Journal of Data Science 24 (2), 394–410
April 2026
DOI: 10.6339/26-JDS1220
Data Science in Action

# Investigating Spatial Dependence in the Degree of Asymptotic Dependence between a Satellite Precipitation Product and Station Data in the Northern US Rocky Mountains

Brook T. Russell[1,*], Kelie Marline Momo Nizegha[1], Dulshan Malshika[1], and Whitney K. Huang[1]

[1]*School of Mathematical and Statistical Sciences, Clemson University, Clemson, South Carolina, USA*

**Abstract**

Satellite precipitation products have the potential to be employed for the purpose of better understanding extreme precipitation events in remote mountainous terrain, where weather stations and radar data tend to be sparse. For this reason, it is crucial to assess how closely satellite estimates agree with ground observations during extreme events, and how that agreement varies across such regions. We use asymptotic dependence from multivariate extreme value theory as the primary tool in this study. After presenting two measures of asymptotic dependence and their associated estimators, we illustrate these ideas using simulated data. We then model the level of asymptotic dependence between PERSIANN-CDR and SNOTEL station data over the US Northern Rocky Mountains. We consider both asymptotic dependence estimators, and based on hypothesis tests and visual diagnostics, both estimates of asymptotic dependence indicate positive spatial dependence. We also investigate whether geographical factors influence the levels of asymptotic dependence over this region. Using a spatial correlation analysis, we find that elevation is negatively correlated with both asymptotic dependence estimators and average summer temperature is positively correlated with both asymptotic dependence estimators. However, we did not find any geographical covariates to be statistically significant in the model.

**Keywords** *bivariate regular variation; extreme value theory; PERSIANN-CDR; Snow Telemetry Station Network*

## 1 Introduction

Satellite precipitation products (SPP) make it possible to reliably reproduce surface precipitation globally. These SPP often perform well in terms of characterizing typical precipitation events, but have been shown to model extreme precipitation events less well (AghaKouchak et al., 2011). One way to assess this capability is through the lens of asymptotic dependence. Asymptotic dependence is a concept from extreme value theory (EVT) — the branch of statistics the aims to model the far upper tails of response distributions — that seeks to model the degree to which the components of a random vector are able to be at extreme levels simultaneously (Coles, 2001, Section 8.4).

In this work, we build on the research of Russell et al. (2024), and similarly choose to focus on one particular SPP, PERSIANN-CDR (Ashouri et al., 2015). Although the findings presented

*Corresponding author. Email: brookr@clemson.edu.

© 2026 The Author(s). Published by the School of Statistics and the Center for Applied Statistics, Renmin University of China. Open access article under the CC BY license.
Received August 15, 2025; Accepted February 11, 2026

here may or may not generalize to other SPPs, we believe the results are interesting and informative, and that the approach can be applied to other spatial regions. To this end, we model the ability of PERSIANN-CDR to capture extreme precipitation events in the remote US northern Rocky Mountains of Wyoming, Idaho, and Montana. We choose this region as other works have shown that the ability of SPP to model EPE is diminished over high elevation mountainous terrain (Bharti and Singh, 2015; Hirpa et al., 2010; Derin and Yilmaz, 2014), making for an interesting case study. As the density of weather stations is sparse throughout this remote region, we leverage data from the Snow Telemetry (SNOTEL) network of stations as the ground truth.

We pursue three related research objectives in this work. First, we assess the degree to which PERSIANN-CDR is asymptotically dependent with Snow Telemetry (SNOTEL) precipitation data over the study region using two well-known asymptotic dependence estimators, namely the extremal dependence coefficient $\chi$ (Coles et al., 1999), the most widely used measure under asymptotic dependence due to its clear interpretation and broad adoption, and $\bar{\gamma}$ (Russell et al., 2024), which offers improved angular resolution and more stable finite-sample performance. We make the a priori assumption that these two variables exhibit asymptotic dependence rather than the degenerate asymptotic independence case. This assumption is reasonable, given that the probability of PERSIANN-CDR detecting an extreme precipitation event (EPE) is not expected to be zero.

In this context, while alternative extremal dependence measures have been proposed (see, e.g., Cooley et al., 2012 and Section 2 of Dey et al., 2016 for a more comprehensive review), and extremal dependence has also been examined in spatial network settings (Huang et al., 2019; Engelke and Hitz, 2020; Gong et al., 2024), we restrict our analysis to an asymptotic dependence framework. More broadly, there is a growing literature on multivariate and spatial extreme-value models that accommodate both asymptotic dependence and asymptotic independence, thereby relaxing the asymptotic dependence assumption (see the survey by Huser and Wadsworth, 2022 and recent work on geometric extremes in Wadsworth and Campbell, 2024). Much of this work, however, emphasizes model development rather than the proposal of new dependence summaries, with dependence properties instead derived from fitted models. Consequently, our conclusions are not intended to depend sensitively on the specific dependence measure used; alternative extremal dependence measures are therefore not pursued in detail in order to avoid unnecessary complexity.

Second, we examine whether the resulting asymptotic dependence estimates exhibit spatial dependence across the region by conducting hypothesis tests and fitting spatial models. Specifically, we investigate the spatial structure of SNOTEL/PERSIANN-CDR tail-dependence summaries, assessing whether dependence levels display geographic clustering and how they relate to broad physiographic and climatological gradients. This spatial perspective provides scientific insight beyond treating station-level dependence estimates as spatially independent. Here, "spatial dependence" refers to spatial correlation among estimated dependence measures across stations, rather than the spatial occurrence of extreme precipitation events happening simultaneously across multiple locations. Finally, we examine whether a set of geographic covariates can explain variability in the estimated asymptotic dependence measures. Together, these objectives build on and extend the findings of Russell et al. (2024).

This manuscript is organized in the following manner. In Section 2 we begin by providing the reader a brief overview of two common asymptotic dependence parameters, and then give the corresponding estimators. We conclude Section 2 by illustrating these ideas by exploring simulated data. We open Section 3 by describing the data used in our analysis, and conclude by addressing our research aims. We conclude the paper by providing a discussion in Section 4.

## 2 Bivariate Asymptotic Dependence

### 2.1 Asymptotic Dependence Parameters

Assume that we have a bivariate random vector $(Y_1, Y_2)^T$. Its components are deemed to be asymptotically dependent (or tail dependent) if the parameter

$$\chi = \lim_{u \to 1} P\big(F_2(Y_2) > u \mid F_1(Y_1) > u\big) > 0. \tag{1}$$

If (1) does not hold, we say that the components of the random vector are asymptotically independent. In essence, the components will be asymptotically dependent if there is a non-zero probability that they can both be at extreme levels at the same time.

The parameter $\chi$ defined in (1) serves as one of the most ubiquitous asymptotic dependence parameters (Coles, 2001, Section 8.4). However, several other useful asymptotic dependence parameters are used in practice. In this work we make use of $\chi$, but we also employ an asymptotic dependence parameter based on the works of Russell et al. (2016) and Russell et al. (2024). This parameter is a special case of an extremal dependence measure (EDM) (Resnick, 2004; Larsson and Resnick, 2012), and relies on the framework of multivariate regular variation from EVT. Multivariate regular variation is an important and useful framework from the world of multivariate extremes; the interested reader is referred to Resnick (2007) for details.

In the regular variation framework, multivariate observations are typically transformed into pseudo-polar coordinates and the resulting radial and angular components are utilized in further analysis. The level of asymptotic dependence, assuming that the data generating process is not in the degenerate asymptotic independence case, can be fully characterized by the angular measure. The angular measure, denoted $H$, can be thought of as providing a distribution for the angle of extreme observations over the unit interval. As a special case of the EDM of Larsson and Resnick (2012), we also consider the asymptotic dependence parameter suggested by Russell et al. (2024),

$$\bar{\gamma} = 1 - \int_{[0,1]} |2w - 1| \, dH(w). \tag{2}$$

Here, $w \in [0, 1]$ represents the angular component in the pseudo-polar decomposition.

Importantly, the asymptotic dependence parameter $\chi$ in (1) and $\bar{\gamma}$ in (2) are both constrained to take on values in the unit interval, and larger values of both imply a higher degree of asymptotic dependence. For both parameters, a value of 1 corresponds with perfect asymptotic dependence while a value of 0 corresponds to the degenerate asymptotic independence case.

### 2.2 Asymptotic Dependence Estimators

The parameter $\chi$ has a straightforward estimator based on the sample analogue of (1), given by

$$\hat{\chi}(u) = \frac{\sum_{t=1}^{N} I\{\hat{F}_1(Y_{t,1}) > u\} I\{\hat{F}_2(Y_{t,1}) > u\}}{\sum_{t=1}^{N} I\{\hat{F}_1(Y_{t,1}) > u\}}, \tag{3}$$

where $Y_{t,1}$ and $Y_{t,2}$ are iid realizations of $Y_1$ and $Y_2$, $I(\cdot)$ is the indicator function, $u \to 1$ is a high probability level, $F_1$ and $F_2$ are the distribution functions of $Y_1$ and $Y_2$, $\hat{F}_1$ and $\hat{F}_2$ are the corresponding estimated distribution functions, and $N$ is the sample size. The distribution functions $\hat{F}_1$ and $\hat{F}_2$ are typically taken to be empirical distribution functions, which implies that the estimator in (3) is simply an empirical proportion.

Now consider the sequence of iid bivariate regularly varying random vectors

$$\boldsymbol{Z_t} = (Z_{t,1}, Z_{t,2})^T, \quad t = 1, \ldots, N,$$

obtained by marginally transforming $(Y_{t,1}, Y_{t,2})^T$ to a common marginal (e.g., unit Fréchet). This marginal standardization allows the joint tail behavior of the data to be described using the framework of multivariate regular variation. We define the pseudo-polar coordinate transformation where the radial component $R_t = ||\boldsymbol{Z_t}||_1$ and angular component $W_t = R_t^{-1} Z_{t,1}$.

In practice, the marginal transformations are estimated using the empirical distribution functions of $Y_1$ and $Y_2$. The pseudo-polar representation is then applied to observations exceeding a high radial threshold $r_0$, in direct analogy to estimating $\chi$ using exceedances above a high probability level $u$. We can then estimate the asymptotic dependence parameter $\bar{\gamma}$ via

$$\hat{\bar{\gamma}}(r_0) = 1 - \int_0^1 |2w - 1| \, d\hat{H}(w) = \frac{\sum_{t=1}^N I(R_t > r_0)|2W_t - 1|}{\sum_{t=1}^N I(R_t > r_0)}, \tag{4}$$

where $r_0$ is the threshold for the radial components and $\hat{H}$ is the empirical estimator of the angular measure $H$.

Threshold selection in extreme-value analysis is both critical and challenging, and reflects a familiar bias–variance tradeoff. Lower thresholds yield more observations and hence lower variance, but may introduce bias by including non-extreme values. Conversely, higher thresholds reduce bias by focusing on more extreme observations, but increase variance due to smaller effective sample sizes.

A common practical strategy is the *threshold stability* approach, in which the threshold is chosen beyond the point where parameter estimates appear stable. This idea is well established in the univariate setting (Coles, 2001, Section 4.3) and becomes more challenging in the multivariate context (Resnick, 2007, Section 9.2.4). In practice, the threshold $r_0$ is often selected as a high empirical quantile $\hat{Q}(u)$ of the observed radial components. The resulting estimator, denoted by $\hat{\bar{\gamma}}(\hat{Q}(u))$, can be viewed as a surrogate for asymptotic dependence and is adopted here as a pragmatic choice.

## 2.3 Illustrating via Simulation

In order to help the reader get a better feel for the idea of asymptotic dependence and the ways in which it can be estimated, we leverage simulated data using `R`. For illustrative purposes, we first work with two simulated data sets. Both data sets are simulated from a bivariate extreme value distribution with logistic dependence structure (Coles, 2001, Section 8.2). The logistic dependence function has one dependence parameter, $\alpha \in (0, 1)$, and is a relatively simple model that allows for a wide range of asymptotic dependence levels. Importantly, a lower value of $\alpha$ implies a higher degree of asymptotic dependence for this dependence model. For illustrative purposes, we choose two parameter settings: $\alpha = 0.35$ for the first simulated data set and $\alpha = 0.85$ for the second simulated data set. The marginal distributions are taken to be unit Gumbel, which is a special case of the generalized extreme value (GEV) distribution. As is typically done in analyses employing the framework of regular variation, the realizations are first transformed to have unit Frèchet marginal distributions. The unit Frèchet distribution is another special case of the GEV distribution, and is commonly utilized for this purpose in multivariate extremes. We also extract the empirical angular components from the unit Frèchet

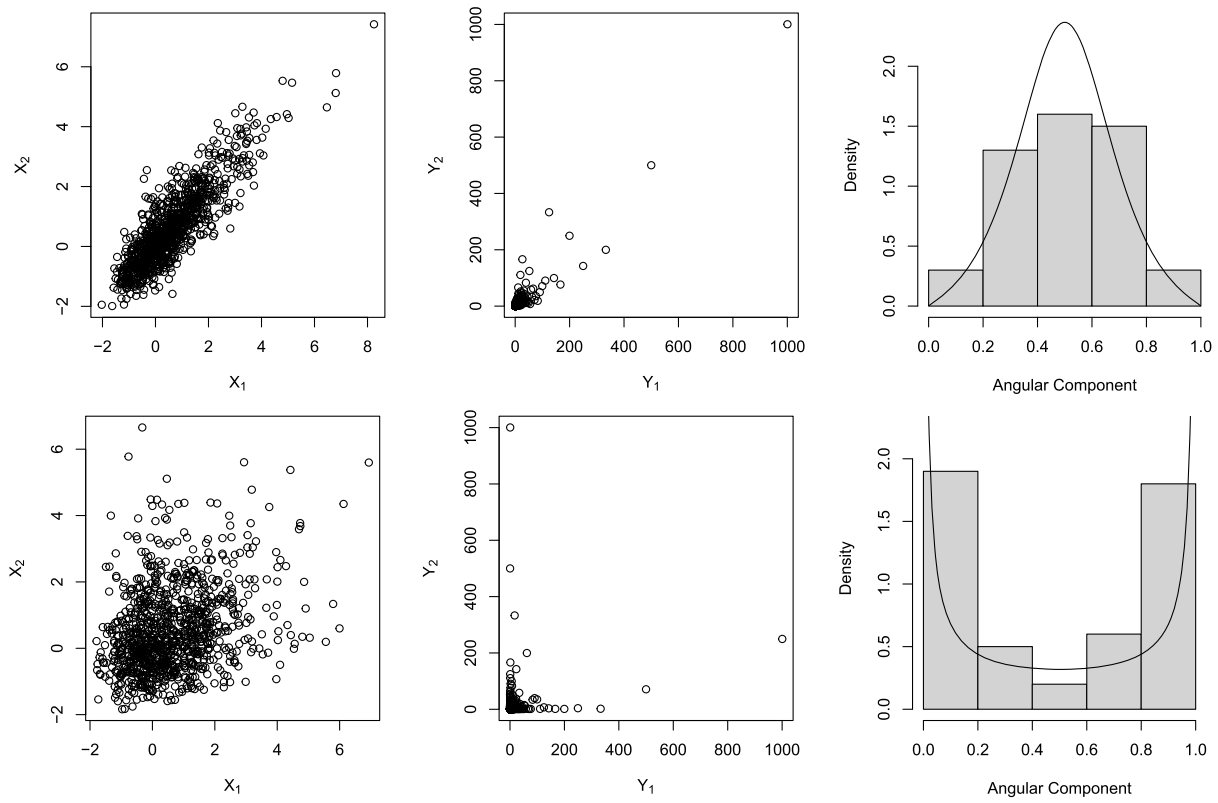


Figure 1: First row: realizations on the Gumbel scale (left), unit Fréchet scale (middle), and a histogram of empirical angles for the top 5% of radial values with the true angular density (right); second row: analogous plots for the second simulated dataset.

transformed data, emphasizing that these data should provide an empirical estimator of the true angular measure $H$.

Figure 1 plots the data from both simulated data sets. Plots of realizations from the process with the higher level of asymptotic dependence are shown in the top row of Figure 1, while the bottom row displays plots using data from the process with the lower degree of asymptotic dependence. The scatterplot in the top left panel plots the data on the original (Gumbel) scale, and is consistent with a relatively high degree of asymptotic dependence as the largest observations for $X_1$ coincide with the largest observations for $X_2$. Using a probability integral transformation (PIT) to transform the marginal distributions to unit Frèchet results in the plot in the top center panel. Because the unit Frèchet distribution is extremely heavy tailed, the vast majority of the points are mapped very close to the origin, highlighting the extreme observations. Again, it is evident from this plot that the largest observations for $Y_1$ coincide with the largest observations for $Y_2$. Extracting the empirical angular component for the points with the top 5% empirical radial components yields the histogram in the top right panel. The fact that the largest observations for $X_1$ coincide with the largest observations for $X_2$ results in many empirical angular components close to 0.5, which is evident in this histogram. For reference, the true angular measure's density function is also plotted on this histogram.

In contrast, the second data set is generated with $\alpha = 0.85$, implying that the degree of asymptotic dependence is relatively low. The data on the original (Gumbel) scale are plotted

in the bottom left panel, and the relatively low degree of asymptotic dependence is evident as the largest observations for $X_1$ do not tend to coincide with the largest observations for $X_2$. Transforming the marginal distributions to unit Frèchet results in the plot in the bottom center panel. Most of the points with large radial components are close to the axes, which is consistent with the largest observations for $Y_1$ not coinciding with the largest observations for $Y_2$. Extracting the empirical angular component for the points with the top 5% empirical radial components results in the histogram in the bottom right panel, and high degree of density close to 0 and 1 coincides with the largest observations for $X_1$ not coinciding with the largest observations for $X_2$. The true angular measure's density function is plotted on this histogram as well; visual inspection reveals a noticeably different shape for this density function compared to that of the other process.

Not surprisingly, the values of the estimators defined in (3) and (4) also reflect what we observe in Figure 1. For the first simulated data set, plotted in the top row of Figure 1, $\hat{\bar{\gamma}}(0.95) \approx 0.70$ and $\hat{\chi}(0.95) \approx 0.63$. For the second simulated data set, plotted in the bottom row of Figure 1, $\hat{\bar{\gamma}}(0.95) \approx 0.18$ and $\hat{\chi}(0.95) \approx 0.23$. Interestingly, for the logistic dependence model it can be shown that $\chi = 2 - 2^{\alpha}$, which implies that $\chi = 0.73$ for the first data set and $\chi = 0.20$ for the second data set. As the estimators are much higher for the first data set, they are consistent with a higher degree of asymptotic dependence in the first data set. Finally, it is important to point out that these estimators, while both do a reasonable job of reflecting asymptotic dependence levels, do not estimate the same parameter value, and should not be directly compared.

In order to better understand the sampling distribution for these estimators for various sample sizes, we simulate bivariate data sets of sizes ranging from 1,250 to 10,000. For each sample size, we simulate 10,000 samples and calculate $\hat{\bar{\gamma}}$ and $\hat{\chi}$ for each. The code for this simulation study is available in the GitHub repository located at the URL https://github.com/brooktrussell/InvestInvestigatingSpatialDependence. Figure 2 gives boxplots of the 10,000 estimates for each combination of sample size, asymptotic dependence estimator, and true asymptotic dependence parameter are shown for $\hat{\bar{\gamma}}$ (left column) and $\hat{\chi}$ (right column), based on simulations from a bivariate extreme value distribution with unit Fréchet marginals and logistic dependence ($\alpha = 0.35$ in the top row and $\alpha = 0.85$ in the bottom row). For reference, each true asymptotic dependence parameter is plotted with a horizontal line. Under the logistic dependence function, the parameter $\chi = 2 - 2^{\alpha}$ is straightforward to calculate. The parameter $\bar{\gamma}$ is less straightforward to calculate and therefore the right hand side of Equation (2) is integrated numerically using `R`.

## 3 Asymptotic Dependence Levels in the Northern Rockies

### 3.1 Description of the Data

The North American Rocky Mountains is a mountain range that extends from the US state of New Mexico to the Canadian provinces of Alberta and British Columbia. The US Northern Rocky Mountains, which includes Wyoming, Idaho, and Montana, are not as quite as high as in Colorado, but are equally rugged and even more remote. For reference, Figure 3 plots the elevation of the northern Rocky Mountain states of Wyoming, Idaho, and Montana using the grid and elevation data from the PRISM data product (Daly et al., 1997). Over this region, winter temperatures are extremely cold, while summer days can be quite warm with cool night time temperatures. The region has a much lower density of weather stations compared to the eastern US, and therefore it is challenging to identify gauge data to use to ground truth SPP data. For

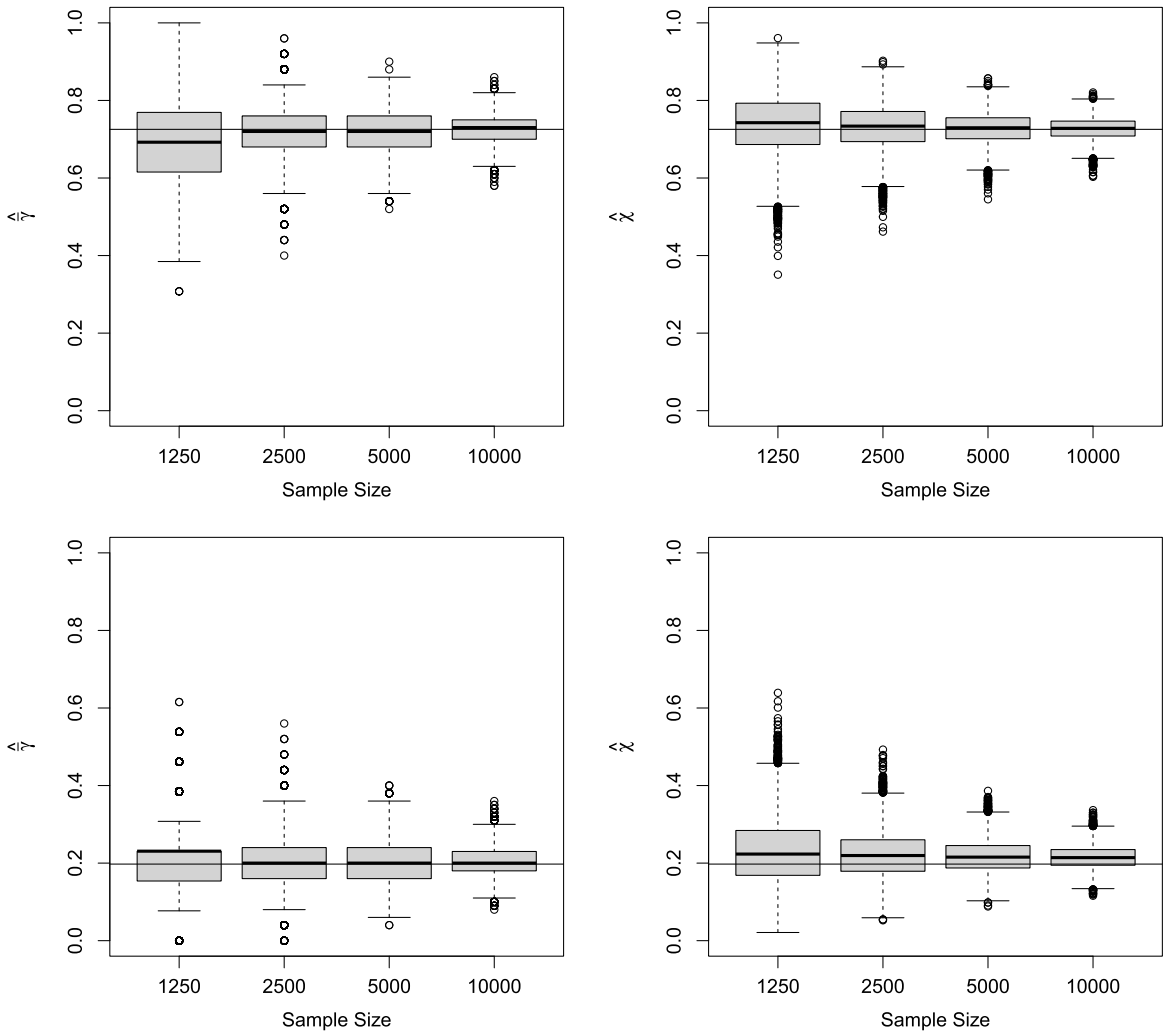


Figure 2: Boxplots of 10,000 estimates for each combination of sample size, asymptotic dependence estimator, and true asymptotic dependence parameter are shown for $\hat{\hat{\gamma}}$ (left column) and $\hat{\chi}$ (right column), based on simulations from a bivariate extreme value distribution with unit Fréchet marginals and logistic dependence ($\alpha = 0.35$ in the top row and $\alpha = 0.85$ in the bottom row). For reference, each true asymptotic dependence parameter is plotted with a horizontal line.

this reason, we employ data from the SNOTEL network that can be found at https://www.nrcs.usda.gov/resources/data-and-reports/snow-and-water-interactive-map and is funded by the US Department of Agriculture. Stations in the SNOTEL network are unmanned autonomous stations that are located in remote mountainous areas. Monitoring snow pack data is the primary mission of the SNOTEL station network, but many of the stations also collect temperature and precipitation data year-round.

Our primary objective is to characterize a SPP's ability to reproduce summer (defined as June, July, and August) EPE in the US northern Rocky Mountains, and to better understand how geographical factors may be related to this ability. For this purpose, we employ data from the years ranging from 2005 to 2015 in the US states of Wyoming, Idaho, and Montana. We use

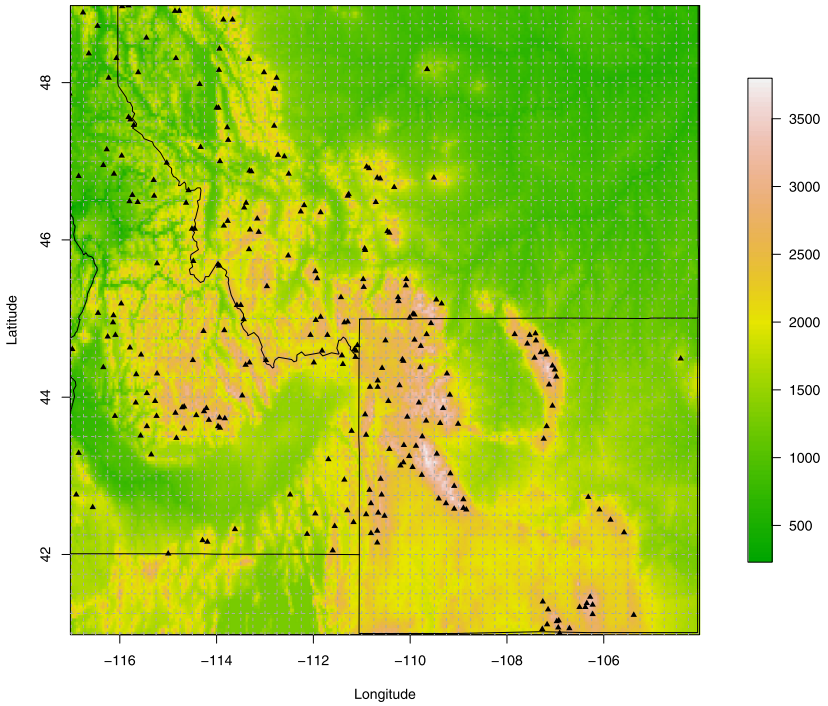


Figure 3: SNOTEL stations in Wyoming, Montana, and Idaho with highlighted PERSIANN-CDR grid cells; background shows elevation (m).

the SPP PERSIANN-CDR which uses data processing approaches to reproduce surface precipitation based on satellite observations. PERSIANN-CDR is calculated on a resolution of 0.25° at latitudes between $60°S$ and $60°N$. Figure 3 plots all 261 Wyoming, Idaho, and Montana SNOTEL station locations, with the PERSIANN-CDR grid cells indicated by gray lines. Data from the SNOTEL network of stations are point-referenced, whereas PERSIANN-CDR is a gridded data product. The raw data that we use in this analysis are the same as analyzed in Russell et al. (2024), and include 11 years worth of daily data for a total of 4015 observations. This manuscript analyzes summer data exclusively, resulting in a total of 1,012 data points for a station without missing observations. Whereas Russell et al. (2024) employ a Gibbs posterior based approach to do analysis within the tail dependence regression framework, we take a different approach, and focus on comparing asymptotic dependence estimators and better understanding their spatial distribution.

The focus of this work is to better understand dependence in the far upper joint tail of these two heterogeneous data products; therefore, we purposefully do not employ any methods to deal with spatial misalignment as these approaches may attenuate the presence of asymptotic dependence. Furthermore, better understanding the nature of degree to which PERSIANN-CDR is associated with EPEs may be able to provide information about precipitation related natural disasters, such as flash flooding. We also emphasize that we conduct this analysis through the framework of EVT, which may have provide additional insights compared to a traditional correlation analysis. It is important to keep in mind that the underlying model is assuming that the bivariate realizations are independent, when in reality this is likely not the case. However, we believe that our analysis still conveys useful information as temporal dependence in daily precipitation is gen-

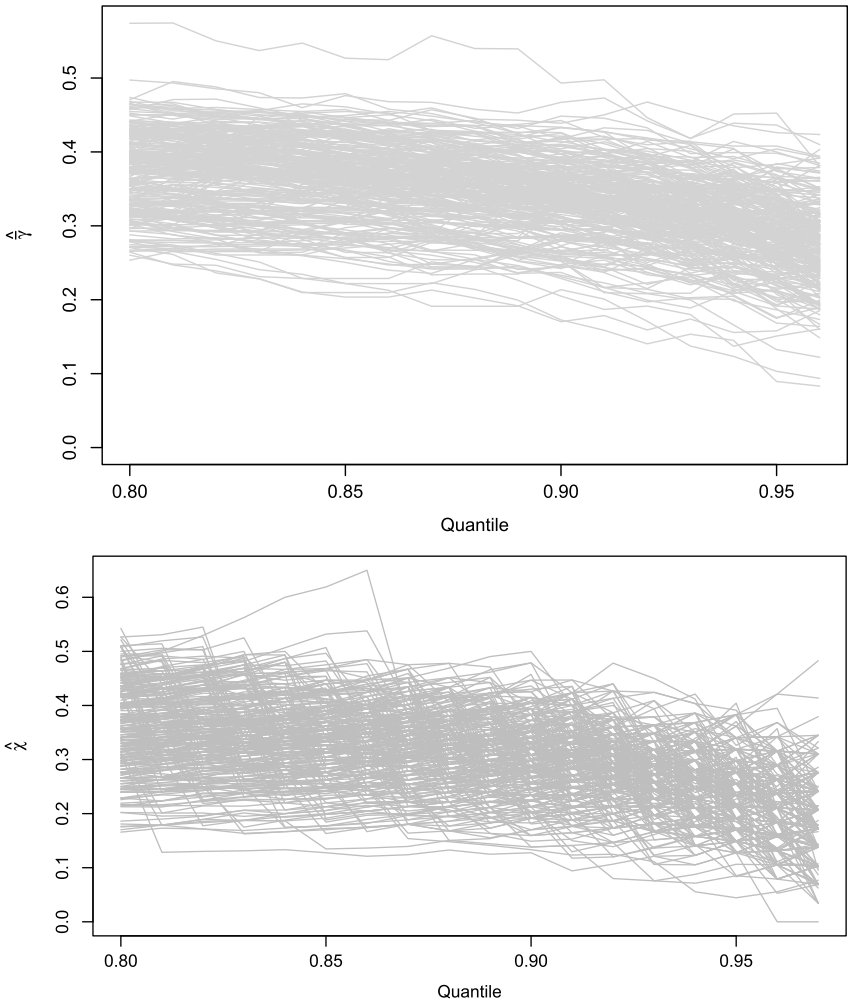


Figure 4: For threshold quantiles ranging from 0.8 to 0.975 across all stations, the top panel shows the resulting sequence of $\hat{\bar{\gamma}}$ estimates, and the bottom panel shows the corresponding sequence of $\hat{\chi}$ estimates.

erally mild compared with many other environmental processes. Moreover, our primary interest is in bivariate extremal dependence, for which any remaining temporal dependence is expected mainly to reduce the effective sample size rather than to substantially bias the dependence estimates. The code and data to reproduce the following analysis is publicly available in a GitHub repository found at https://github.com/brooktrussell/InvestInvestigatingSpatialDependence.

### 3.2 Asymptotic Dependence Estimates

In the top left panel of Figure 5, we plot the value of $\hat{\bar{\gamma}}(0.95)$ for each SNOTEL station location in Wyoming, Idaho, and Montana. Similarly, in the middle left panel of Figure 5, we plot the value of $\hat{\chi}(0.95)$ at the same locations. For reference, we plot the differences $\hat{\chi}(0.95) - \hat{\bar{\gamma}}(0.95)$ for each station in the bottom left panel of Figure 5. The empirical quantile of 0.95 was selected after conducting exploratory analysis, and seemed to provide a good balance of retaining sufficient data yet being far enough out in the upper joint tail. Our choice of threshold was based on

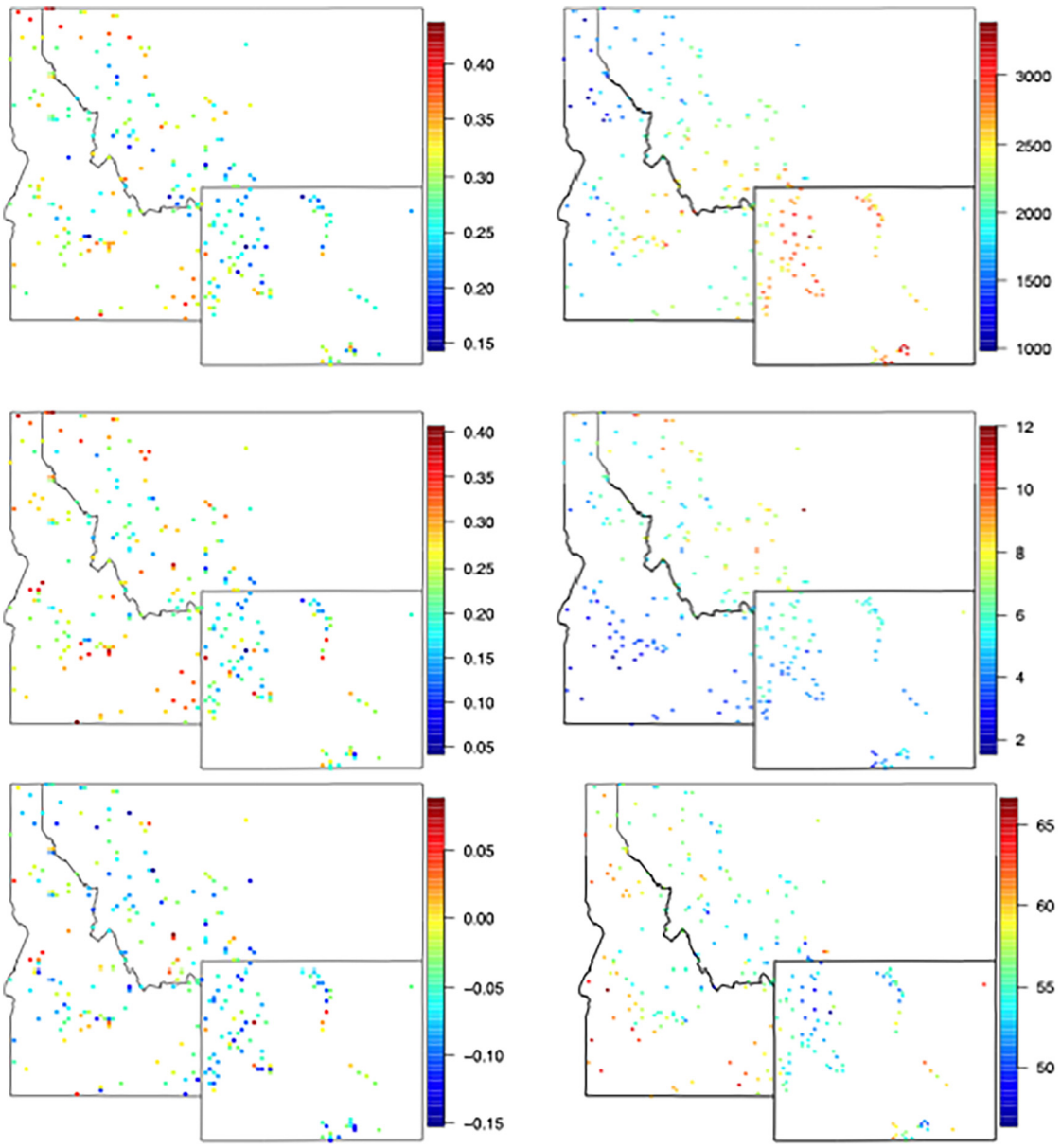


Figure 5: The top left and middle left panels show $\hat{\tilde{\gamma}}(0.95)$ and $\hat{\chi}(0.95)$, respectively, for each SNOTEL station in Wyoming, Idaho, and Montana, summarizing asymptotic dependence between SNOTEL and PERSIANN-CDR precipitation. The bottom left panel plots $\hat{\chi}(0.95) - \hat{\tilde{\gamma}}(0.95)$ for reference. The top right panel shows station elevation (m). The middle and bottom right panels display PRISM 30-year mean summer precipitation (inches) and mean summer air temperature (°F), respectively.

examining the plots in Figure 4, where we plot the asymptotic dependence estimates at all stations for a sequence of quantiles ranging from 0.8 to 0.975. Although there is a great deal of uncertainty involved in threshold selection, we believe that these plots may support the choice of the empirical 0.95 threshold as the plots indicate that the estimators may have stabilized by

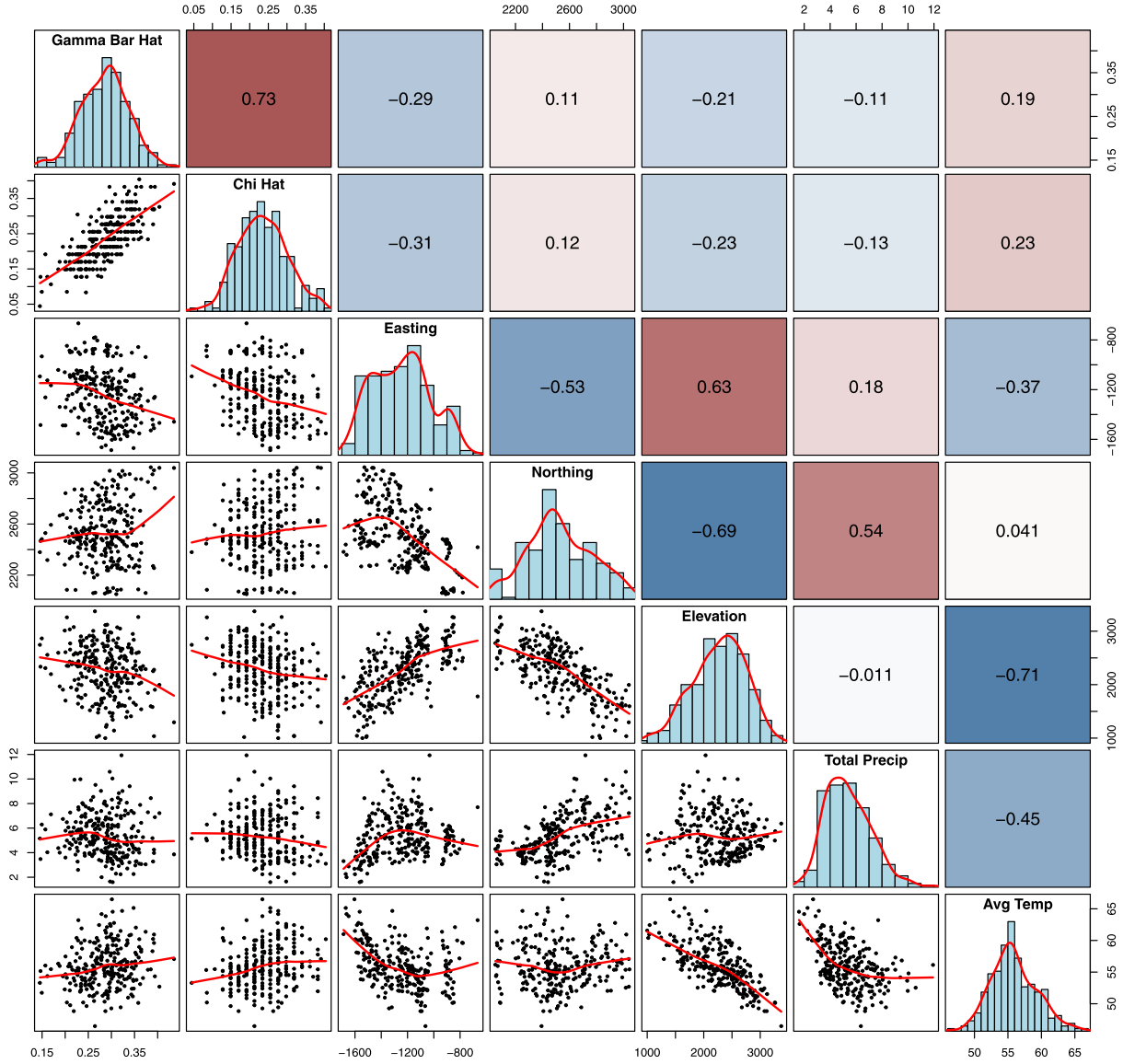


Figure 6: Pairwise scatterplots with the corresponding LOESS smooth (bottom left panels), correlation coefficients (top right panels), and kernel density estimates (diagonal panels) of the asymptotic dependence estimators and geographical covariates considered.

that point, taking the aforementioned "threshold stability" approach. One potential limitation of this approach is that the same threshold is applied to both estimators across all stations, and the stability would not necessarily occur at the same level at all stations. We keep this potential limitation in mind as we proceed with analysis.

Although the estimators $\hat{\bar{\gamma}}$ and $\hat{\chi}$ — defined in Equations (4) and (3) (respectively) — are both asymptotic dependence estimators, they differ in terms of the way that they do this. Importantly, they should not be directly compared or interpreted in terms of their numeric values. That being said, we would expect the resulting estimates to be positively correlated, and observe this in the top scatterplot in Figure 6. To this end, we also note that the estimates have a

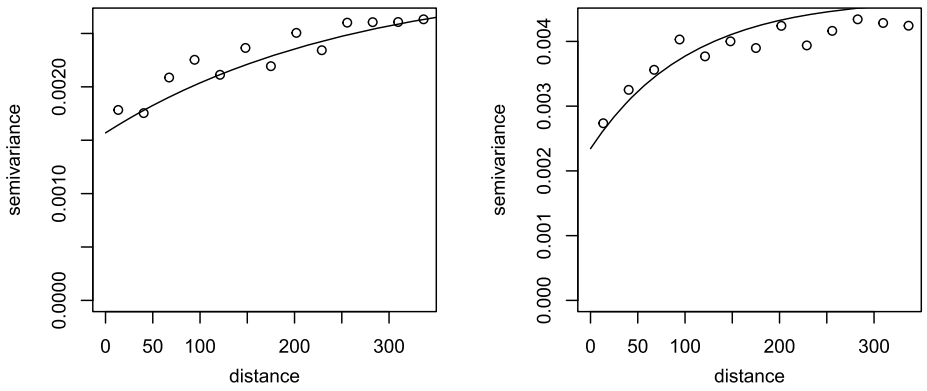


Figure 7: Empirical variograms of the station-level estimates $\hat{\bar{\gamma}}(0.95)$ (left) and $\hat{\chi}(0.95)$ (right) with distances of separation in km. For reference, the curves show the covariance functions implied by the best-fitting least-squares models selected by AIC.

moderately high positive correlation of approximately 0.73. Interesting, careful visual inspection of these two plots suggests that there may be some level of spatial dependence present.

## 3.3 Investigating Spatial Dependence in Asymptotic Dependence Estimates

### 3.3.1 Testing for Spatial Dependence

In order to better understand whether the level of asymptotic dependence throughout this region exhibit spatial dependence, we undertake the following analysis. It is first necessary to determine the distance of separation between each pair of stations. To this end, we calculate distance among station locations, after projecting latitude and longitude (in degrees) to the Contiguous Albers Equal Area Conic projection. This results in axes with units in meters. Hereafter, we refer to these projected station coordinates as 'easting' and 'northing'.

In order to test for the presence of spatial dependence of the $\hat{\bar{\gamma}}(0.95)$ and $\hat{\chi}(0.95)$ values, we using Moran's I test (Moran, 1950). In conducting these tests, we use inverse distance weighting among stations separated by a maximum of 300 km. The tests for positive spatial correlation results in p-values both less than 0.0001. These results suggest the presence of spatial dependence in the asymptotic dependence estimates $\hat{\bar{\gamma}}(0.95)$ and $\hat{\chi}(0.95)$. Additionally, Figure 7 presents empirical variograms for $\hat{\bar{\gamma}}(0.95)$ in the left panel and $\hat{\chi}(0.95)$ in the right panel. Both plots also indicate at least a moderate level of spatial dependence in these estimates, which is consistent with the results of the Moran's I tests.

### 3.3.2 Investigating the Relationships with Geographical Covariates

At each station location, we also consider three spatial/geographical covariates, in addition to easting and northing: elevation above sea level, the 30-year normal summer precipitation total (based on the PRISM model), and the average summer air temperature (also based on the PRISM model). In Figure 6, for the two asymptotic dependence estimators and the three spatial covariates mentioned above, we plot all pairwise scatterplots with the corresponding LOESS smooth in the bottom left panels, all univariate histograms with a corresponding estimated density in the panels on diagonal, and all pairwise correlation coefficients in the top right panels. We plot the station elevation (in m above sea level) for each station in the top right panel of

Figure 5. We plot the 30-year PRISM total summer precipitation and the 30-year PRISM mean summer temperature in the middle right and bottom right panels of Figure 5, respectively.

The output in Figure 6 provides for an exploratory correlation analysis. For these data, concluding that the true correlation parameters differ from zero at the 5% level (individually) requires the corresponding sample correlations exceed 0.12 in absolute value. Considering the tests individually suggests that easting and elevation are negatively correlated with $\hat{\bar{\gamma}}(0.95)$ and $\hat{\chi}(0.95)$, while average temperature is positively correlated with both asymptotic dependence estimates. Northing and total precipitation are not quite significant for $\hat{\bar{\gamma}}(0.95)$, but marginally significant for $\hat{\chi}(0.95)$.

We next consider models that account for spatial dependence in the $\hat{\bar{\gamma}}(0.95)$ and $\hat{\chi}(0.95)$ estimates. Spatial linear models are fit using maximum likelihood via the `geoR` package (Ribeiro and Diggle, 2025) in the statistical computing environment `R` (R Core Team, 2024). Model selection is performed based on the Akaike information criterion (AIC) (Akaike, 1998). Interestingly, the best models for both asymptotic dependence estimates utilize the exponential covariance function and utilize a constant mean function. This indicates that the geographical covariates considered here — while showing a moderate association with the asymptotic dependence estimates — appear to be less important than the correlation analysis above may suggest. For reference, we add the estimated covariance functions for the best fitting spatial models based on AIC to the variograms in Figure 7.

## 4 Discussion

As SPP may be employed in order to understand EPE over remote mountainous regions, it is critical to understand the ways in which levels of asymptotic dependence may vary over such a region. As a case study, we model the level of asymptotic dependence between PERSIANN-CDR and SNOTEL station data, and consider two asymptotic dependence estimators, $\hat{\bar{\gamma}}$ and $\hat{\chi}$. We conduct analysis of the resulting estimates that are based on several years worth of data. Through formal hypothesis testing and visual diagnostics, we find that it is reasonable to believe that there is positive spatial dependence in these asymptotic dependence estimates. This implies that stations that are separated by less distance will tend to have asymptotic dependence levels that are more similar, and vice versa. This is an important idea as it indicates that there could possibly be geographical factors influencing the levels of asymptotic dependence.

Similar to the findings of Russell et al. (2024), our analysis indicates that elevation is negatively correlated with both asymptotic dependence estimators. Additionally, we find that the average summer temperature is positively correlated with both asymptotic dependence estimators. However, it is important to keep in mind that the average summer temperature is strongly related to elevation, but possibly also the location's latitude. Despite the interesting findings that the correlation analysis brought to light, none of these geographical covariates were significant in the model.

In this work, we employ two different asymptotic dependence parameters and their corresponding estimators. We take this approach for illustrative purposes and to be able to compare and contrast two particular ways to assess asymptotic dependence; however, in practice this type of approach is at least somewhat atypical. Normally, an analyst will choose to work with one particular asymptotic dependence parameter a priori. Making inference on $\chi$ is much more common, as it has a clear interpretation, does not require marginal transformation, and has a straightforward estimator. EDMs in general, and $\bar{\gamma}$ in particular a less commonly employed,

but may be more useful in particular cases. For example, when an analyst is seeking a function of covariates that yields an optimal level of asymptotic dependence with a response variable is a research aim — as in the tail dependence regression framework of Russell et al. (2024) — an EDM will likely be useful. In multivariate extremes, the tail pairwise dependence matrix (TPDM) of Cooley and Thibaud (2019) seeks to characterizes pairwise asymptotic dependence and is analogous to the covariance matrix in classical statistics. Cooley and Thibaud (2019) define the TPDM using EDMs that rely on the $L_2$ norm; others, such as Jiang et al. (2020) and Russell and Hogan (2018) have employed this version of the TPDM.

Despite the results of the modeling procedure in this work, we still believe that there are geographical factors that we were unable to obtain that are associated with asymptotic dependence levels over this region. We speculate that more detailed geographical information system (GIS) data — such as information about aspect and the terrain — may account for more variability in the responses. This work provides an initial exploratory perspective, but we hope to pursue a more formal modeling framework that accounts for spatial dependence, nonlinear effects, or covariate interactions in the future. We also hope to be able to employ spatial hierarchical modeling approaches that may be able to better capture the complicated spatial structure, and also consider additional environmental covariates that may account for variability in asymptotic dependence levels. Another potential issue could be related to spatial nonstationarity when applying the semivariogram in practice. We hope to also be able to consider these ideas in future research.

## Supplementary Material

The authors have compiled a GitHub repository that contains files and data related to this manuscript, that is publicly available and located at https://github.com/brooktrussell/InvestInvestigatingSpatialDependence. More specifically, this repository contains the following materials:

1. Analysis.R — the R file that contains the code used in this analysis,
2. Objects.RData — an R workspace file that contains the data used in this analysis,
3. PRISMelevation — an R workspace file that contains the PRISM data used in this analysis, and
4. SimStudyCode.R — the R file that contains the code used in the simulation study.

## A Appendix

For the spatial models considered in Section 3.3.2, we rely on AIC to perform model selection. Additionally, we note that exploratory analysis indicates that the exponential covariance function is preferred for both asymptotic dependence metrics (again based on AIC). Although we do not present results here, we also note that we do not find quadratic terms involving the geographic coviates to be useful in these models. To this end, Table 1 gives the AIC values for the spatial models considered in the analysis of $\hat{\bar{\gamma}}$. Analogously, Table 2 gives the AIC values for the spatial models considered in the analysis of $\hat{\chi}$. As mentioned in Section 4, we believe that a more comprehensive spatial hierarchical modeling approach may be able to capture the complicated underlying spatial signals in these data. We plan to pursue this approach in future work.

Table 1: We define the models and report the resulting AIC values for all spatial models considered in the analysis of $\hat{\tilde{\gamma}}$.

| Model | Const | Elev | Precip | Temp | AIC |
|---|---|---|---|---|---|
| M1 | X | | | | −866.43 |
| M2 | X | X | | | −864.46 |
| M3 | X | | X | | −864.77 |
| M4 | X | | | X | −864.47 |
| M5 | X | X | X | | −863.04 |
| M6 | X | X | | X | −862.47 |
| M7 | X | | X | X | −863.15 |
| M8 | X | X | X | X | −861.15 |

Table 2: We define the models and report the resulting AIC values for all spatial models considered in the analysis of $\hat{\chi}$.

| Model | Const | Elev | Precip | Temp | AIC |
|---|---|---|---|---|---|
| M1 | X | | | | −722.83 |
| M2 | X | X | | | −721.08 |
| M3 | X | | X | | −721.91 |
| M4 | X | | | X | −721.13 |
| M5 | X | X | X | | −719.94 |
| M6 | X | X | | X | −719.15 |
| M7 | X | | X | X | −719.91 |
| M8 | X | X | X | X | −717.95 |

## References

AghaKouchak A, Behrangi A, Sorooshian S, Hsu K, Amitai E (2011). Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research: Atmospheres*, 116(D2): D02115.

Akaike H (1998). Information theory and an extension of the maximum likelihood principle. In: Emanuel Parzen, Kunio Tanabe, Genshiro Kitagawa (Eds.), *Selected Papers of Hirotugu Akaike*, 199–213. Springer.

Ashouri H, Hsu KL, Sorooshian S, Braithwaite DK, Knapp KR, Cecil LD, et al. (2015). PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society*, 96(1): 69–83. https://doi.org/10.1175/BAMS-D-13-00068.1

Bharti V, Singh C (2015). Evaluation of error in TRMM 3B42V7 precipitation estimates over the Himalayan region. *Journal of Geophysical Research: Atmospheres*, 120(24): 12458–12473. https://doi.org/10.1002/2015JD023779

Coles S (2001). *An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics*. Springer-Verlag, London.

Coles S, Heffernan J, Tawn J (1999). Dependence measures for extreme value analyses. *Extremes*, 2(4): 339–365. https://doi.org/10.1023/A:1009963131610

Cooley D, Cisewski J, Erhardt RJ, Jeon S, Mannshardt E, Omolo BO, et al. (2012). A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects. *REVSTAT Statistical Journal*, 10(1): 135–165.

Cooley D, Thibaud E (2019). Decompositions of dependence for high-dimensional extremes. *Biometrika*, 106(3): 587–604. https://doi.org/10.1093/biomet/asz028

Daly C, Taylor G, Gibson W (1997). The PRISM approach to mapping precipitation and temperature. In: *Proceedings, 10th AMS Conference on Applied Climatology*, 20–23.

Derin Y, Yilmaz KK (2014). Evaluation of multiple satellite-based precipitation products over complex topography. *Journal of Hydrometeorology*, 15(4): 1498–1516. https://doi.org/10.1175/JHM-D-13-0191.1

Dey DK, Jiang Y, Yan J (2016). Multivariate extreme value analysis. In: *Extreme Value Modeling and Risk Analysis: Methods and Applications* (DK Dey, J Yan, eds.), 23–39. Chapman and Hall/CRC. Chapter 2.

Engelke S, Hitz AS (2020). Graphical models for extremes. *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, 82(4): 871–932. https://doi.org/10.1111/rssb.12355

Gong Y, Zhong P, Opitz T, Huser R (2024). Partial tail-correlation coefficient applied to extremal-network learning. *Technometrics*, 66(3): 331–346. https://doi.org/10.1080/00401706.2024.2304334

Hirpa FA, Gebremichael M, Hopson T (2010). Evaluation of high-resolution satellite precipitation products over very complex terrain in Ethiopia. *Journal of Applied Meteorology and Climatology*, 49(5): 1044–1051. https://doi.org/10.1175/2009JAMC2298.1

Huang WK, Cooley DS, Ebert-Uphoff I, Chen C, Chatterjee S (2019). New exploratory tools for extremal dependence: $\chi$ networks and annual extremal networks. *Journal of Agricultural, Biological, and Environmental Statistics*, 24(3): 484–501. https://doi.org/10.1007/s13253-019-00356-4

Huser R, Wadsworth JL (2022). Advances in statistical modeling of spatial extremes. *Wiley Interdisciplinary Reviews: Computational Statistics*, 14(1): e1537. https://doi.org/10.1002/wics.1537

Jiang Y, Cooley D, Wehner MF (2020). Principal component analysis for extremes and application to US precipitation. *Journal of Climate*, 33(15): 6441–6451. https://doi.org/10.1175/JCLI-D-19-0413.1

Larsson M, Resnick SI (2012). Extremal dependence measure and extremogram: The regularly varying case. *Extremes*, 15(2): 231–256. https://doi.org/10.1007/s10687-011-0135-9

Moran PA (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2): 17–23. https://doi.org/10.1093/biomet/37.1-2.17

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Resnick S (2004). The extremal dependence measure and asymptotic independence. *Stochastic Models*, 20(2): 205–227. https://doi.org/10.1081/STM-120034129

Resnick S (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, 1st edition. *Springer Series in Operations Research and Financial Engineering*. Springer, New York.

Ribeiro Jr PJ, Diggle P (2025). *geoR: Analysis of Geostatistical Data*. R package version 1.9-5.

Russell BT, Cooley DS, Porter WC, Reich BJ, Heald CL (2016). Data mining to investigate the meteorological drivers for extreme ground level ozone events. *Annals of Applied Statistics*, 10(3): 1673–1698. https://doi.org/10.1214/16-AOAS954

Russell BT, Ding Y, Huang WK, Dyer JL (2024). Characterizing asymptotic dependence be-

tween a satellite precipitation product and station data in the northern US Rocky Mountains via the tail dependence regression framework with a Gibbs posterior inference approach. *Environmetrics*, 35(8): e2890. https://doi.org/10.1002/env.2890

Russell BT, Hogan P (2018). Analyzing dependence matrices to investigate relationships between National Football League Combine event performances. *Journal of Quantitative Analysis in Sports*, 14(4): 201–212. https://doi.org/10.1515/jqas-2017-0086

Wadsworth JL, Campbell R (2024). Statistical inference for multivariate extremes via a geometric approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(5): 1243–1265. https://doi.org/10.1093/jrsssb/qkae030