pairwise comparison.”

In addition, the paragraph in the first Results section (4.1 Spatial Agreement among Precipitation Products) has been modified as follows:

> “Mean annual precipitation totals and spatial patterns were broadly consistent among the six precipitation products, with Daymet producing slightly higher totals and nClimGrid slightly lower totals than the others (Fig. 2 and 3). Mean annual totals for 1980–2024 ranged from 1,348 mm for nClimGrid to 1,434 mm for Daymet, with precipitation totals smallest (~1,100 mm) across central Georgia, South Carolina, and North Carolina and largest (~2,000 mm) in the Blue Ridge and Cumberland Mountains (Fig. 2). Inter-product differences in mean annual totals were remarkably small: among the 15 pairwise combinations, 67% to 98% of grid cells differed by no more than ±5%, with a mean of 89% of cells meeting this threshold (Fig. 3). The largest discrepancies were concentrated in the Appalachian and Cumberland Mountains and in southeastern Florida, although differences exceeding 15% were rare even in these areas. On average, Daymet produced totals 4.7% higher on average than the other products, whereas nClimGrid produced totals 2.3% lower on average, indicating modest wet and dry biases relative to the other products.”

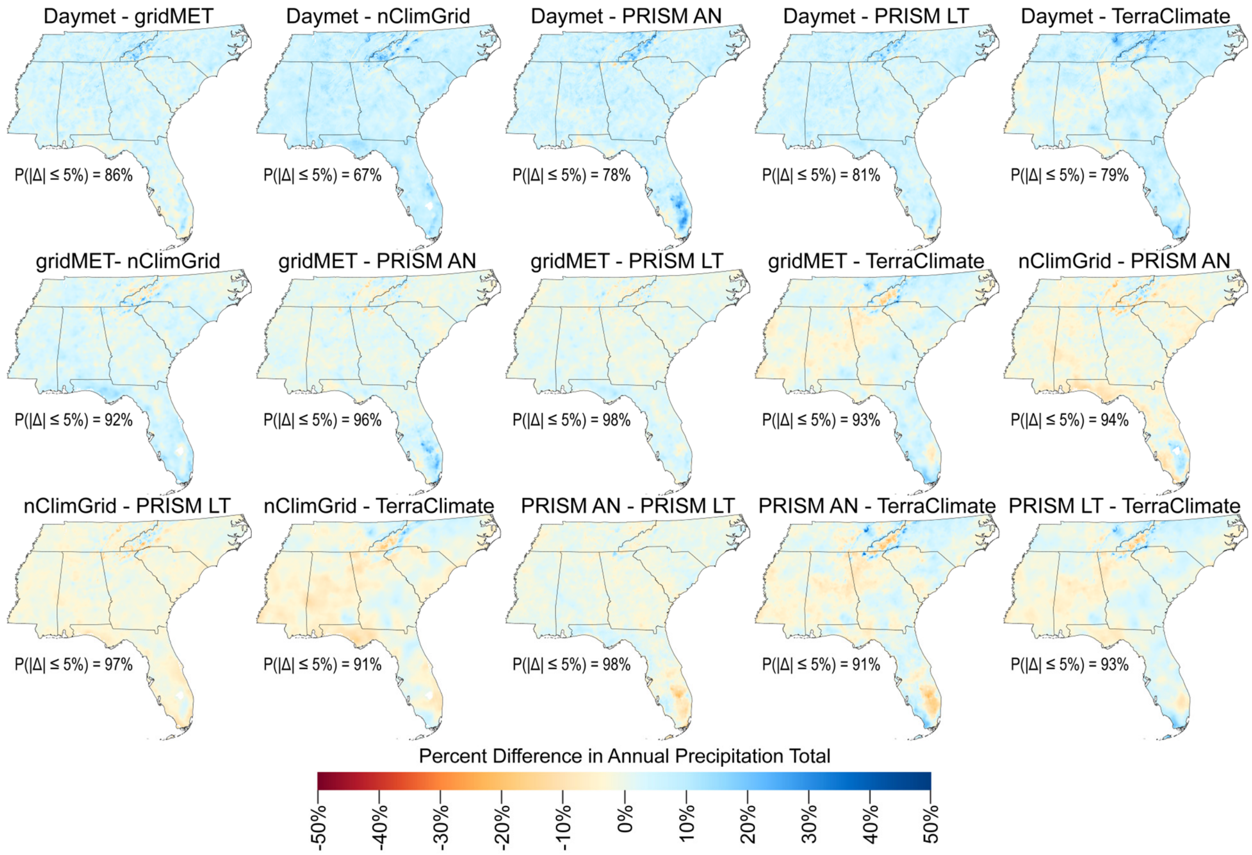


Figure 3. Percent difference in mean annual precipitation totals (1980-2024) between pairs of products. P(|Δ| ≤ 5%) is the percentage of grid cells with percent differences between −5% and 5%.

Comment: As COOP stations were used for the evaluation, datasets that integrate more of respective stations are expected to show better performance than other products. Can you include a discussion about this effect and is it possible to create an independent evaluation dataset that was not used for the creation of the respective products? Does the combination of datasets increase the proportion of evaluation stations and therefore, leads to improved results? Please discuss these issues

Response: Greater inclusion of COOP stations does not necessarily ensure improved temporal performance, particularly because the specific COOP stations incorporated into each product vary over time. Indeed, several products that rely heavily on COOP still exhibited significant discontinuities and divergent trends, indicating that shared gauge input alone does not guarantee agreement with the reference series. Moreover, the improved stability observed in certain product combinations does not reflect an increase in evaluation stations, but rather the offsetting of wetting and drying biases associated with differing network influences. Fully independent evaluation is challenging for multi-decadal U.S. precipitation products because most gridded datasets assimilate COOP observations, and no spatially extensive, high-quality network independent of the products used in the study extends back several decades.

Comment: Regarding the statement that the combination Daymet–nClimGrid is superior to other datasets: this may be based on the strong inclusion of reference stations in these datasets. Without independent evaluation (e.g. test in other area with independent stations) this statement is not very well founded.

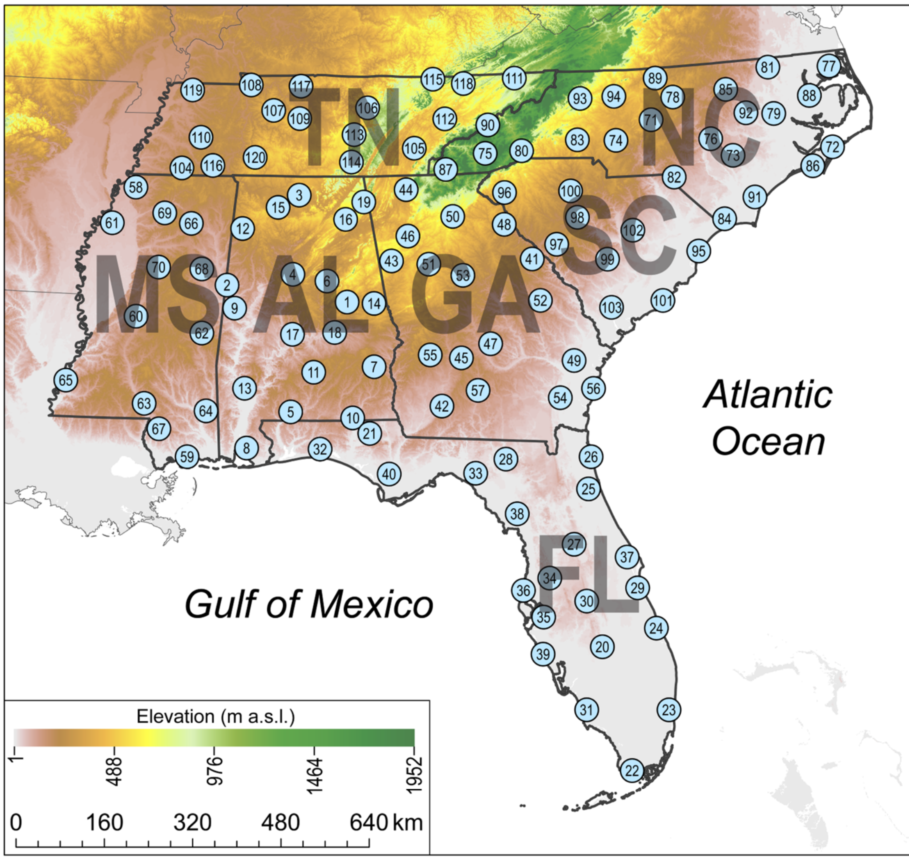
Response: Thank you for this important comment. I agree that any statement implying overall superiority must be carefully bounded, particularly given the lack of independent validation outside the study domain. My conclusions are based solely on agreement with the reference dataset for annual precipitation totals for the southeastern United States. The section has been rewritten to avoid implying superiority and to clarify the criteria used for product suitability. The revised paragraph is provided below.

> "Multiple products are suitable for multi-decadal analyses of annual precipitation totals for the southeastern United States. An optimal product should have three characteristics: (1) a relatively small cumulative residual total in the residual-mass analysis; (2) no statistically significant discontinuities; and (3) a precipitation trend within 10% of the reference trend. Two product combinations—Daymet–nClimGrid and Daymet–nClimGrid–PRISM LT—meet these criteria. In these combinations, the wetting bias in Daymet is offset by the drying biases in nClimGrid and, to a lesser extent, PRISM LT. The result is a temporally stable series that exhibits no detectable discontinuities and produces a trend closely aligned with the reference time series. A limitation of these combined datasets is the reduction in spatial resolution resulting from use of nClimGrid, which has ~4-km grid cells that are at least 16 times larger than the ~1-km cells of Daymet and PRISM LT. For applications requiring finer spatial detail, this loss of resolution may be undesirable."

> "PRISM LT also merits consideration as a viable individual product for multi-decadal precipitation analyses. Although the product has a statistically significant discontinuity in 1993, it produced the smallest cumulative residual total among the individual products, and its decadal trend (27 mm dec^{-1}) is within 10% of the reference trend. The 1993–2024 period can be homogenized by applying a multiplicative adjustment factor of 1.0095 derived using the mean-ratio approach outlined by Peterson et al. (1998). However, a potential long-term limitation of PRISM LT is the continued decline in COOP gauge coverage, as the product is dominated by COOP observations and further reductions in that network could decrease spatial representativeness and potentially reduce accuracy over time."

Comment: Figure 1: Please change the color scale here. In the text it says "The gauges ranged in elevation from 1 m a.sl. to 668 m a.sl.;" The yellowish parts in the figure occur at three different elevations in the legend. This is confusing. Use clearly separable colors.

Response: A different color scheme, where yellow appears only once, is now used to show elevation gradients in Figure 1.



ID	COOP#	ID	COOP#	ID	COOP#
1	010160	41	090311	81	314456
2	010184	42	091500	82	314860
3	010655	43	091640	83	314996
4	010764	44	091863	84	315116
5	011080	45	092266	85	315123
6	011620	46	092485	86	315830
7	011725	47	092966	87	316001
8	012813	48	093060	88	316853
9	013160	49	093538	89	317202
10	013251	50	093621	90	319123
11	013519	51	094700	91	319467
12	013645	52	095863	92	319476
13	014193	53	095988	93	319555
14	014502	54	096219	94	319675
15	015635	55	097087	95	381093
16	017207	56	097808	96	381770
17	017366	57	098703	97	384607
18	018038	58	220237	98	385200
19	018469	59	220792	99	386527
20	080236	60	221389	100	387722
21	081544	61	221707	101	388405
22	083020	62	221860	102	388440
23	083163	63	221865	103	389469
24	083207	64	224966	104	400137
25	083874	65	226177	105	400284
26	084366	66	227106	106	402009
27	085076	67	227128	107	402489
28	085099	68	228374	108	402589
29	085612	69	229079	109	403280
30	085973	70	229743	110	404561
31	086078	71	310286	111	404858
32	086240	72	311606	112	404946
33	087025	73	311881	113	405882
34	087851	74	311975	114	406162
35	088782	75	312200	115	406493
36	088824	76	312500	116	408160
37	088942	77	312719	117	408562
38	089120	78	313555	118	408868
39	089176	79	313638	119	409219
40	089566	80	313976	120	409502

Figure 1. Locations of the 120 reference gauges in the southeastern United States. The seven states that comprise the southeastern United States are Alabama (AL), Florida (FL), Georgia (GA), Mississippi (MS), North Carolina (NC), South Carolina (SC), and Tennessee (TN).All gauges are part of the U.S. Cooperative Observer (COOP) network. In the table, the two leading zeros have been removed from all COOP identification numbers.

Comment: Figure 3 and Figure 4: the captions are switched and generally, should include all necessary information and readability should be improved. How were the differences calculated?

Response: I am sorry about the switched captions in the first version of the manuscript. Those figures are now Figures 4 and 5. The differences are the mean annual precipitation totals from the monthly product minus the mean annual precipitation totals from the daily product. The caption has been changed to the following: “Differences in annual precipitation between monthly and daily products, 1980–2024.”

Comment: Figure 7: please clearly elaborate how these differences were derived in the methods section. Are this pixel means covering all stations vs stations means? What is the reference area?

Response: The differences shown in Figure 7 are based on regional mean precipitation values derived from grid-cell averages, which were compared with the mean precipitation of the 120 COOP gauges. To clarify this procedure and the reference domain, the final sentence of the second paragraph of the Data section has been revised as follows: "When calculating regional means, which were compared with the mean of the 120 COOP gauges, approximately 2.5% of grid cells (those exceeding 668 m a.s.l.) were excluded to restrict the analysis to low- and mid-elevation portions of the Southeast and minimize precipitation inaccuracies associated with mountainous areas."