White James (Orcid ID: 0000-0002-4404-633X)

Title: Using Bayesian Statistics to Detect Trends in Alaskan Precipitation

Short Title: Using Bayesian Statistics to Detect Trends in Alaskan Precipitation

Authors:

James H.R. White

University of Alaska Fairbanks, Department of Atmospheric Sciences

2160 N. Koyukuk Dr.

Fairbanks, AK 99775

jwhite119@alaska.edu

330-322-7223

John E. Walsh

University of Alaska Fairbanks, Alaska Center for Climate Assessment and Policy

Richard L. Thoman, Jr.

University of Alaska Fairbanks, Alaska Center for Climate Assessment and Policy

Keywords: Observational data analysis, Statistical methods, Climate, Rainfall, Atmosphere, Polar, Alaska

Country Keywords: United States, Canada

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/joc.6946

Abstract

Air temperature has exhibited a clear positive trend over the past several decades throughout the arctic, including Alaska. Other variables, such as precipitation, have much more uncertain trends due to inhomogeneities in measurement and high internal variability. The use of linear regression to analyze precipitation in Alaska has resulted in often contradictory results. This paper proposes the use of Bayesian models such as the R package Rbeast to allow for the more nuanced analysis. The examples given in this paper show how Bayesian analysis can be used to detect subtle changes and better constrain the disagreement between data sources. Applied to gridded data, Bayesian analysis shows how precipitation has changed overtime across Alaska. Change has accelerated over the past decade, but only precipitation increase on the North Slope can be assigned high confidence. Overall, this analysis highlights how Bayesian techniques may be uniquely useful to climate research in regions with heterogeneous data sources and substantial internal variability.

Introduction

Alaska, like much of the Arctic, is experiencing unprecedented environmental change (Walsh et al., 2011). While changes in some climate variables, like temperature, exhibit easily identifiable trends, changes in other variables are much more unclear (Bieniek et al., 2014). Precipitation is one such variable because of its many complex processes, large spatial variability, high inter-annual to decadal variability and even difficulties in basic measurement in some environments. Yet precipitation is a highly consequential variable driving winter snowpack, river discharge and associated floods, and glacier mass accumulation. Precipitation is also a key determinant of soil wetness, drought, and wildfire potential. In these and other ways, precipitation impacts infrastructure, ecosystems, and humans.

While climate models largely agree that Alaska's precipitation, particularly summer precipitation, is likely to increase by 2100, observed trends over the past century are much less conclusive (Lader et al., 2017). Most past studies have shown mixed results and uncertain trends for both average and extreme observed precipitation around the state over the past century (McAfee, Guentchev and Eischeid, 2013; McAfee, Guentchev and Eischeid, 2014; Bennett and Walsh, 2014). These issues are enhanced by the spotty and inconsistent station data records in Alaska. Heterogeneities in precipitation histories arise not only from the intermittency of observations at some stations, but also from changes in instrumentation and changes in the measurement locations (Scaff et al., 2015). Given these heterogeneities and the importance of precipitation as a high-impact climate variable, there is a need for more rigorous assessments of variations and trends in precipitation over regions such as Alaska.

The present study brings more rigorous analysis techniques to assess historical variations of precipitation. While most prior studies have utilized different forms of linear regression which may not be well suited to precipitation data, methods such as the Bayesian analysis technique utilized here have the potential to yield more detailed confidence information and estimation of change (Hobbs, 1997). Regardless of whether such analysis methods improve confidence in trends, these methods can provide useful insight to precipitation time series, especially to changes in the mean, seasonal cycle, and trend.

Background and Motivation

McAfee, Guentchev and Eischeid (2013; 2014) provided comprehensive summaries of previous studies that have attempted to detect changes in Alaska's precipitation through the early 2010s. Past analyses based on both station and gridded data have used different types of linear regression (most often ordinary least squares regression) or occasionally spline fitting analysis. Linear regression has many advantages such as simplicity and reproducibility. However, it faces many challenges including strict assumptions that are often not met in Alaska and is highly sensitive to outliers and missing data (Wilks, 2011). Furthermore, linear regression can risk oversimplifying trends, missing important short-term variations, and giving little information on uncertainty at any given time. Spline fitting can give a more wholistic picture but struggles to provide the clean, objective results for which linear regression is so often used. These shortcomings have led past studies of Alaska precipitation to often divergent conclusions about historical and ongoing trends (McAfee, Guentchev and Eischeid, 2013).

In Alaska, sparse and discontinuous station data confound assessments of trends. To alleviate this problem, recent research has utilized climate regions that attempt to aggregate data together into a more complete dataset. The Alaska climate divisions were originally developed by Bieniek et al. (2012) and are now used operationally by the National Centers for Environmental Information (NCEI). The climate divisions for Alaska can be seen in Figure 1. NCEI obtains divisional climate data by area-weighted averaging of a gridded dataset based on continuous NCEI station records and a high-resolution climatology (Daly et al., 2012). These data are aggregated into the current 13 Alaska climate regions in the nCLIMDIV database (Vose et al., 2014). While the underlying gridded data likely suffer from many of the issues seen in other gridded datasets that rely on interpolation, the regional aggregation may help reduce the impacts of heterogeneities at individual locations.

As an alternative to station-based products, reanalysis can provide an independent, continuous source of precipitation data. Additionally, many reanalysis products, especially ERA (European Center for Medium Range Weather Forecasting Re-Analysis), have proven useful in Alaska (Bieniek et al., 2016). Reanalysis, however, carries its own set of issues, including bias compared to observations, inability to resolve local effects, and uncertainties associated with model physics. These biases have been documented for Alaska by Lindsay et al. (2014) and Lader et al. (2016). Given the often-discordant results of observational studies, reanalysis still offers a useful independent source to consider, especially since the complete spatial coverage is achieved by physically based methods.

To illustrate how different datasets compare using traditional linear regression, trends from two reanalysis products were evaluated against the NCEI regional observational data. The recent ERA5 reanalysis (Copernicus, 2017) and a downscaled version of ERA-Interim using the WRF (Weather Research and Forecasting) model (Bieniek et al., 2016) were chosen due to their skill in depicting precipitation (Lader et al., 2016). These reanalysis products were aggregated into the climate divisions using the same area averaging methodology as the NCEI data. From this data, total annual precipitation was graphed, and an ordinary least squares linear regression was then applied.

The time series for selected regions (Figure 2) illustrates several notable features. First, both reanalysis products show a clear bias compared to the observationally-based dataset, although ERA5 is generally closer to observations than the downscaled ERA-I. Such reanalysis bias has been well documented in past research (Lader et al., 2016; Marshall et al., 2018). Second, linear regression results vary widely. While generally not statistically significant when using a Wald test, linear trend lines based on the various sources converge in some regions, diverge in others, and largely agree in some. Of the 3 datasets and 13 regions over the 1979-2018 time period, the trends show statistical significance at the 95% level in only two instances: the NCEI data for the South-Central and West Coast regions. These results contrast with recent local observations of extreme precipitation and predictions for the Arctic as a whole (Min, Zhang and Zwiers, 2008), whose model results suggest a widespread moistening of the Arctic over the second half of the 20th Century. In order to get a more complete picture, this same analysis can be extended back to 1925 for the NCEI data, although both reanalysis products currently only go back to 1979. The linear regression results for this longer time period are shown in Table 1. This method still only produces two statistically significant trends; interestingly in two different regions than previously noted.

Past research has noted the strong influence of teleconnections such as the Pacific Decadal Oscillation (PDO) on Alaskan precipitation (Wendler, Gordon, and

Stuefer, 2017). Because a well-known PDO shift occurred in the mid-1970s, such an influence is likely to lead to a breakpoint in the center of the NCEI timeseries (Hartmann and Wendler, 2005). To account for the possibility of breakpoints in the data, a segmented linear analysis (Muggeo, 2016) was applied to the long-term NCEI data. If a breakpoint was found to be statistically significant at the 95% level, it was added to the regression line. An example of this analysis for the same regions in Figure 2 is presented in Figure 3, and the overall numerical results for all regions are summarized in Table 2. The inclusion of breakpoints greatly enhances the significance of trends in several regions, particularly coastal regions and in Southern Alaska. Most Interior regions do not exhibit confident breakpoints and show the same trends as in Table 1.

Overall, these examples of linear regression are not meant to serve as a comprehensive comparison of reanalysis and observations nor as a detailed analysis of the merits of linear regression. Rather, these examples show how linear regression can often produce differing results on different datasets and time periods, confounding its utility for meaningful trend detection. Although subtle changes may exist, linear regression gives only a single answer over a large swath of time. Furthermore, these examples show how breakpoints may exist in the data that can be used to improve the informativeness of linear regression. The Bayesian approach used in the rest of this paper helps alleviate many of the issues of simple linear regression by adding an uncertainty analysis, allowing for the presence of breakpoints, and determining trends at instantaneous points in time. Thus, the Bayesian method described in the following section provides a more comprehensive analysis and avoids many of the pitfalls associated with linear regression.

Methods

Bayesian modeling leverages prior information to infer model structure. Using prior knowledge, one can set the general structure of a model, but exact model parameter values are left unknown (though they also may be constrained with prior knowledge). Parameter values are then inferred using new observations and Bayesian inference to create probability distributions of likely parameter values. Rbeast is a Bayesian model developed to analyze time series and identify change points in those time series (Zhao et al., 2019). While originally developed for applications of remote sensing to vegetation, the model can be applied to any time series of data that meets its assumptions. The model assumes that an input time series can be decomposed into four separate components: a seasonal component modeled with a harmonic function, a background component modeled with a piecewise linear regression function, some number of possible change points for both the seasonal and background components, and some amount of random noise. In the present application, total monthly precipitation values were used for all Rbeast modeling to ensure that all these assumptions were met. Monthly precipitation throughout Alaska has clearly identifiable seasonal cycles, some climate-based background average, the possibility of having changepoints, and some amount of noise due to internal variability or the chaotic component of the climate system. Monthly precipitation totals were retrieved from the nCLIMDIV database for regional (i.e. climate divisional) information (Vose et al., 2014), the Copernicus datastore for ERA5 data (Copernicus, 2017), and the GSOM version 1 dataset for station data (Lawrimore et al., 2016).

A roadmap for the Rbeast model is shown in Figure 4. A distribution of possible parameters is created by the model as a prior distribution. Prior distributions reflect prior knowledge of the parameter values. In this case, the model initially assumes no general knowledge and hence creates flat distributions. Prior distributions are created for each of the model components as in Figure 4. These distributions can be influenced in a few key ways by the user. Most importantly, the harmonic component period was set at 12 to reflect the actual seasonal cycle. For this study, the maximum number of change-points in the seasonal component was set to 6 and the maximum number of background change-points was set to 12. The minimum changepoint separation was set to 1 year to help minimize the influence of outlier events. The exact values of these parameters do not have a large effect on the results but can help improve model consistency and confidence. These distributions describe many different possible piecewise linear regression and harmonic components, alongside changepoint point number and positions in those components, that could be used to describe a time series. After prior distributions are created, the actual observed (or reanalysis) precipitation data is ingested to create a single posterior Bayesian distribution. The posterior distribution essentially 'weights' the prior distributions based on their ability to accurately fit the time series. The prior distribution defines the space of parameters that could be used to describe a time series and the posterior distribution weights those functions according to their likelihood to actually describe a given time series. The resulting distribution is analytically intractable. In order to understand the results, this distribution is then sampled using a complex Monte Carlo procedure. For determining sample size, this study follows the guidelines of past sensitivity analysis (Zhao et al., 2013). For individual examples, 4 sampling chains were used with 60,000 samples each using a burn-in period of 10,000 samples. Sample chains represent the number of sample runs completed, sample size is the length of each chain, and burn-in is a number of samples discarded at the start of a chain. Because the process is inherently stochastic, exact results can vary between model runs. The parameters chosen here minimize this variability and differences between runs generally do not affect the results presented in this paper. No further processing was done to Rbeast output; rather the model results are visualized "as is" to provide insights to the precipitation trends.

Results

Rbeast outputs a number of parameters that describe the Bayesian distribution. Figure 5 shows a detailed example of Rbeast output. For easy comparison to linear analysis, this example uses NCEI monthly precipitation totals for the Northwest Gulf region (Figure 1). The results show a clear decline in background precipitation from the 1920s until the late 1970s. At that point, a period of rapid increase occurs, followed by several decades of unchanging background precipitation. The uncertainty around the period of rapid change is much greater than at other points in the record. The probability distribution indicates that changepoints are most likely placed in the late 1970s but that there may have been another period of change during the late 1950s. Changepoints are highlighted simply by taking the average number of changepoints in the posterior distribution and then picking the highest probability points in the changepoint probability distribution up to that number of changepoints with at least 1 year separation. Specific changepoints are highlights to add context for analysis, but it may often be more useful to look at the shape of the changepoint probability distribution itself. The presence of one identified changepoint means that the sampled Bayesian models most often used one breakpoint, but some used more, and others used fewer. The seasonal component (bottom half of the figure) has no indicated changepoints. Overall, this series implies a period of rapidly changing background precipitation for the Northwest Gulf Coast during the late 1970's followed by a period of no significant change in the background precipitation nor in the seasonal cycle.

Piecewise linear analysis for Northwest Gulf region's data for 1925-2018 showed an overall decreasing trend followed by an increasing trend with a breakpoint around the early 1950's. The Bayesian Analysis finds a similar overall shape, but the decreasing trend continues until 1970 with a rapid increase over only about 5 years. Where linear analysis implies a gradual increase in precipitation, the Bayesian analysis indicates a much more rapid shift. The Bayesian analysis places its changepoints later than the linear breakpoint (the mid-1970s rather than the early 1950s) but the probability distribution implies a subtle change in the 1950s as well. The shift identified in the Bayesian analysis is likely related to the breakpoint placement in the linear analysis. The mid 1970s changepoint in the Bayesian analysis aligns well with past research which places a PDO associated climate shift around this time period (Hartmann and Wendler, 2005). The example illustrates how RBeast can add significant context to changepoint and trend detection. Compared to linear regression, its results cannot be simply aggregated into a table, as doing so would eliminate much of the context that Bayesian analysis adds.

Beyond producing Rbeast output for each NCEI time series, Rbeast analysis was used to compare reanalysis and some station data records. These comparisons showcase many of the unique strengths of a Bayesian approach. Due to the length of the reanalysis, these comparisons are constrained to the time period (1979-2018). Several notable examples are now presented.

Figure 6 displays a comparison between the University Experimental Station and its corresponding ERA5 grid cell. The University Experimental Station is a long-term weather station record that has been consistently maintained near the University of Alaska Fairbanks campus. Excluding a short, wet period around 1989, the station data record shows a period of relatively little change from 1979 until 2012. It rapidly increases from about 2012 through 2015 before flattening out at this higher level. This increase is associated with a changepoint, however, because the credible interval before and after the changepoint overlap, the change is unlikely to be significant. The ERA5 results show an overall similar, but less pronounced change than the station data. The wet period in the early 1990s appears more prolonged in the reanalysis and the recent wetting does not appear as strongly. The changepoint distribution is similar, though exact placement is much less confident for the reanalysis data. The Fairbanks example shows a location where reanalysis and observations largely agree in terms of general trends. The muted shift in reanalysis likely stems from reanalysis data showing relatively lower extreme highs compared to observations. This likely illustrates the importance of convective and terrain-associated precipitation in the Fairbanks area which is often not well resolved in reanalysis models (de Leeuw, Methven and Blackburn, 2014).

Figure 7 is a similar comparison for King Salmon in Southwest Alaska. The station data for King Salmon shows that precipitation has slightly but consistently increased linearly over the record. A changepoint is placed around 2010 indicating a recent increase in the precipitation trend. Though consistent, the confidence level of this change is not high. The reanalysis results differ substantially from the station-based results. The reanalysis exhibits little change until 2010 with gradually increasing precipitation appearing more recently. It has a changepoint placed around 2017 associated with the recent precipitation increase, though the probability distribution shows very low confidence in this change point. The reanalysis change in precipitation has even lower confidence than the station data. In contrast to the preceding comparison for Fairbanks, the two series here show relatively little agreement. Other comparisons in Southwest Alaska, such as Bethel, show similar results to King Salmon. The exact mechanisms for this disagreement are unclear, but similar to Fairbanks, much of the warm-season rainfall in this area is controlled via small-scale (unorganized) convection which is often not well captured by reanalysis.

Figure 8 shows a comparison for Kuparuk, which is an NWS cooperative station on the Alaskan north coast, west of Prudhoe Bay. The station data show a slightly positive increase until around 2010 where the precipitation increase appears to accelerate. A changepoint is placed around 2012 alongside an acceleration of the precipitation increase. The reanalysis data exhibit a very similar trend to the station data but with a much wider confidence interval. A changepoint occurs in 2012 alongside the station data changepoint, though the probability distribution indicates that the reanalysis changepoint is less confident. Unique among the examples displayed here, the red line in the reanalysis series indicates a changepoint in the seasonal harmonic component. This suggests that the North Slope data may be exhibiting some change in seasonality most likely driven by increasingly wet summers in contrast to dry winters. Similar precipitation increases and agreement between reanalysis and station data is seen in other North Slope stations such as Utqiaġvik.

Figures 9 and 10 show how Rbeast can provide insights to precipitation in Southeast Alaska. In the results for Juneau (Figure 9), the station data indicate a rapid, short term increase in precipitation between 1990 and 1995. This increase is associated with 3 changepoints, although the probability distribution indicates little confidence in their exact placement. The wide uncertainty interval during this period indicates little overall confidence in the background component and hence the trend. The background remains flat for most of the rest of the record with only a recent decrease. The reanalysis series differs greatly, exhibiting a nearly flat trend with only very recent drying. The confidence interval is very wide, indicating little confidence in the recent change. Although one changepoint is placed, the relatively flat probability distribution indicates very little confidence. Similar disagreement, though with different timing, occurs in some other Southeast Alaska stations such as Ketchikan. These similar stations are all first order, automated stations. Figure 10 shows the results for different Southeast Alaska station in Auke Bay, just a few miles from Juneau. In contrast to the Juneau station, this data is collected by NWS cooperative observers, not an automated station. The station record here indicates essentially no change until 2010 where there is a subtle increase followed by a decrease after 2015. An uncertain changepoint is associated with the initial change in 2010. The reanalysis data is very similar, although no increase is observed in 2010. The magnitude of station-reanalysis disagreement for Auke Bay is much smaller than for Juneau. For both cases, the wide confidence interval suggests notable though not significant change. In contrast to Juneau, several other cooperative stations in Southeast Alaska, such as Little Port Walter and Petersburg, show similar agreement to Figure 10. Disagreements between station data and reanalysis, as illustrated by the 1990 spike in Juneau, present a clear issue, as such disagreement is present at several automated stations but wholly absent from nearby Co-Op stations. The NCEI regional data includes both Co-Op sites and first order automated stations, and hence may be influenced by these incongruent shifts. Rbeast analysis for the Central Panhandle region in Figure 11 exhibits an increase in uncertainty after 1990 around the time of the spike seen in Juneau. This time series also exhibits a further increase around 2000 that can be seen in some other areas of the Central Panhandle region but not in Juneau. The 1990 spike has been noted as an inhomogeneity by (McAfee, Guentchev and Eischeid, 2013). These results suggest that the 1990 spike may be caused by station changes rather than actual precipitation, though station metadata does not report any specific change around 1990 (Lawrimore et al., 2016). This example shows a situation where reanalysis may provide a more accurate picture than station data alone. It should be noted that all data sources in Southeast Alaska exhibit a recent decrease in precipitation, likely associated with widespread recent drought (Bathke et al., 2020).

More broadly, reanalysis data allows for exploration on a larger geographic scale in order to assess broad-scale trends. For this purpose, we use the spatially complete ERA5 data. The native ERA5 grid is used and Rbeast is run on each grid cell individually; no extra information is ingested form surrounding cells. Figure 12 shows estimated changes over various timeframes in background precipitation using RBeast based on ERA5. Relative to 1979, the 1990s (panel b) and 2000s (panel c) saw drier conditions in much of the interior and west coast while wetter conditions are apparent on the North Slope. In the 2010s (panel d), much of the Interior and Southwest Coast was wetter than preceding decades and the North slope saw further intensification of a wetter climate. Southeast Alaska, on the other hand, experienced drying. There is also a gradual wetting trend observed in the far western Aleutian Islands and much of the Chukchi Sea throughout the record. In contrast, the Bering Sea experienced more mixed trends, with recent wetting in the west but little change in the central and eastern Bering Sea. Over the Arctic Ocean, there is a decrease north of the Chukchi Sea, peaking in intensity in the early 2000s. Areas of high confidence change are generally limited to the 2010s and to the North Slope and west of the far Aleutian Islands. High confidence change only occurs in areas of increased precipitation; no area of precipitation decrease is highlighted with high confidence.

Discussion

In situ observations often provide a more detailed and confident record for a given location, but reanalysis enables investigation over large areas where observations are not present. Figure 12 illustrates how Bayesian analysis may be applied to a gridded dataset. The details afforded by Bayesian analysis allow for identification of trend intensity over time. Furthermore, the uncertainty statistics allow for areas of high certainty to be highlighted. While trends vary greatly, some definite conclusions can be made. Precipitation on much of the North Slope has significantly increased over the past 40 years. Additionally, precipitation has significantly increased over the western Bering Sea, although little observational data exists to verify the reanalysis there. No area of decrease is marked as significant, so the broad trends point to a wetter climate. This finding aligns well with current climate projections (IPCC, 2013; Lader et al, 2017).

While overall trends are positive, the trends have changed unevenly around Alaska in the last 40 years, and many areas, such as the Interior, have only recently seen increased precipitation. Furthermore, short term climate events such as the drought in Southeast Alaska (Bathke et al., 2020) can have large effects on the Rbeast results, so these trends that are not statistically significant are not necessarily related to climate change. That said, areas highlighted as significant change are more likely associated with climate change rather than interannual variability. The Kuparuk location shown in Figure 8 is within the area of statistical significance in Figure 12d. In this case, the two data sources agree that precipitation has gradually increased over 40 years, though the increase in the reanalysis data is somewhat less confident. The narrow confidence interval in this example suggests that the significance of this change is due to both the large magnitude of the change and the narrower confidence interval generated from less variable precipitation.

The results presented here highlight differences between Bayesian analysis and traditional linear regression. Perhaps the most obvious practical difference is that Bayesian analysis allows for direct decomposition of background and seasonal signals. Furthermore, Bayesian statistics allow for trends to be analyzed at specific points in time rather than as a single linear regression spread over a large period. This allows the Bayesian analysis to identify short-term or subtle trends that are missed by linear analysis. Alongside instantaneous trend identification, Bayesian statistics also allow for a varying confidence range, giving a more nuanced picture of uncertainty. Breakpoint analysis in linear regression yields a single deterministic location, while Bayesian analysis can identify many possible changepoints while providing probability distributions for their placement. Bayesian analysis also enables a more temporally detailed analysis while giving an explicit picture of component uncertainty. While interpretation may be less straightforward than linear analysis, the added context can provide significant nuance. The examples explored here showed often very wide uncertainty ranges, likely stemming from the large inter-annual variability in precipitation. Less variable parameters like temperature show narrower confident ranges, allowing Bayesian analysis to detect subtle changes that linear analysis may miss.

Conclusions

The examples given in this paper show how Bayesian analysis can be used to detect subtle changes in precipitation and better constrain the disagreement between different sources of precipitation data. For example, the application of Rbeast shows how Bayesian methods can be used to detect the start and end dates of trends and, more importantly, to assign confidence levels changes in trend components. The method can also identify changes in the seasonal cycle, although there was little evidence for such changes in the precipitation data examined here.

This analysis drew upon several data sources, including records from individual stations, precipitation data aggregated into climate divisions, and a state-of-the-art atmospheric reanalysis, ERA5. Reanalysis has the advantage of spatial and temporal completeness, and it arguably provides the best avenue to a robust assessment of trends over time. Application of the Bayesian method to the different sources of data for a specific location can lead to the identification of spurious heterogeneities in the station data. Examples presented here for Southeast Alaska showed that the change to the automated observing system in the early 1990s resulted in such a heterogeneity in the station data at some locations, demonstrating further the insights that can be provided by the Bayesian method.

When applied to ERA5 gridded precipitation data for Alaska and the surrounding seas, Rbeast shows how precipitation has changed spatially over time. Recent increases are found over northern Alaska, parts of the Interior, nearshore regions of the Beaufort and Chukchi Seas, and a large area of the western Bering Sea. Except for the northern Alaska coastal region, these changes emerged only with the addition of the most recent decade (2010-2019) to the ERA5 reanalysis. However, only northern Alaska's increase of precipitation can be assigned high confidence. The emergent character of this increase over much of the Alaska domain is consistent with global climate model projections of increased precipitation in high latitudes due to anthropogenic forcing. In this respect, the Bayesian method appears to be an ideal candidate for systematically monitoring a variable that has widespread impacts in Alaska, as noted in the Introduction. The results presented here show the potential of the Bayesian method to not only monitor but to diagnose the trends in climate variables. Applications to other variables such as temperature, wind speed, and even cryosphere variables such as sea ice and snow cover appear to have merit, especially since the seasonal cycle of these variables can be stronger than the seasonal cycle of the precipitation, which was the focus of this study.

Acknowledgments

This work was supported by the Climate Program Office of the National Oceanic and Atmospheric Administration through Grant NA16OAR4310162 to the Alaska Center for Climate Assessment and Policy. We would like to thank the reviewers for their constructive comments on this paper.

References

Bathke, D.J., Prendeville, H.R., Jacobs, A., Heim, R., Thoman, R. and Fuchs, B. (2019). Defining Drought in a Temperate Rainforest. Bulletin of the American Meteorological Society, 100(12), pp.2665–2668.

Bennett, K.E. and Walsh, J.E. (2014). Spatial and temporal changes in indices of extreme precipitation and temperature for Alaska. International Journal of Climatology, 35(7), pp.1434–1452.

Bieniek, P.A., Bhatt, U.S., Thoman, R.L., Angeloff, H., Partain, J., Papineau, J., Fritsch, F., Holloway, E., Walsh, J.E., Daly, C., Shulski, M., Hufford, G., Hill, D.F., Calos, S. and Gens, R. (2012). Climate Divisions for Alaska Based on Objective Methods. Journal of Applied Meteorology and Climatology, 51(7), pp.1276–1289.

Bieniek, P.A., Bhatt, U.S., Walsh, J.E., Rupp, T.S., Zhang, J., Krieger, J.R. and Lader, R. (2016). Dynamical Downscaling of ERA-Interim Temperature and Precipitation for Alaska. Journal of Applied Meteorology and Climatology, 55(3), pp.635–654.

Bieniek, P.A., Walsh, J.E., Thoman, R.L. and Bhatt, U.S. (2014). Using Climate Divisions to Analyze Variations and Trends in Alaska Temperature and Precipitation. Journal of Climate, 27(8), pp.2800–2818.

Copernicus Climate Change Service (C3S) (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate . Copernicus Climate Change Service Climate Data Store (CDS), [Accessed: 2019 October 15]. https://cds.climate.copernicus.eu/cdsapp#!/home

Daly, C., Widrlechner, M.P., Halbleib, M.D., Smith, J.I. and Gibson, W.P. (2012). Development of a New USDA Plant Hardiness Zone Map for the United States. Journal of Applied Meteorology and Climatology, 51(2), pp.242–264.

de Leeuw, J., Methven, J. and Blackburn, M. (2014). Evaluation of ERA-Interim reanalysis precipitation products using England and Wales observations. Quarterly Journal of the Royal Meteorological Society, 141(688), pp.798–806.

Hartmann, B. and Wendler, G. (2005). The Significance of the 1976 Pacific Climate Shift in the Climatology of Alaska. Journal of Climate, 18(22), pp.4824–4839.

Hobbs, B.F. (1997). Bayesian Methods for Analysing Climate Change and Water Resource Uncertainties. Journal of Environmental Management, 49(1), pp.53–72.

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley(eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Lader, R., Bhatt, U.S., Walsh, J.E., Rupp, T.S. and Bieniek, P.A. (2016). Two-Meter Temperature and Precipitation from Atmospheric Reanalysis Evaluated for Alaska. Journal of Applied Meteorology and Climatology, 55(4), pp.901–922.

Lader, R., Walsh, J.E., Bhatt, U.S. and Bieniek, P.A. (2017). Projections of Twenty-First-Century Climate Extremes for Alaska via Dynamical Downscaling and Quantile Mapping. Journal of Applied Meteorology and Climatology, 56(9), pp.2393–2409.

Lawrimore, Jay H.; Ray, Ron; Applequist, Scott; Korzeniewski, Bryant; Menne, Matthew J. (2016): Global Summary of the Month (GSOM), Version 1. [indicate subset used]. NOAA National Centers for Environmental Information. https://doi.org/10.7289/V5QV3JJ5. [Accessed 2019 October 15].

Lindsay, R., Wensnahan, M., Schweiger, A. and Zhang, J. (2014). Evaluation of Seven Different Atmospheric Reanalysis Products in the Arctic. Journal of Climate, 27(7), pp.2588–2606.

Marshall, G.J., Kivinen, S., Jylhä, K., Vignols, R.M. and Rees, W.G. (2018). The accuracy of climate variability and trends across Arctic Fennoscandia in four reanalyses. International Journal of Climatology, 38(10), pp.3878–3895.

McAfee, S.A., Guentchev, G. and Eischeid, J.K. (2013). Reconciling precipitation trends in Alaska: 1. Station-based analyses. Journal of Geophysical Research: Atmospheres, 118(14), pp.7523–7541.

McAfee, S., Guentchev, G. and Eischeid, J. (2014). Reconciling precipitation trends in Alaska: 2. Gridded data analyses. Journal of Geophysical Research: Atmospheres, 119(24), pp.13,820-13,837.

Min, S.-K., Zhang, X. and Zwiers, F. (2008). Human-Induced Arctic Moistening. Science, 320(5875), pp.518–520.

Muggeo, V.M.R. (2016). Testing with a nuisance parameter present only under the alternative: a score-based approach with application to segmented modelling. Journal of Statistical Computation and Simulation, 86(15), pp.3059–3067.

Scaff, L., Yang, D., Li, Y. and Mekis, E. (2015). Inconsistency in precipitation measurements across the Alaska–Yukon border. The Cryosphere, 9(6), pp.2417–2428.

Vose, R.S., Applequist, S., Squires, M., Durre, I., Menne, M.J., Williams, C.N., Fenimore, C., Gleason, K. and Arndt, D. (2014). Improved Historical Temperature and Precipitation Time Series for U.S. Climate Divisions. Journal of Applied Meteorology and Climatology, 53(5), pp.1232–1251.

Walsh, J.E., Overland, J.E., Groisman, P.Y. and Rudolf, B. (2011). Ongoing Climate Change in the Arctic. AMBIO, 40(S1), pp.6–16.

Wendler, G., Gordon, T. and Stuefer, M. (2017). On the Precipitation and Precipitation Change in Alaska. Atmosphere, 8(12), p.253.

Wilks, D.S. (2011). Statistical methods in the atmospheric sciences. Waltham, Ma: Academic Press.

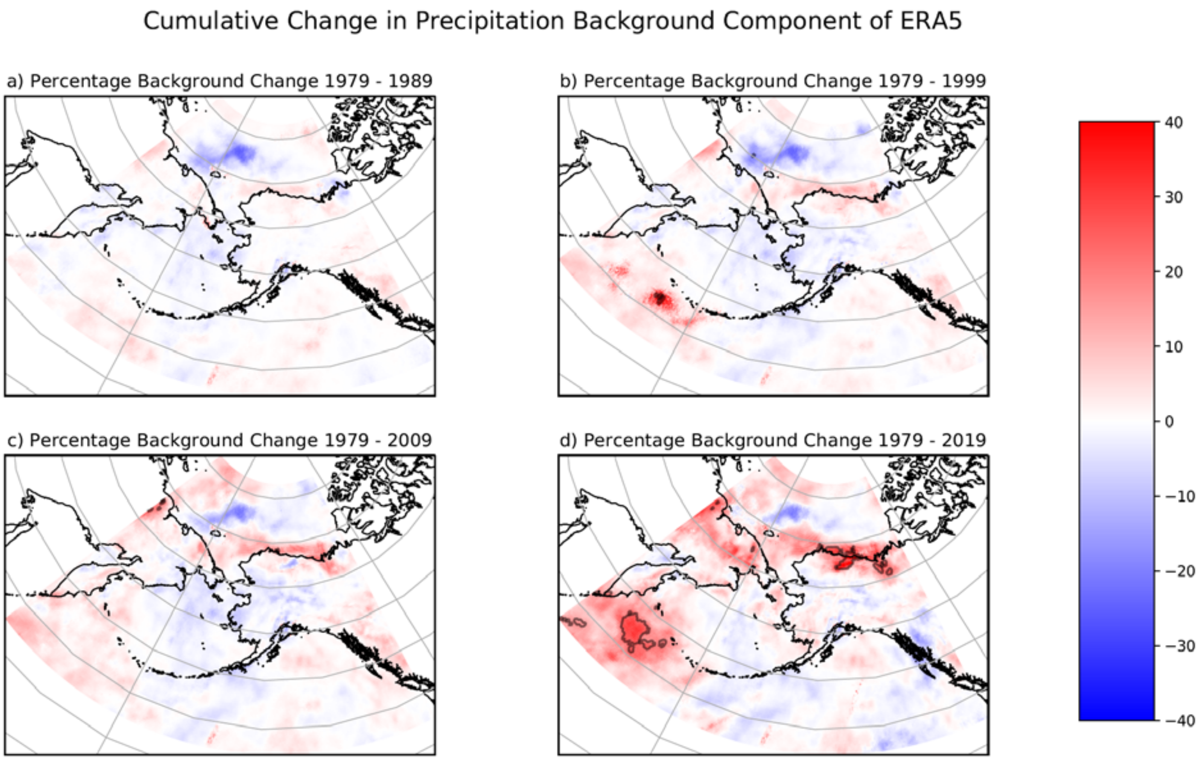
Zhao, K., Valle, D., Popescu, S., Zhang, X. and Mallick, B. (2013). Hyperspectral remote sensing of plant biochemistry using Bayesian model averaging with variable and band selection. Remote Sensing of Environment, 132(1), pp.102–119.

Zhao, K., Wulder, M.A., Hu, T., Bright, R., Wu, Q., Qin, H., Li, Y., Toman, E., Mallick, B., Zhang, X. and Brown, M. (2019). Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. Remote Sensing of Environment, 232(1), p.111181.

Title: Using Bayesian Statistics to Detect Trends in Alaskan Precipitation

Authors: James H.R. White*, John E. Walsh, Richard L. Thoman Jr.

Graphical Abstract



(Note that this is the same as Figure 12 in the manuscript)

The large variability in occurrence and measurement of precipitation in Alaska has complicated trend identification via traditional linear regression. This paper applies recently developed Bayesian analysis techniques to better understand how precipitation in the state has changed over time. When compared to linear regression, the analysis presented here gives a more detailed picture of trend uncertainty and aids in the identification of disagreement between data sources.

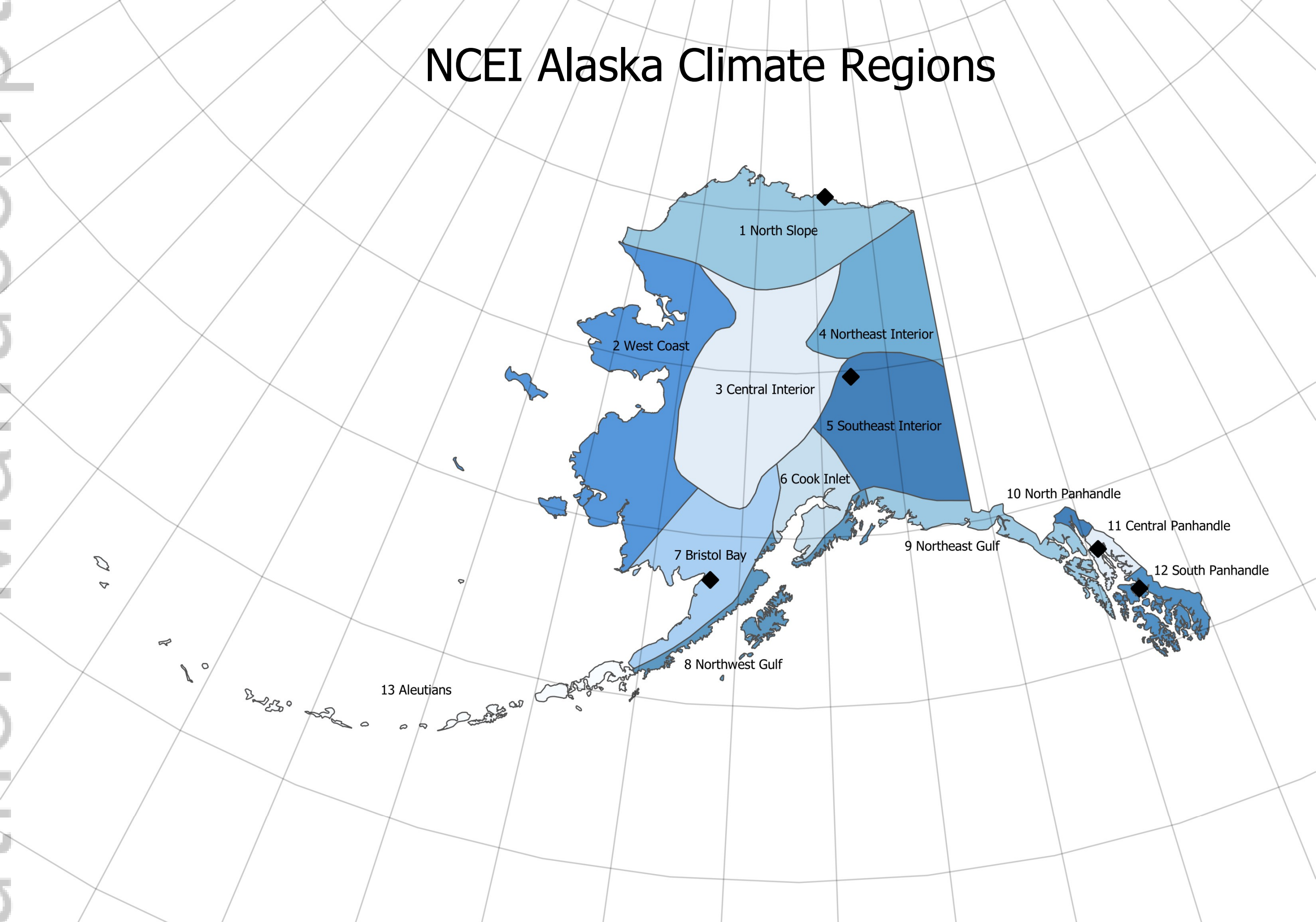


Figure 1: A map of the NCEI climate regions. Diamonds indicate stations that are analyzed later in this paper. Regions from (Bieniek, 2012)

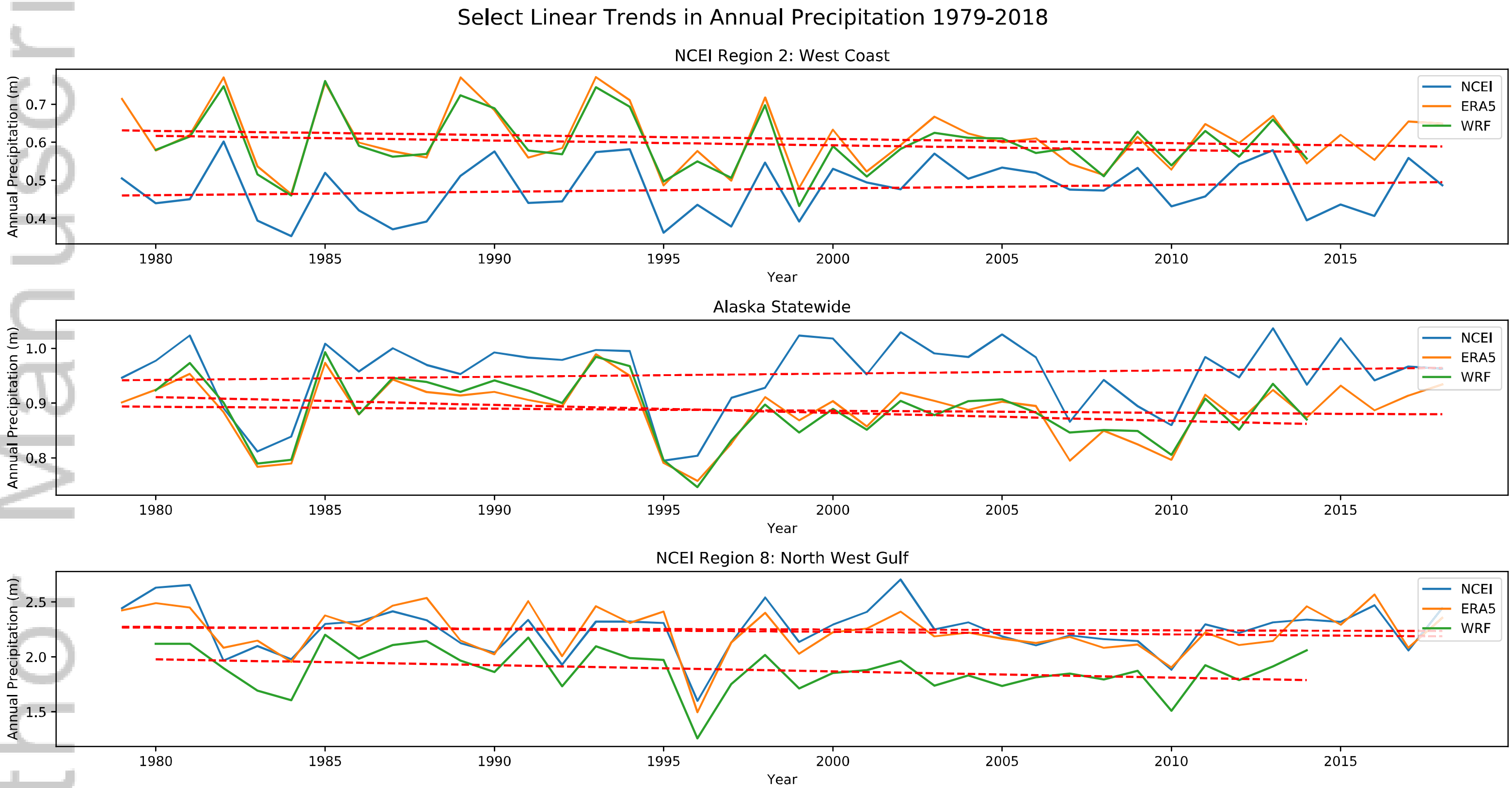


Figure 2: The time series of observational NCEI, UAF downscaled WRF, and ERA5 annual precipitation. Upper panel is for NCEI Region 2 (Alaska west coast), middle panel is for Alaska statewide average, and lower panel is for NCEI Region 8 (northwest Golf of Alaska coast). Data are displayed over 1979-2018 period of reanalysis. Linear regressions were constructed using ordinary least squared method. Regions were chosen to illustrate how linear analysis may indicate differing trends in observations and reanalysis.

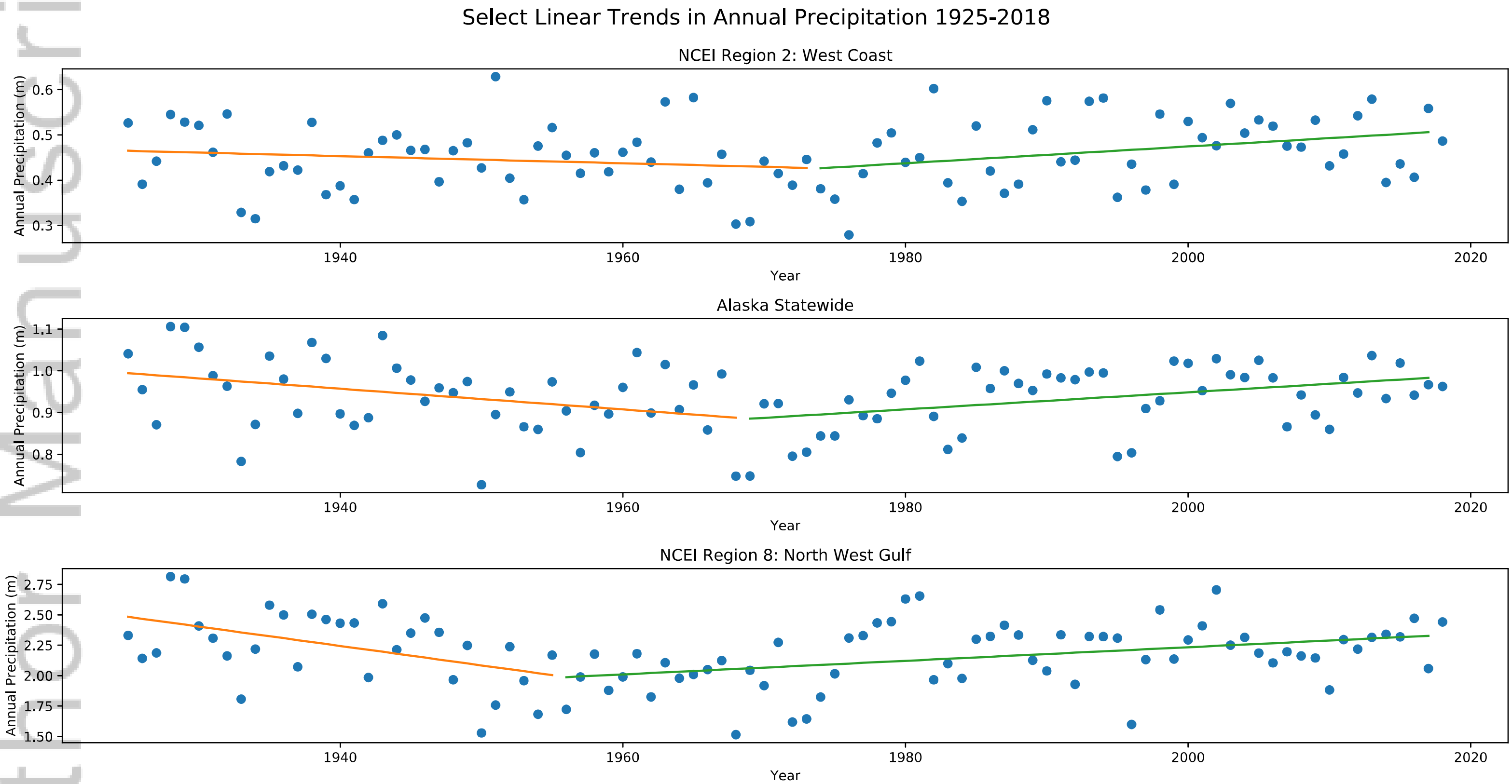


Figure 3: As in Figure 2 but for the entire time period (1925 – 2018) of the NCEI data. A piecewise linear regression model was constructed using methods from (Muggeo, 2016) which determines the validity of a breakpoint by finding the difference in slopes. The fit is improved by the breakpoint analysis.

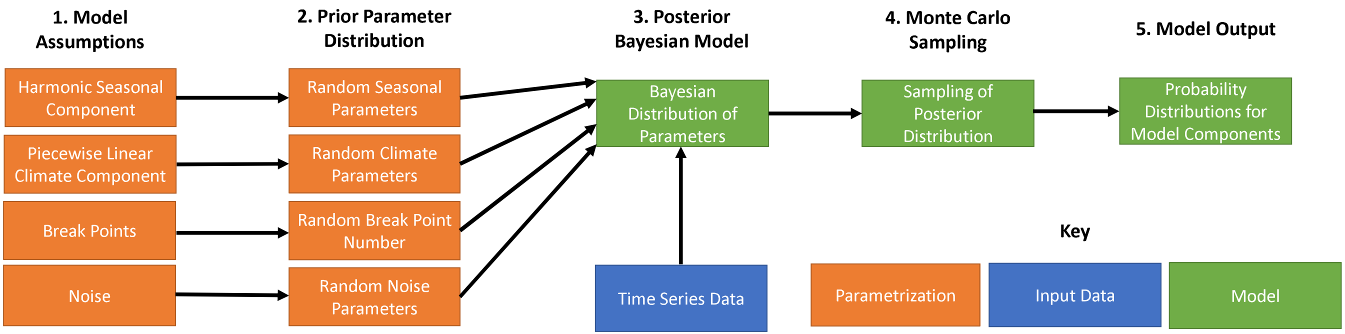


Figure 4: A schematic representation of the Rbeast model.

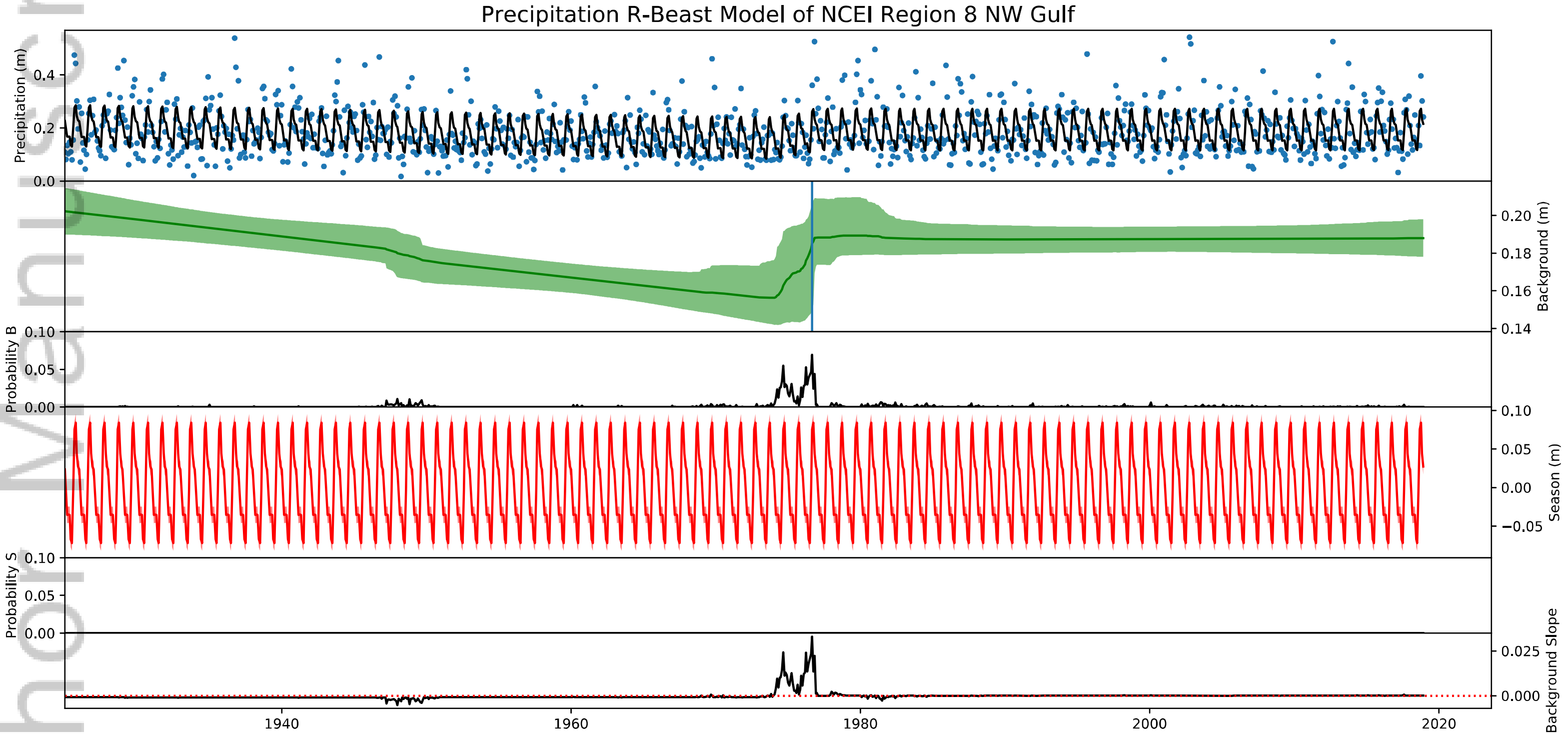


Figure 5: An example of Rbeast output for NCEI region 8 (Northwest Gulf of Alaska) using monthly precipitation totals over the period (1925-2018). The top panel shows the best fit model results alongside the input time series data points. The second panel shows the best fit background component surrounded by a 95% credible interval derived from the sample distribution. Model identified changepoints are marked with vertical blue lines. The third panel shows the probability distribution that any given point represents a changepoint. Note how the peaks align with the vertical blue lines. The next two panels mirror the previous two except they depict the harmonic seasonal component rather than the background climate component. The final panel shows the instantaneous slope of the background component to aid in trend identification.

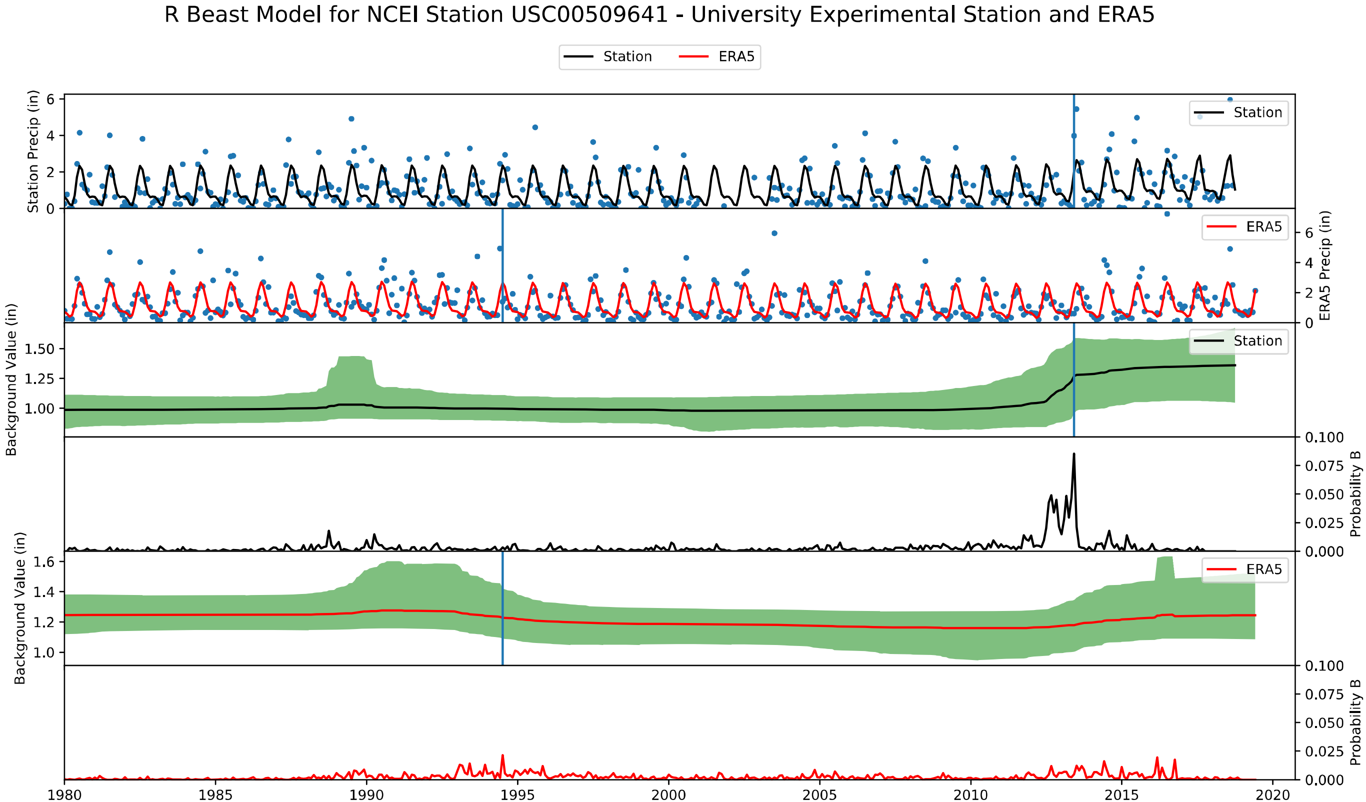


Figure 6: A comparison between the University experimental station and ERA5 reanalysis covering the period (1979-2018). The top panel shows the station data and corresponding Rbeast model fit while the second panel displays the same information for the reanalysis data. The middle two panels show the background component and corresponding changepoint probability distribution for the station data while the bottom two panels show the same data for the reanalysis output. Note that there is some missing data left blank in the station data series in 2003.

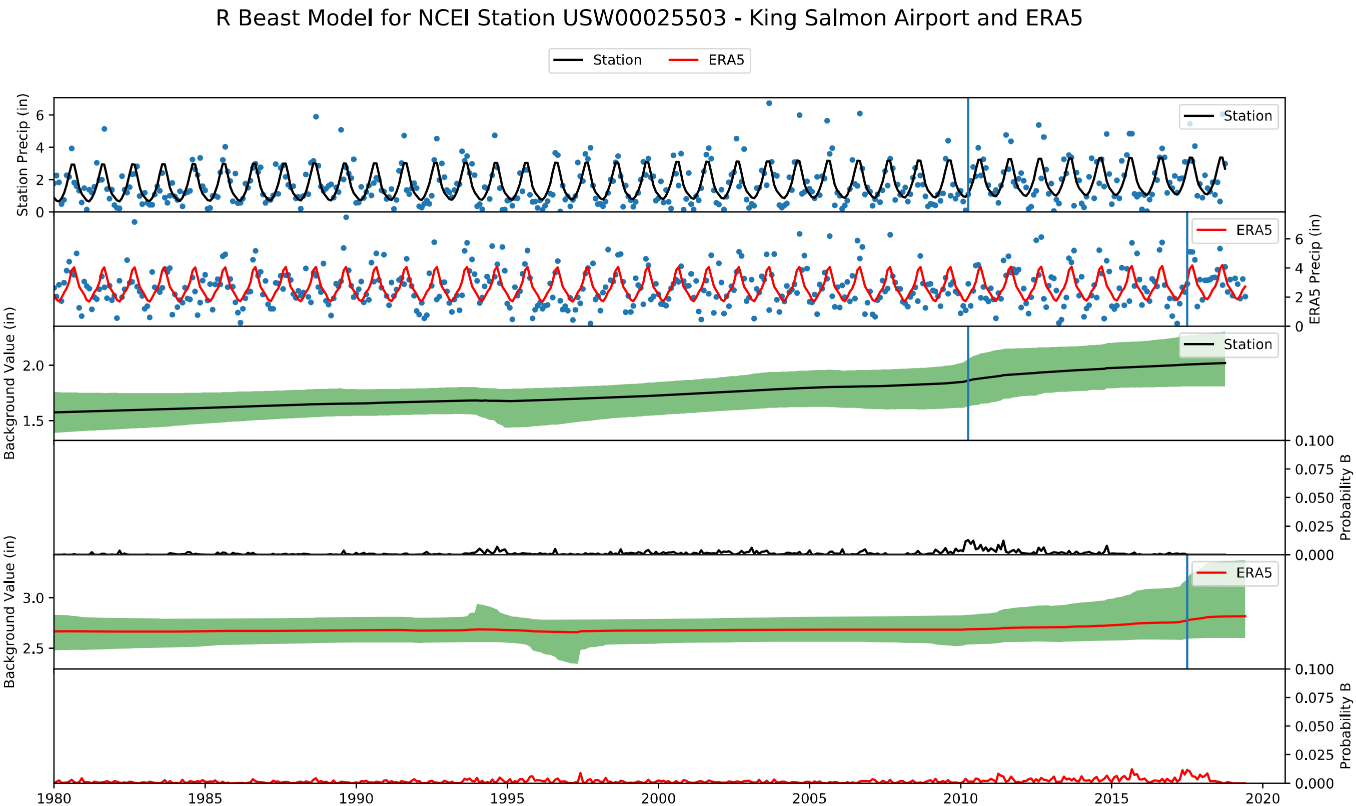


Figure 7: As in Figure 6, but for the King Salmon airport station and ERA5 reanalysis.

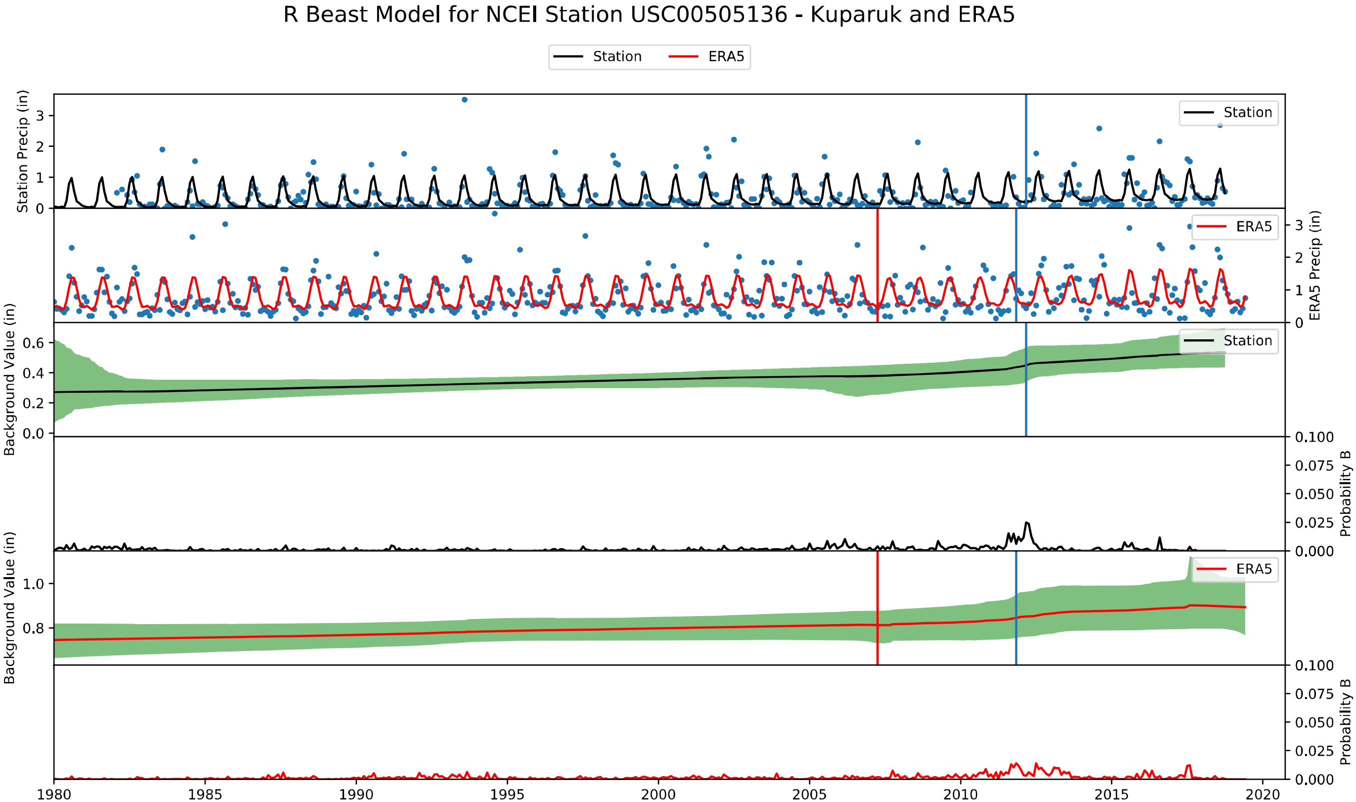


Figure 8: As in Figure 6, but for the Kuparuk station and ERA5 reanalysis. The red line indicates a change-point in the seasonal harmonic component. Note the wide station credible interval near the start of the record is due to missing data.

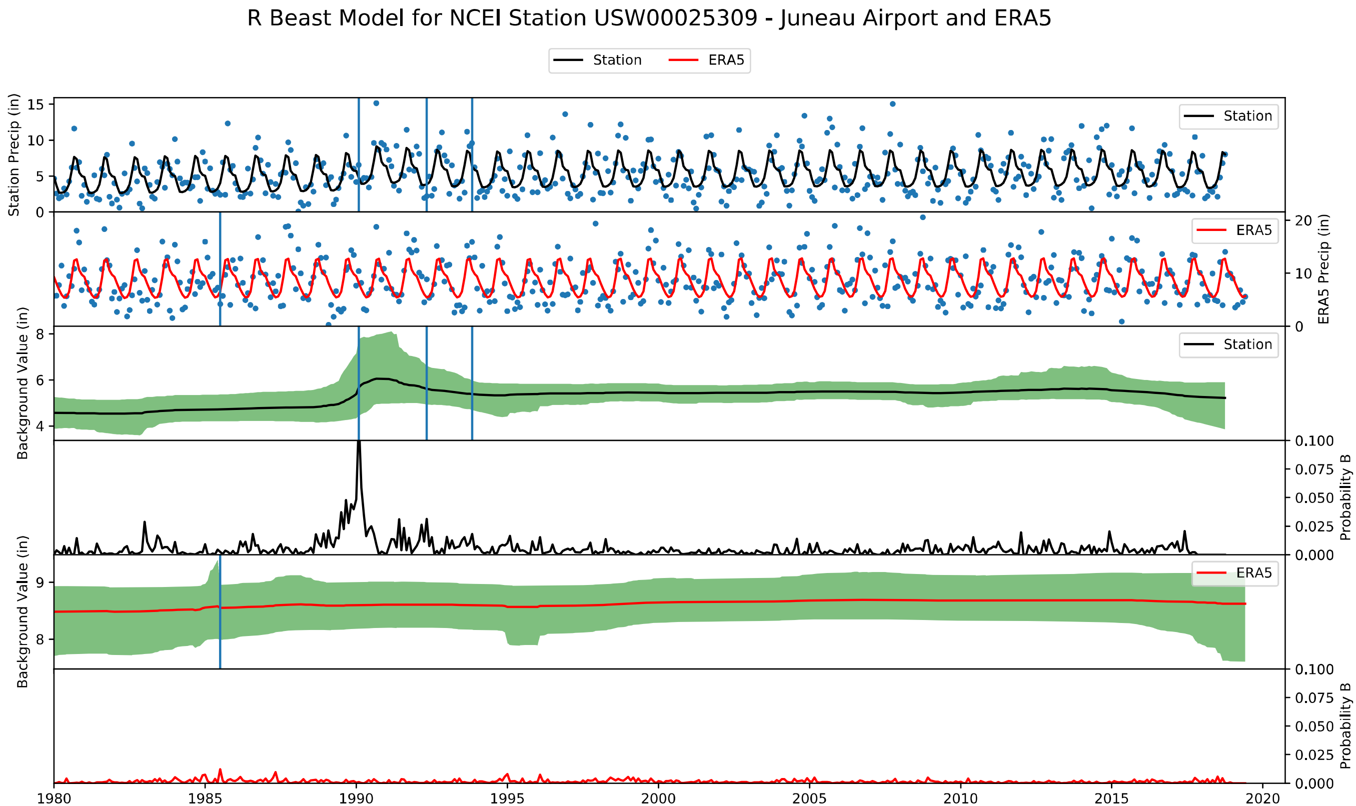


Figure 9: As in Figure 6, but for the Juneau airport station and ERA5 reanalysis.

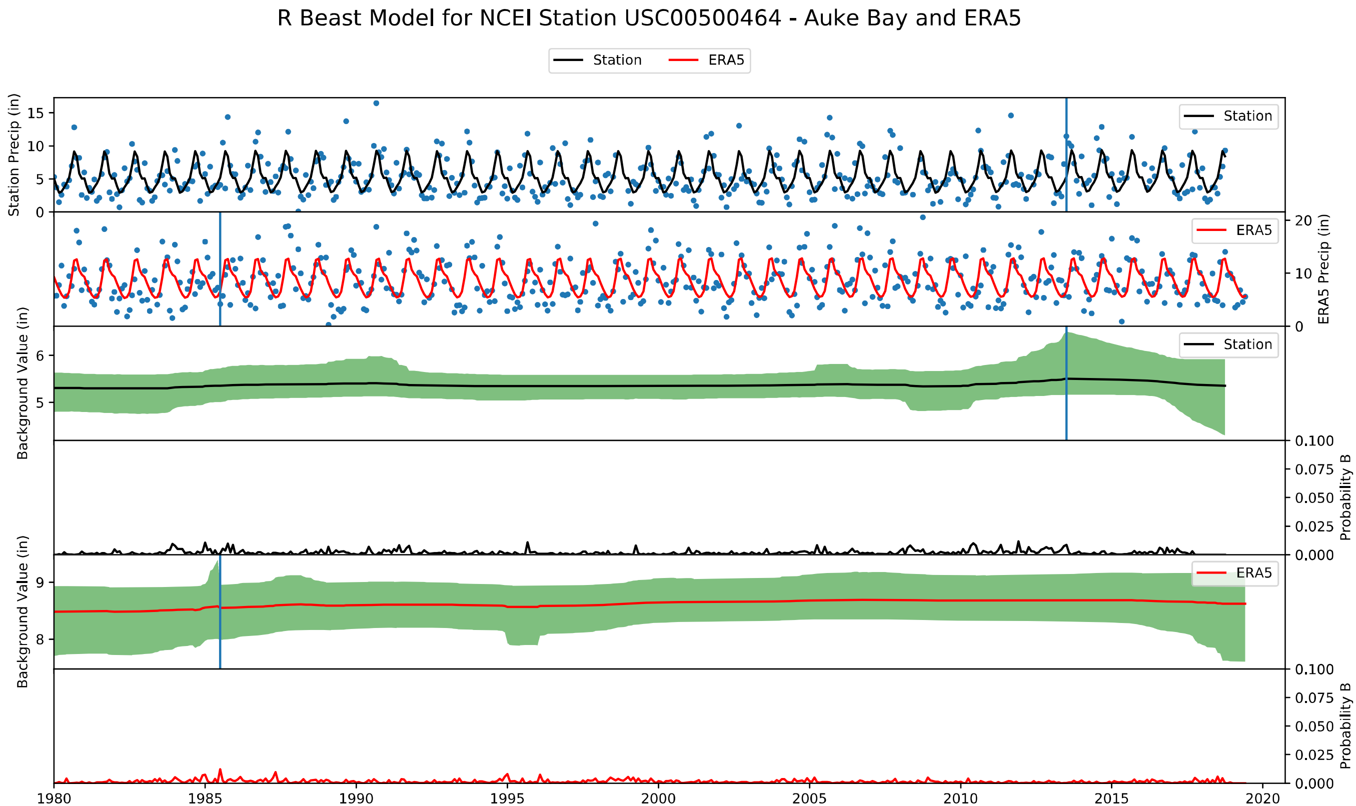


Figure 10: As in Figure 6, but for the Auke Bay station and ERA5 reanalysis.

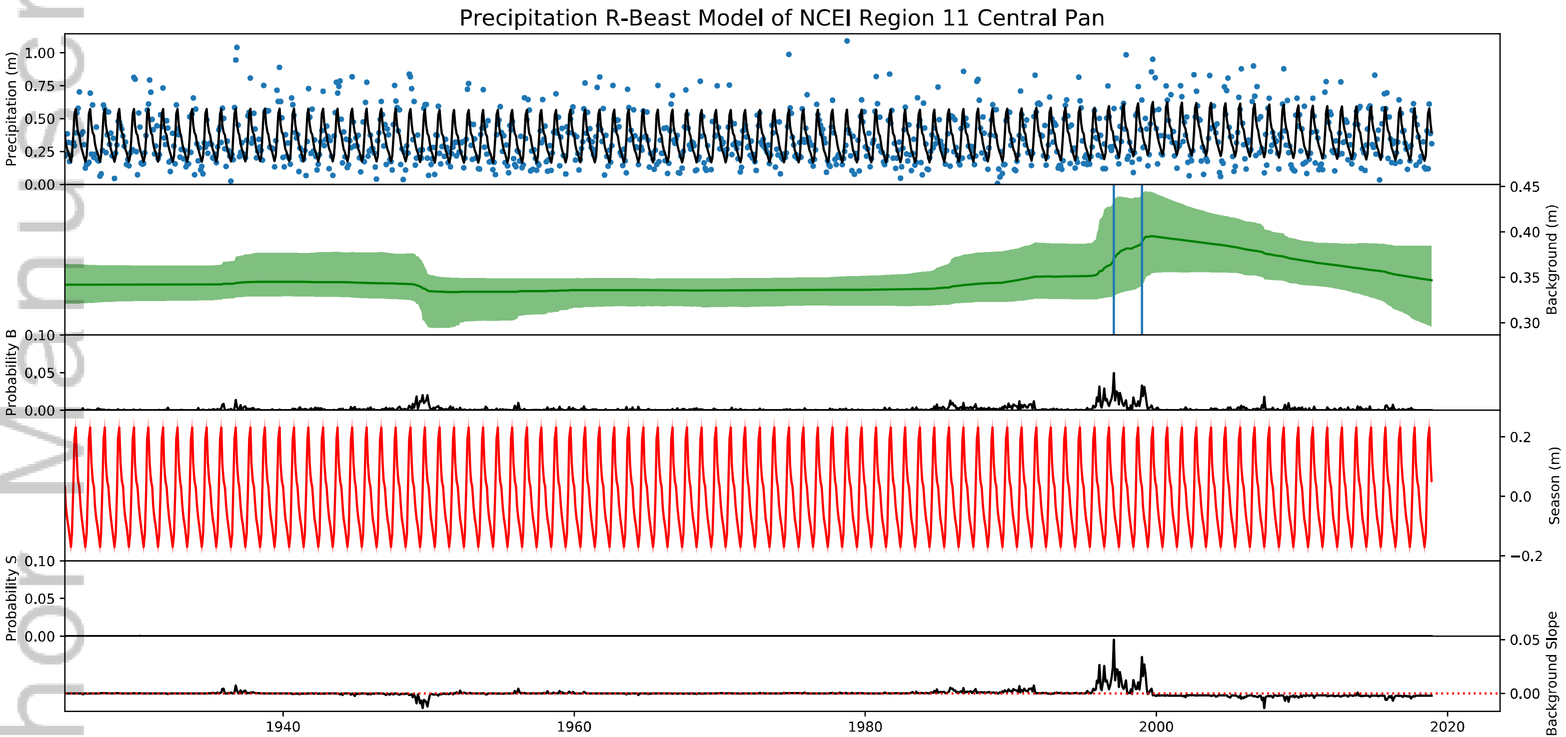


Figure 11: As in Figure 5, but for the Central Panhandle region.

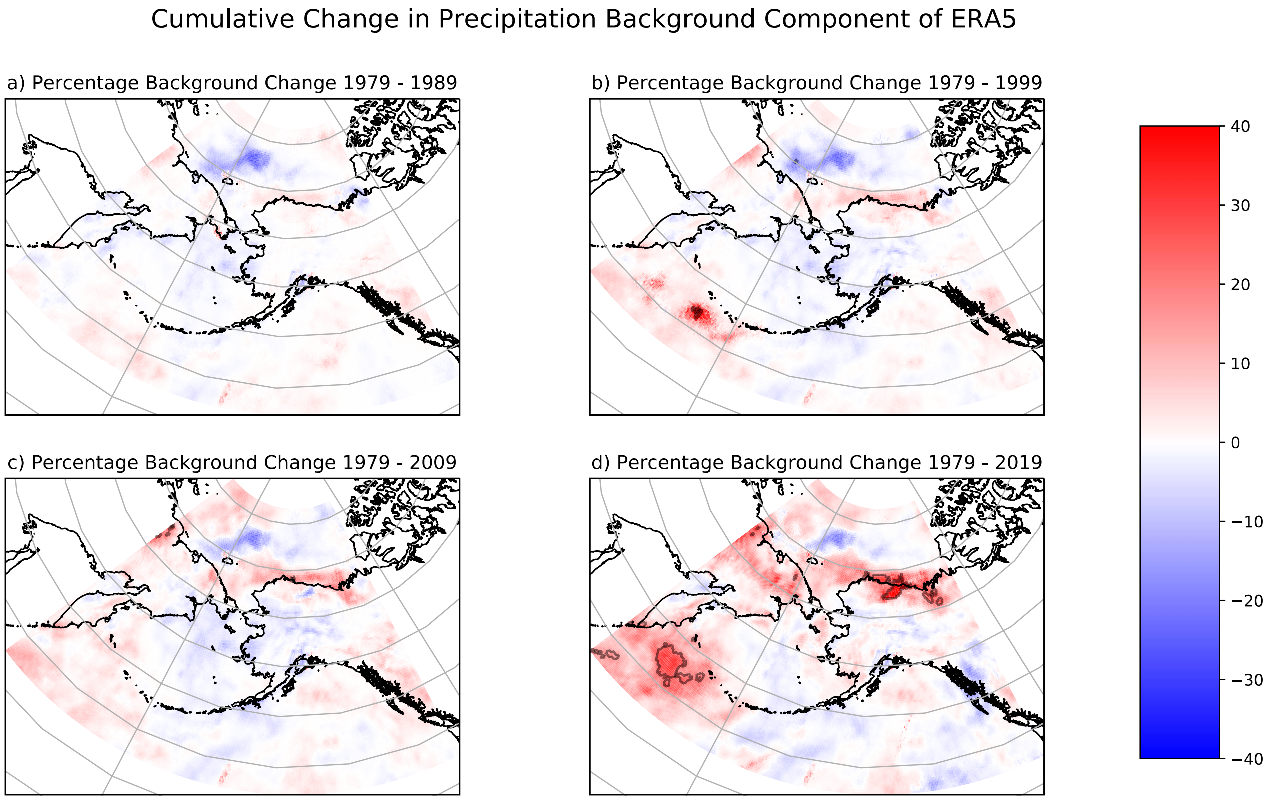


Figure 12: Maps showing percentage change in background precipitation over time according to Rbeast. Each panel shows the percentage change in the background component relative to the record start in 1979 and the end year indicated above the panel. Grid cells where the 95% credible interval no longer overlaps (indicating high confidence in change) are outlined in black.

LS Linear Regression			
Region	Slope	p	Region
	Annual		
1-Nslope	0.10	0.53	1-Nslope
2-Wcoast	0.50	0.11	2-Wcoast
3-Cint	-0.10	0.58	3-Cint
4-NEInt	0.40	0.04	4-NEInt
5-SEInt	0.00	1.00	5-SEInt
6-Cook	-0.60	0.36	6-Cook
7-Bristol	0.20	0.68	7-Bristol
8-NWCoast	-1.00	0.93	8-NWCoast
9-NECoast	-3.20	0.13	9-NECoast
10-Npan	-1.50	0.29	10-Npan
11-Cpan	5.00	0.03	11-Cpan
12-Span	2.20	0.16	12-Span
13-Aluet	-6.00	0.45	13-Aluet
Statewide	0.58	0.52	Statewide

Table 1: A table containing the ordinary least squared regression results applied for the annual NCEI regional precipitation data over the period (1925-2018). Slope is listed in (mm/year) and colored according to its magnitude (green being more positive). P values are calculated via Wald test and represent the confidence in the slope being non-zero. P-values near significance are highlighted in yellow.

Single Breakpoint Regression							
Region	Break p	Break	Slope 1	Slope 2	p1	p2	Region
	Annual						
1-Nslope	0.34		0.11	0.00	0.54		1-Nslope
2-Wcoast	0.09		0.48	0.00	0.10		2-Wcoast
3-Cint	0.10		-0.15	0.00	0.58		3-Cint
4-NEInt	0.08		0.40	0.00	0.04		4-NEInt
5-SEInt	0.05		-0.02	0.00	0.94		5-SEInt
6-Cook	0.39		-0.60	0.00	0.35		6-Cook
7-Bristol	0.01	1991.31	-0.93	4.70	0.15	0.08	7-Bristol
8-NWCoast	0.00	1956.00	-16.55	5.52	0.00	0.00	8-NWCoast
9-NECoast	0.01	1954.00	-34.51	5.74	0.01	0.10	9-NECoast
10-Npan	0.01	1971.98	-9.43	6.45	0.01	0.08	10-Npan
11-Cpan	0.03	1971.42	-7.12	16.67	0.26	0.01	11-Cpan
12-Span	0.28		1.96	0.00	0.21		12-Span
13-Aluet	0.20		-0.59	0.00	0.49		13-Aluet
Statewide	0.00	1969.00	-2.48	2.04	0.01	0.01	Statewide

Table 2: A table containing the breakpoint regression results from (Muggeo, 2016) applied for the annual NCEI regional precipitation data over the period (1925-2018). Slope is listed in (mm/year) and colored according to its magnitude (green being more positive). “Break P” represents the significance of a breakpoint where “P1” and “P2” represent the confidence of the first and second slope respectively being non-zero. P values are colored according to significance (greener values being more significant).