NIST Special Publication 500-301

Maps of Non-hurricane Non-tornadic Wind Speeds With Specified Mean Recurrence Intervals for the Contiguous United States Using a Two-dimensional Poisson Process Extreme Value Model and Local Regression

Adam L. Pintar
Emil Simiu
Franklin T. Lombardo
Marc Levitan

This publication is available free of charge from:
http://dx.doi.org/10.6028/NIST.SP.500-301

National Institute of Standards and Technology
U.S. Department of Commerce

NIST Special Publication 500-301

Maps of Non-hurricane Non-tornadic Wind Speeds With Specified Mean Recurrence Intervals for the Contiguous United States Using a Two-dimensional Poisson Process Extreme Value Model and Local Regression

Adam L. Pintar
Information Technology Laboratory
Statistical Engineering Division

Franklin T. Lombardo
Engineering Laboratory
Materials and Structural Systems Division
and University of Illinois

Emil Simiu
Engineering Laboratory
Materials and Structural Systems Division

Marc Levitan
Engineering Laboratory
Materials and Structural Systems Division

This publication is free of charge from:
http://dx.doi.org/10.6028/NIST.SP.500-301

November 2015

U.S. Department of Commerce
Penny Pritzker, Secretary

National Institute of Standards and Technology
Willie May, Under Secretary of Commerce for Standards and Technology and Director

National Institute of Standards and Technology Special Publication 500-301
Natl. Inst. Stand. Technol. Spec. Publ. 500-301, 102 pages (November 2015)
CODEN: NSPUE2

This publication is available free of charge from:
http://dx.doi.org/10.6028/NIST.SP.500-301

Abstract

This report describes a procedure for creating maps of non-hurricane, non-tornadic wind speeds for a set of recurrence intervals of interest between 10 and 100 000 years over the contiguous United States. The procedure is carried out in two stages. In the first stage a peaks over threshold extreme value model is fitted to observed wind speeds from meteorological stations spread across the contiguous United States. The fitted model is then used to estimate wind speeds for all mean recurrence intervals of interest at each station. Potential differences between thunderstorm and non-thunderstorm wind speeds are accounted for by the model. The estimated wind speeds are then spatially smoothed using local regression to produce maps and associated uncertainties. Software was developed to facilitate plotting the maps and extracting values from them.

Key Words: Extreme Winds; Poisson Process; Local Regression; Mean Recurrence Interval; Map; Contiguous United States

Contents

List of Figures

List of Tables

1 Introduction

The goal of this report is to describe a procedure for taking time histories (irregular time series) of non-hurricane, non-tornadic wind speeds from meteorological reporting stations across the contiguous United States and create maps of wind speeds for various mean recurrence intervals between 10 and 100 000 years. A photograph of a such a station at Milford, Utah is found in Fig. 1. The tall red and white striped structure is the anemometer. A wind speed with a mean recurrence of interval of N years is henceforth referred to as an N-year return value or simply a return value when specifying the number of years is unnecessary. The procedure is performed in two stages. In the first stage, a peaks over threshold approach to modeling extreme values is used to estimate return values at each station separately. The potential impact of thunderstorm versus non-thunderstorm winds on the estimated return values is accounted for by the model. In the second stage, spatial smoothing with local regression is used to create the maps as well as calculate standard uncertainties and upper confidence bounds.

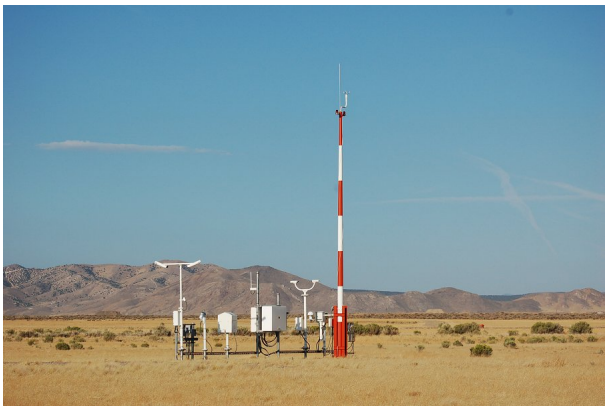


Figure 1: Automated Surface Observing Systems (ASOS) station in Milford, Utah.

There are two main approaches to modeling extreme values and estimating return values: the classical approach and the peaks over threshold approach. The classical approach models yearly maxima with a generalized extreme value distribution, and is based on the theory originally developed in [1] and [2]. There are two major approaches within the peaks over threshold methodology. The first models the exceedances of a chosen high threshold as independent Generalized Pareto observations. It was justified originally by the theory in [3]. The times of the exceedances are then separately modeled as a Poisson process. This approach is applied as a statistical procedure in [4], for example. The second approach jointly models the time and magnitude of the threshold exceedance using a two-dimensional Poisson process. The theoretical basis for this approach was originally developed in [5], and it was employed as a statistical procedure in [6], for example. The two-dimensional Poisson

process approach is more flexible than the Generalized Pareto approach (see [7] page 91), but they are related, and in App. A a specific connection between them is derived. In this work, the peaks over threshold approach is chosen over the classical approach to increase the number of observations available for estimation. Further, the Poisson process approach is selected over the Generalized Pareto approach for its flexibility, as indicated earlier.

In all figures, except for the maps in Apps. B and C, the units for wind speeds are provided in both miles per hour and kilometers per hour. For wind speeds found in the text, the same is true. The maps in Apps. B and C provide units for wind speeds only in miles per hour. Providing units for wind speeds in miles per hour is essential because the goal of this work is to support the next revision of *Minimum Design Loads for Buildings and Other Structures* published by the American Society of Civil Engineers, and the units in that publication are miles per hour.

Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

This report is divided into six main sections and three appendices. Section 2 describes the available data, and Secs. 3 and 4 describe the procedure for creating the maps. Section 5 introduces an R package [8] that was developed for plotting the maps as well as extracting values from them. The main points of the report are revisited in Sec. 6. The first Appendix draws a connection between the Poisson process and Generalized Pareto approaches to extreme value analysis, showing that under certain conditions they are equivalent. Appendix B considers a variation on the intensity function of the Poisson process used in this work (the intensity function used in this work is described in Sec. 4). Finally, in App. C the collection of maps is presented.

2 Data

Information about the original wind speed data obtained from the National Climatic Data Center and its compilation to create the raw data set used in this report is available at http://www.itl.nist.gov/div898/winds/NIST_TN/nist_tn.htm. In addition to standardization of the data for averaging time and anemometer height as described on that web site, a terrain roughness correction is also applied. Wind speed values in the raw data set represent measured or equivalent 3-second peak gust wind speeds at a height of 10 m (33 ft) above the ground in flat, open terrain (defined on page 251 of [9] as Exposure C). The processed data can be downloaded from http://www.itl.nist.gov/div898/winds/NIST_TN/final_qc_data.htm.

To account for the important effects of terrain roughness, a first order approximation was employed to estimate a station average roughness length weighted by the directionality of the station's peak gusts and then convert peak gusts to equivalent Exposure C (roughness length $z_0 = 0.03$ m) values. Roughness length values at most stations were calculated using a procedure similar to that presented in [10]. The z_0 values were estimated over eight 45 degree sectors centered on 0, 45, ..., 315 degrees. Comparison of these z_0 values with estimates obtained from visual review of aerial photographs for sample stations and from

examples provided in [10] showed a reasonable match. A 'global roughness length' was then calculated for each station. This global roughness length was estimated by weighting the z_0 in each sector by the frequency of peak gusts in the database from that sector. Very high global roughness values ($z_0 > 0.3\,\mathrm{m}$) were observed at several stations. Investigation of aerial photographs and topographic maps of these locations typically showed buildings, trees, and/or topographic features in very close proximity to the anemometer. These stations were removed from the database. For stations where it was not possible to determine a roughness length in the manner described above due to limitations of the original data, each station was assigned a z_0 based on a regional average value. A procedure was then followed to account for the effects of terrain roughness at each station. This was accomplished by increasing or decreasing the peak gusts by half of the change implied by converting from the exposure at each station (as estimated by the global roughness value) to $z_0 = 0.03\,\mathrm{m}$, as use of the full increase at stations with greater surface roughness resulted in what appeared to be anomalously high wind speeds in comparison to surrounding stations in more open terrain. This procedure to account for the effects of terrain roughness was vetted through review by several subject matter experts.

The processed original data, referred to in the remainder of this report as the raw data since they are the data that are directly modeled, are histories or irregular time series of wind speeds above some low threshold at 1180 reporting stations in the contiguous Unites States. The data generating process can be thought of in the following way. A reporting station records the time and speed of winds that exceed a specified low threshold. A third characteristic, wind type, is determined *a posteriori*. Two wind types are considered, *thunderstorm* and *synoptic*. The synoptic winds are often referred to as non-thunderstorm winds in the remainder of this report with the understanding that non-thunderstorm winds are neither from hurricanes, tornadoes, nor thunderstorms. Winds from hurricanes and tornadoes were excluded from the data set during its creation. Figure 2 illustrates the raw data from station 724650 in western Kansas. Here, the terminology "low threshold" is preferred so that the threshold used in data collection is differentiated from the typically much higher threshold used in the estimation of return values, which is discussed in Sec. 3. From Fig. 2 it is clear that the low threshold for data collection changes over time. This is immaterial since the final data used for inference will be thresholded at a typically much higher level. The seasonal effect is clearly visible in Fig. 2. Thunderstorms tend to occur in the summer, not winter.

The locations of the 1180 stations are depicted by the points (both green and black) in Fig. 3. The 575 stations that contribute to the final map are depicted by the green points. The criteria used to winnow the number of stations from 1180 to 575 is described in Sec. 3.1.6.

The data at Station 724650 are "nice" in the sense that they do not exhibit features that complicate their handling such as periods of time with no observations or apparent changes in the rate of observed low threshold crossings. However, these features are present, for example at Station 722320 in southern Louisiana (see Fig. 4). The approaches to handling such challenging features are discussed in Sec. 3.

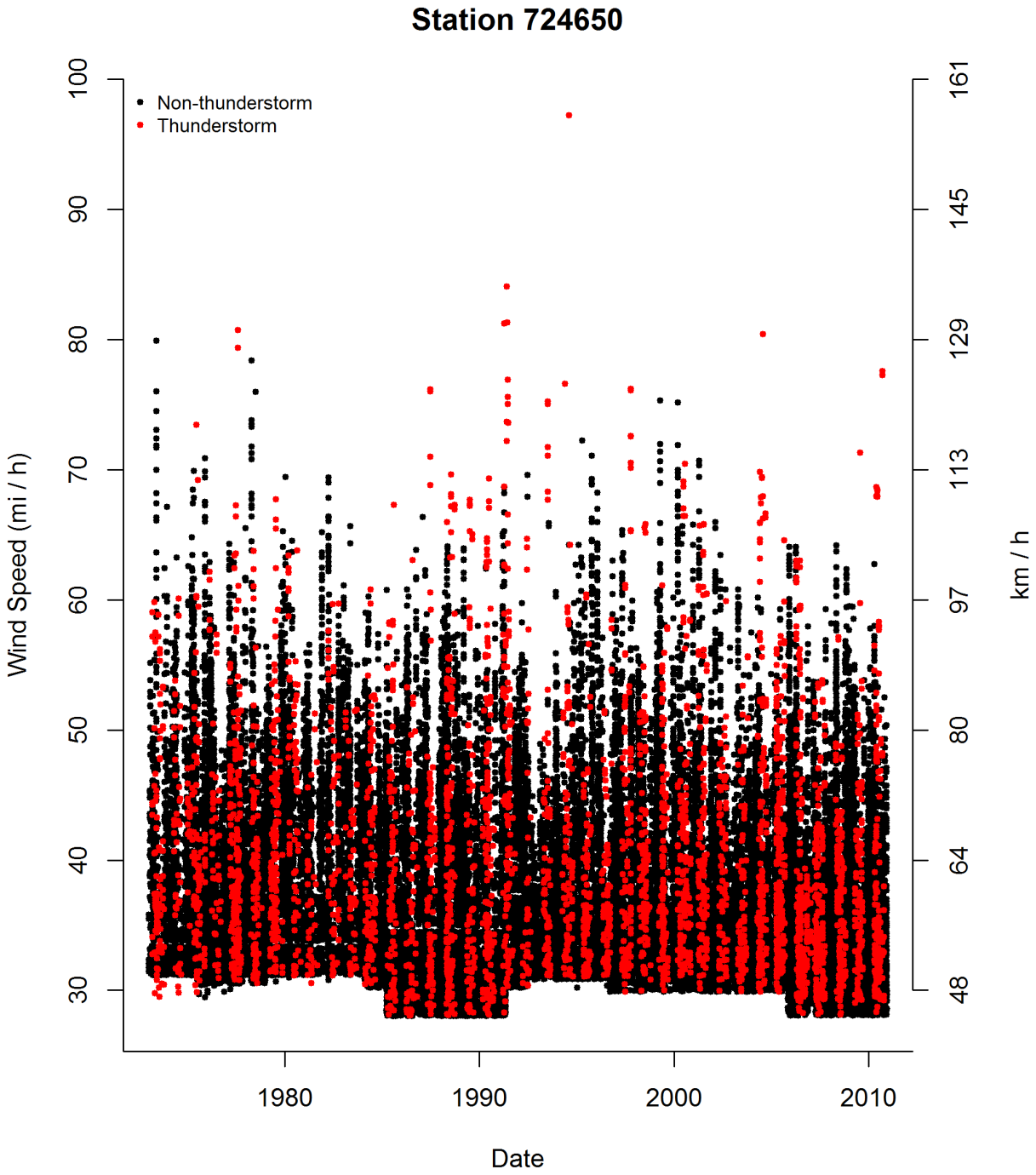


Figure 2: Raw data for station 724650 in western Kansas separated by thunderstorm and non-thunderstorm wind types.

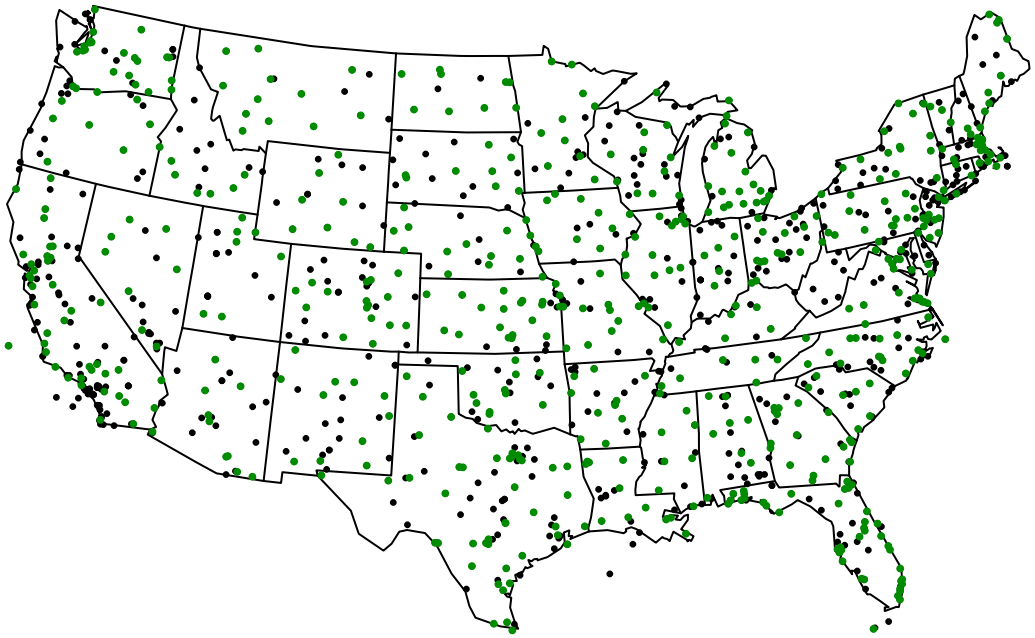


Figure 3: Locations of all 1180 meterological stations in the raw data set (both green and black points). The stations contributing to the final maps are depicted in green.

3 Estimation Approach

The development of return value maps is completed in the following two stages:

1. Fit a two-dimensional Poisson process to each station using maximum likelihood and use that fitted process to estimate return values of interest.

2. Using local regression, interpolate the estimated return values to a regular grid covering the contiguous United States, which are the maps.

The first stage takes the approach of [6] to modeling extreme values. The combination of fitting a model to individual stations in stage 1 and then spatially smoothing in stage 2, resembles the approach taken in Sec. 1.5.4 of [11] to modeling rainfall extremes. The individual stages are described further in Sec. 3.1 and 3.2 that follow, respectively.

3.1 Stage 1

In Stage 1 of this estimation procedure, a Poisson process model is used. Poisson processes are stochastic processes (in one or more dimensions) that are defined (essentially) by their intensity function. They are usually categorized into the classes homogeneous and non-homogeneous (though the homogeneous class can be thought of as a special case of the non-homogeneous class). Homogeneous and non-homogeneous refers to properties of the intensity function. For homogeneous Poisson processes the intensity function is uniform or

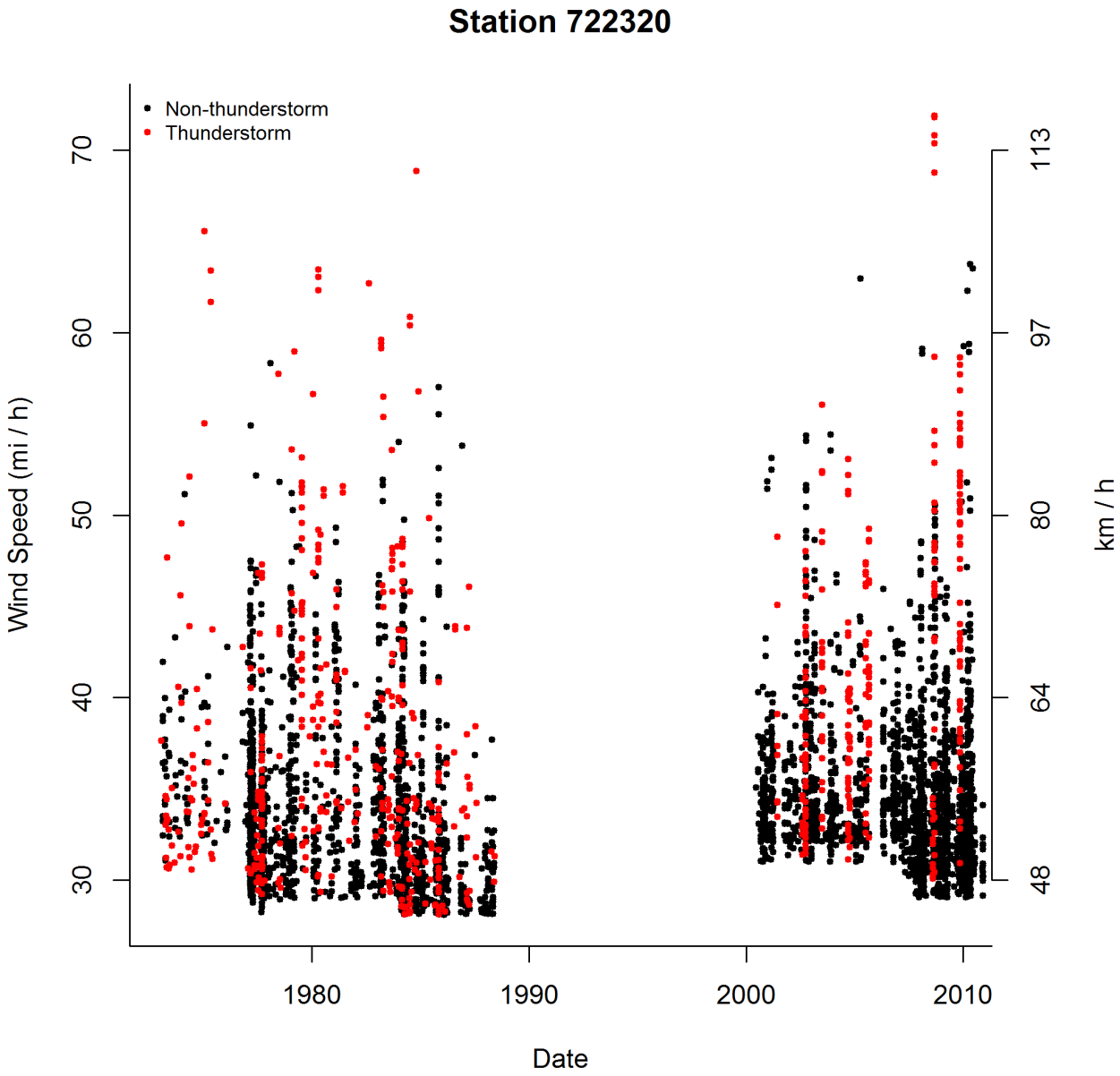


Figure 4: Raw data for station 722320 in southern Louisiana by thunderstorm and non-thunderstorm wind types.

flat over the domain on which the Poisson process is defined. Non-homogeneous then of course implies that the intensity function is not uniform over the domain. One dimensional homogeneous Poisson processes are described in Sec. 2.4 of [12], and Chapter 4 of [13] as well as [14] provide general discussions on point processes, of which the Poisson process is a specific example. Later in this section a rigorous definition of the Poisson process used in this work is provided.

In the first stage, a two-dimensional non-homogeneous Poisson process on the domain $\mathcal{D} = \mathcal{D}_s \bigcup \mathcal{D}_{ns}$ is fitted to the observed data, denoted $\{(t_i,\ y_i)\}_{i=1}^{I}$ at each station, where t_i is the time of a threshold crossing and y_i is the magnitude of the wind speed. The data used in fitting, $\{(t_i,\ y_i)\}_{i=1}^{I}$ is not the entire dataset as depicted in Fig. 2 or 4. Instead, only a subset is used. The subset is obtained by thresholding the observations at a level that is at least as large as the one used in data collection. After re-thresholding the observations, they are further declustered. Thresholding and declustering are discussed further in Sec. 3.1.4 and 3.1.3, respectively. The intensity function of the Poisson process is

$$\lambda(t, y) = \begin{cases} \lambda_s(y) \text{ for } t \text{ in a thunderstorm period} \\ \lambda_{ns}(y) \text{ for } t \text{ in a non-thunderstorm period.} \end{cases} \tag{1}$$

The sub-scripts s and ns stand for *storm* (thunderstorm) and *non-storm* (non-thunderstorm), respectively. The domain, $\mathcal{D}$ is the union of two disjoint sets, $\mathcal{D}_s$ and $\mathcal{D}_{ns}$, which are themselves the union of many disjoint sets. More specifically, suppose that the sequence of intervals $\{(\alpha_k, \beta_k)\}_{k=1}^{K}$ marks the beginning and ending time of all thunderstorms that occur in a finite period of time at some station. That is, α_k marks the beginning of thunderstorm k, β_k marks the ending of thunderstorm k and K thunderstorms occur at the station under consideration. Then, $\mathcal{D}_s = \bigcup_{k=1}^{K}(\alpha_k, \beta_k) \times (b_s, \infty)$ and $\mathcal{D}_{ns} = \left[\bigcup_{k=1}^{K}(\alpha_k, \beta_k)\right]^c \times (b_{ns}, \infty)$, where A^c is the complement of the set A, and the complement is taken with respect to the interval of time (T_m, T_M). So $T_M - T_m$ is the amount of time that the station under consideration has been in service. The process for choosing the *thresholds*, b_s and b_{ns} is described shortly, but note that they will likely be different from the low thresholds used in data collection. A generic intensity function is provided in Eq. (1) because in Sec. 4 three different intensity functions are considered. The Poisson process used in this work is defined by two conditions:

1. For any set $D \subset \mathcal{D}$, the number of observations in D follows a Poisson distribution with mean $\int_D \lambda(t, y)\ dtdy$.

2. If $D_1, D_2 \subset \mathcal{D}$ are disjoint, then the number of observations in D_1 and D_2 are independent random variables.

Using a Poisson process to model extreme value data is justified by limit theorems. See [5], [6] and [11] for details.

Figure 5 illustrates the domain $\mathcal{D}$ on which the Poisson process is defined, and is a generalization of Fig. 1.3 in [11]. In Fig. 5 the number of observations to fall within $D \subset \mathcal{D}$

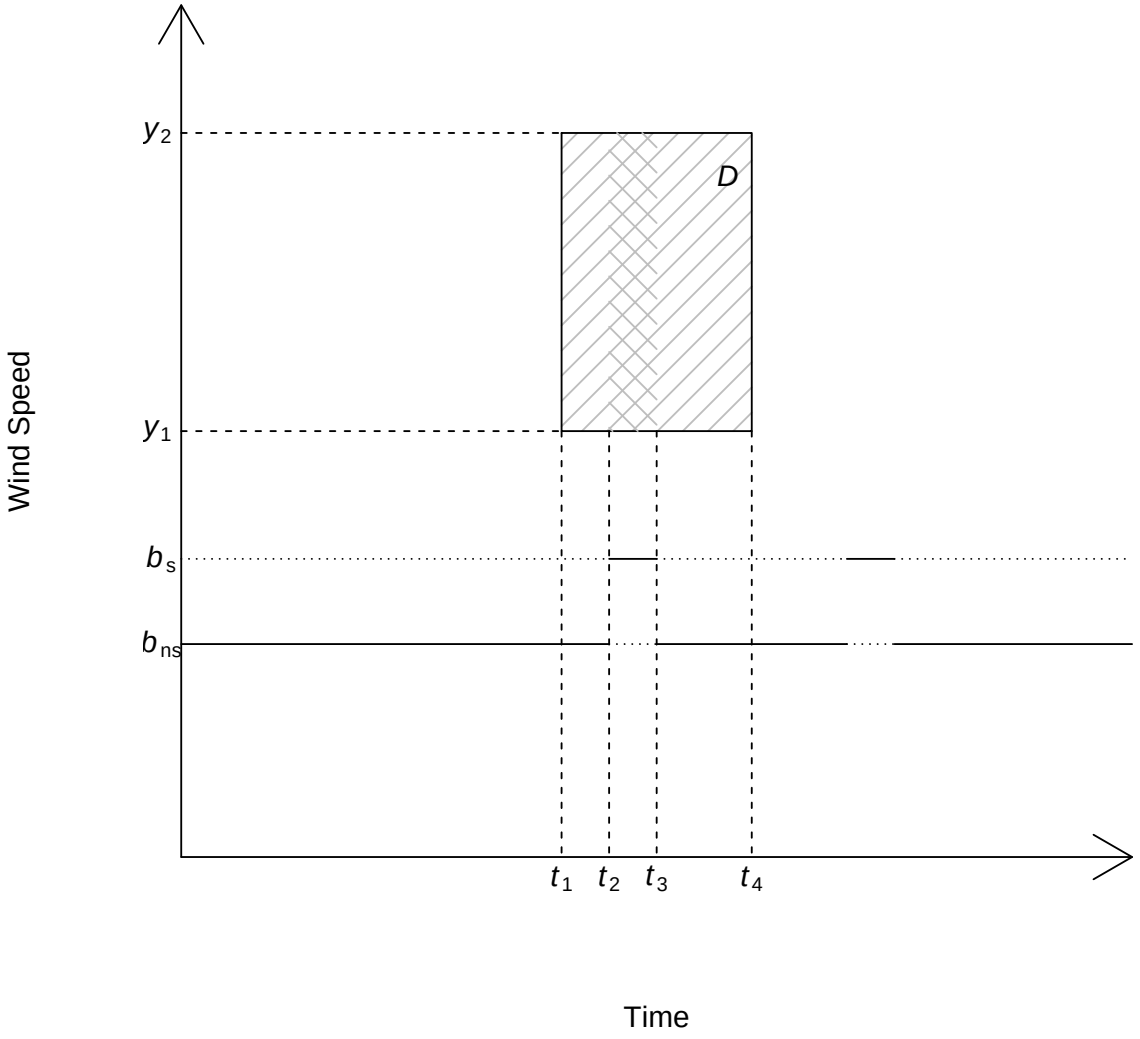


Figure 5: Illustration of $\mathcal{D}$. The number of points to fall in D follows a Poisson distribution with mean $\int_D \lambda(t,y)dtdy$. The time interval (t_2, t_3) is taken to be a thunderstorm time period, and (t_1, t_2) and (t_3, t_4) are taken to be non-thunderstorm time periods.

follows a Poisson distribution with mean

$$\begin{aligned}\Lambda =&(t_2 - t_1) \int_{y_1}^{y_2} \lambda_{ns}(y)\ dy \\ &+(t_3 - t_2) \int_{y_1}^{y_2} \lambda_{s}(y)\ dy \\ &+(t_4 - t_3) \int_{y_1}^{y_2} \lambda_{ns}(y)\ dy.\end{aligned}$$

So the mean of the Poisson distribution is the amount of volume trapped by the intensity function over $D \subset \mathcal{D}$. The volume trapped by the intensity function is the sum of three separate integrals because the time interval (t_2, t_3) is taken to be a thunderstorm time period and (t_1, t_2) and (t_3, t_4) are taken to be non-thunderstorm time periods.

3.1.1 Likelihood

The Poisson process parameters are estimated by maximum likelihood (ML) [see Ref. 15, page 114 for example]. To do this, a likelihood, or the joint probability density of the set of observations, is needed. In Eq. (1.18) of [11] that joint density is given to be

$$L(\boldsymbol{\eta}) = \left(\prod_{i=1}^{I} \lambda(y_i, t_i) \right) \cdot \exp\left\{ - \int_{\mathcal{D}} \lambda(y, t)\ dydt \right\}, \tag{2}$$

which is actually only proportional to the real joint density, but is still perfectly suitable to use as the likelihood. The generic $\boldsymbol{\eta}$ represents the vector of parameters to be estimated, which are presented in Sec. 4. To derive Eq. (2), note that conditional on the number of observations, I, the observations are independent and identically distributed with probability density

$$f(t, y|\boldsymbol{\eta}) = \frac{\lambda(t, y)}{\int_{\mathcal{D}} \lambda(t, y)\ dtdy}$$

and that I is a Poisson random variable with mean $\int_{\mathcal{D}} \lambda(t, y)\ dtdy$

The ML estimate of $\boldsymbol{\eta}$ is the value of $\boldsymbol{\eta}$, $\widehat{\boldsymbol{\eta}}$ that maximizes $L(\boldsymbol{\eta})$. An analytical expression for this optimum is not available for any of the intensity functions considered here; however, $L(\boldsymbol{\eta})$ may be optimized numerically. Specifically, the `optim` function in the `stats` package of R [8] is used, which implements by default the algorithm described in [16].

3.1.2 Return Values

Recall that the goal of stage 1 is to estimate a set of return values of interest at each station. The set of interest is 10, 25, 50, 100, 300, 700, 1200, 1700, 2000, 2500, 3000, 5000, 10 000, 50 000 and 100 000 years. The N-year return value is the y_N such that the probability of exceeding y_N in 1 year is $\frac{1}{N}$. Since the number of observations that fall into the region $[0, 1] \times (y_N,\ \infty)$ follows a Poisson distribution with mean $\int_{y_N}^{\infty} \int_0^1 \lambda(t, y)\ dtdy$ (by the definition

of a Poisson process), y_N is the solution to

$$1 - \exp\left\{-\int_{y_N}^{\infty}\int_{0}^{1} \lambda(t,y)\ dtdy\right\} = \frac{1}{N} \tag{3}$$

The time interval $(0,\ 1)$ is used generically since for any *unobserved* one year period of time, $(t_a,\ t_b)$

$$\int_{y_N}^{\infty}\int_{t_a}^{t_b} \lambda(y,y)\ dtdy = A_s \int_{y_N}^{\infty} \lambda_s(y)\ dy + A_{ns}\int_{y_N}^{\infty} \lambda_{ns}(y)\ dy,$$

where A_s and A_{ns} represent the *typical* amount of thunderstorm and non-thunderstorm time over a year, respectively. The equation

$$\int_{y_N}^{\infty}\int_{0}^{1} \lambda(y,y)\ dtdy = \frac{1}{N}. \tag{4}$$

is solved instead of Eq. (3) because for large N, $1-\exp\{-1/N\}$ is approximately $1/N$. This is also the approach taken in [17].

An analytic solution to Eq. (4) is not available, but it may be solved numerically. Specifically, the function `uniroot` in the `stats` package of R is used. It is necessary to choose values for A_s and A_{ns} to solve Eq. (4), and each station is assigned its own value of A_s and A_{ns}. To calculate these quantities, the average number of thunderstorms per year from the observation period is first tabulated. This is done by counting the number of gaps between thunderstorm observations that are larger than some interval of time (taken to be six hours as in [18]) and dividing by the number of years that station has been in service to give the average number of thunderstorms per year. Then, A_s is the average number of thunderstorms per year multiplied by the average length of a thunderstorm (taken to be one hour as in [19]). It happens that the assumed average thunderstorm length cancels out of the calculation of the N-year return value as long as the same average value is used in both ML estimation and solving Eq. (4). This was checked numerically since it cannot be shown analytically. While the average thunderstorm length cancels out of the final calculation, examining Fig. 3 of [18] implies that it may not be a bad choice. Of course, Fig. 3 of [18] only considers a single station. Since the average length of a thunderstorm could vary spatially (which was not examined here or in [18]), the fact that it cancels out of final calculation bolsters confidence. The assumptions of this procedure are that an observation is recorded for each thunderstorm that occurs at a station and that thunderstorm observations separated by more than six hours arise from different thunderstorms.

3.1.3 Clusters of Observations

Since the recorded observations are governed by underlying environmental processes, they tend to form clusters. That is, a single environmental phenomenon tends to give rise to multiple observations. This makes the Poisson process assumption less justifiable, so the observations are declustered prior to model fitting. Declustering is accomplished by examining the time gaps between adjacent observations. A cluster is defined as a sequence of observa-

tions for which all adjacent observations are separated by some fixed amount of time or less (taken to be six hours for thunderstorm observations and four days for non-thunderstorm observations [18]). Thus if two observations are separated by longer than this fixed amount of time, they belong to separate clusters. The cluster maximums are used in model fitting. This is the same algorithm that is used in [6].

3.1.4 Choice of Threshold

Choosing a pair (thunderstorm and non-thunderstorm) of thresholds for each station is a critical part of the analysis. Recall that the Poisson process models being used are justified by limit theorems, and the limits are taken as time increases to infinity and the thresholds increase to infinity. So if the thresholds are chosen too small, the model may not match the data very well. However, the observing station has only been in service a finite amount of time, so if the thresholds are chosen too large, few data are available for estimation. Threshold choice is often accomplished by using a high percentile of the data. For example, in [17], the 95th percentile is used.

A different approach is taken here. A search for the threshold pair that produces the optimal (in a sense) fit of the data to model is undertaken. A regular grid of threshold pairs on $\mathbb{Z}^2$ (only integer thresholds are considered) is constructed first. To do so, a *minimum* and *maximum* number of thunderstorm and non-thunderstorm observations per year is chosen. If the station has been in service for S years, only threshold pairs that lead to at least $S \cdot minimum$ thunderstorm and non-thunderstorm observations, but not more than $S \cdot maximum$ thunderstorm and non-thunderstorm observations are considered. The operating point $minimum = 4$ and $maximum = 15$ is used. Using $minimum = 4$ is based on the desire to use substantially more observations than classical extreme value methods (a benefit of peaks over threshold methods). Using $maximum = 15$ balances computational burden with exhaustively exploring the space of potential threshold pairs. Note that 4 observations per year times the number of years the station is in service is not the minimum possible sample size for fitting the Poisson process model. There is no guarantee that many observations will be available. However, stations with very few observations are excluded from the final maps. This is discussed further in Sec. 3.1.6.

The basis of the search for the optimal threshold pair is the W-statistic found on page 31 of [11], which leverages the probability integral transform (see for example page 54 of [20]). Given that an observation occurs at time t (it is known to be a thunderstorm or non-thunderstorm observation), the probability that the observation is less than or equal to y is

$$F(y) = P(Y \leq y) = 1 - \frac{\int_b^y \lambda(t,x)\ dx}{\int_b^\infty \lambda(t,x)\ dx} \tag{5}$$

The probability integral transform states that if the random variable Y is distributed according to the distribution function $F(y)$ from Eq. (5), then $U = F(Y)$ is distributed as a Uniform(0, 1) random variable, which then implies that $W = -\log(1-U)$ is an exponential random variable with mean one. So at a station, once the parameters have been estimated for a potential threshold pair, for each observation that exceeds the threshold, a W-statistic

is calculated. The W-statistics are sorted and plotted against percentiles of the exponential distribution with mean one Exp(1). This is referred to as a W-plot from now on. If the points closely follow the 45° line, there is good agreement between the data and the model.

An example of such a plot for station 724110 near Roanoke, Virginia with 58 km/h for the thunderstorm threshold (36 miles per hour (mi/h)) and 79 km/h for the non-thunderstorm threshold (49 mi/h) with the Gumbel type tail length parameter is given in Fig. 6 (the types of tail length parameters are discussed in Sec. 4). In Fig. 6 there is excellent agreement between the data and the model. For some stations the agreement is not as good. One example is Station 722287 between Atlanta and Birmingham, Georgia near Interstate 20 with W-plot in Fig. 7. Figure 7 implies an outlier is causing the majority of the degraded fit, and the boxplots of the thresholded and declustered observations separated by thunderstorm and non-thunderstorm wind types in Fig. 8 shows an apparent outlier in the non-thunderstorm observations. The non-thunderstorm observation that is close to 160 km/h (100 mi/h) is much larger than the rest of the observations that range from about 48 km/h (30 mi/h) to 115 km/h (70 mi/h). Both the outlier and Station 722287 contribute to the final maps since there is no physical reason for excluding them (e.g. a malfunction or that the outlying observation was due to a tornado or hurricane). The criteria used in discarding stations from the final maps are discussed in Sec. 3.1.6.

For each station, there is a collection of W-plots, one for each potential threshold pair. Human examination and comparison of these plots is the ideal way to choose the best threshold pair. However, that is impractical, so an automated procedure for favoring one W-plot over another is implemented. Specifically the vertical distance from each point on the W-plot to the 45° line is calculated, and the summary for a plot is the maximum distance. For each station, the threshold pair that leads to the minimum summary is chosen. This approach is inspired by the minimax estimator (see for example page 261 of [21]). It was also the approach taken in [19], and it is very similar to the approach endorsed in Sec. 2 of [22] for choosing a threshold when fitting data to a generalized Pareto distribution. While the specifics of the approaches here and in Sec. 2 of [22] differ, they both select the threshold that provides the best fit (albeit measured in slightly different ways) of the model to the data. For the Gumbel type tail length, the selected thunderstorm thresholds range from 42 km/h (26 mi/h) to 82 km/h (51 mi/h), and the non-thunderstorm thresholds range from 45 km/h (28 mi/h) to 105 km/h (65 mi/h).

3.1.5 Removing Time Gaps

Recall that the underlying assumption at each station is that the thresholded, declustered observations follow a Poisson process with intensity function given in Eq. (1). Stations that have long time spans with no observations and/or an apparent change in the rate of recording observations call into question this assumption. The raw non-thunderstorm data for Station 722320 in Fig. 4 illustrate both of these artifacts, especially the long time span with no observations. They are artifacts because it is unlikely that the underlying wind climate at these stations would change in their relatively short service life. The long time gaps in observations seem almost certainly due to long periods of time when the station is inoperable. The cause of the change in the observation rate is less clear. Perhaps there are many short periods of time when the station is not recording observations. To account for

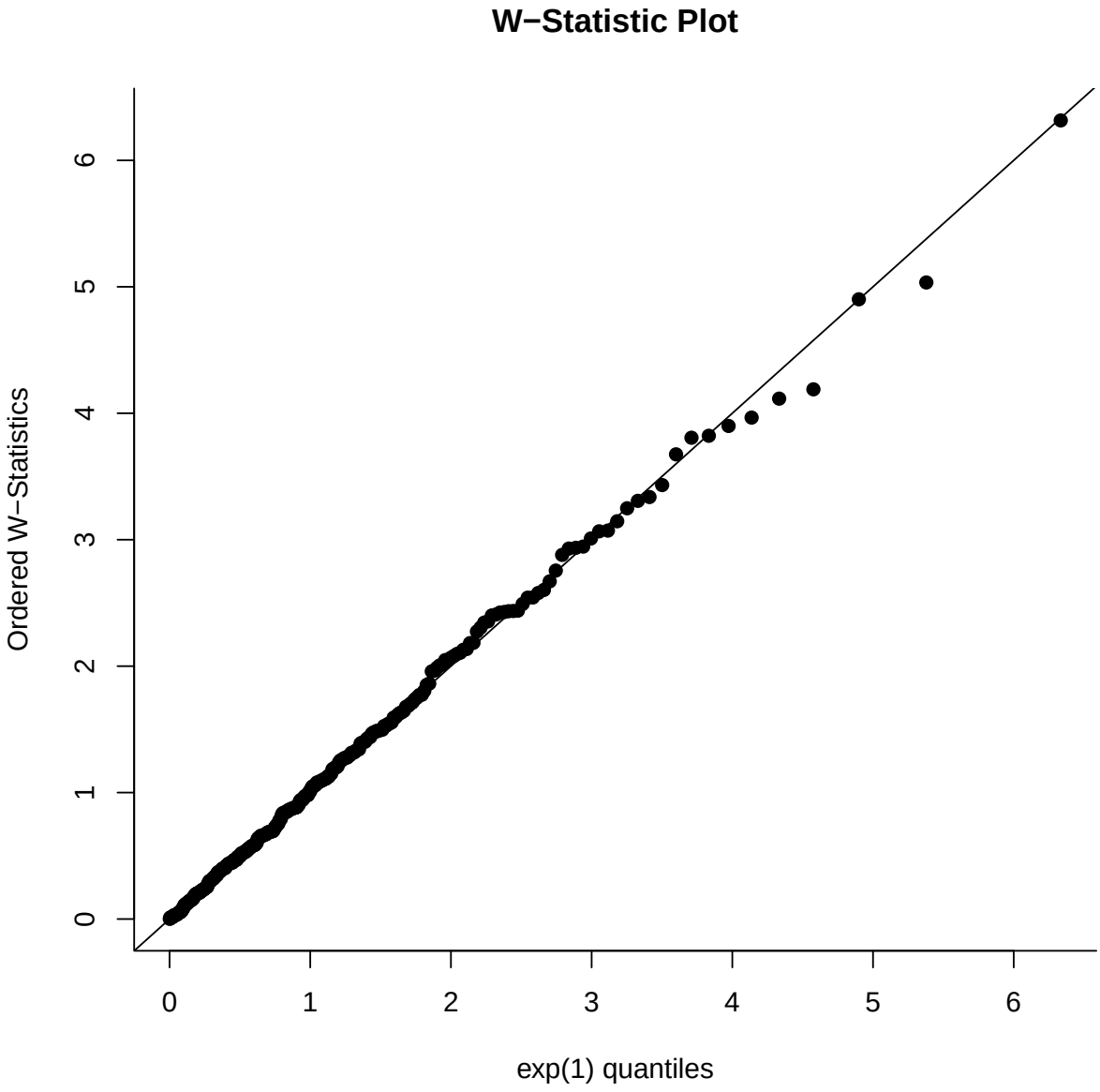


Figure 6: Plot of ordered W-statistics versus Exp(1) percentiles for Station 724110 with a Gumbel like tail.

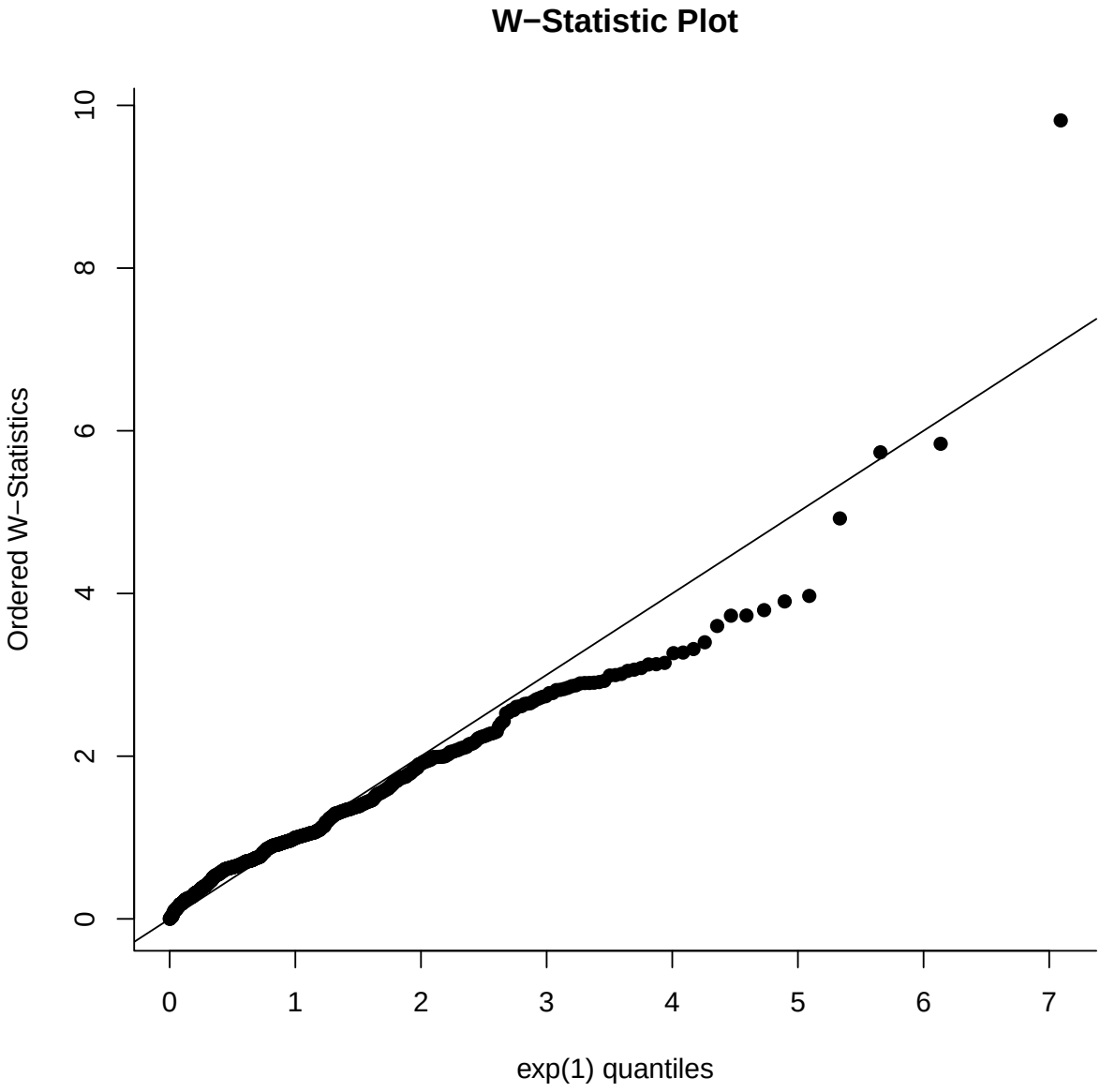


Figure 7: Plot of ordered W-statistics versus Exp(1) percentiles for Station 722287 with a Gumbel like tail.

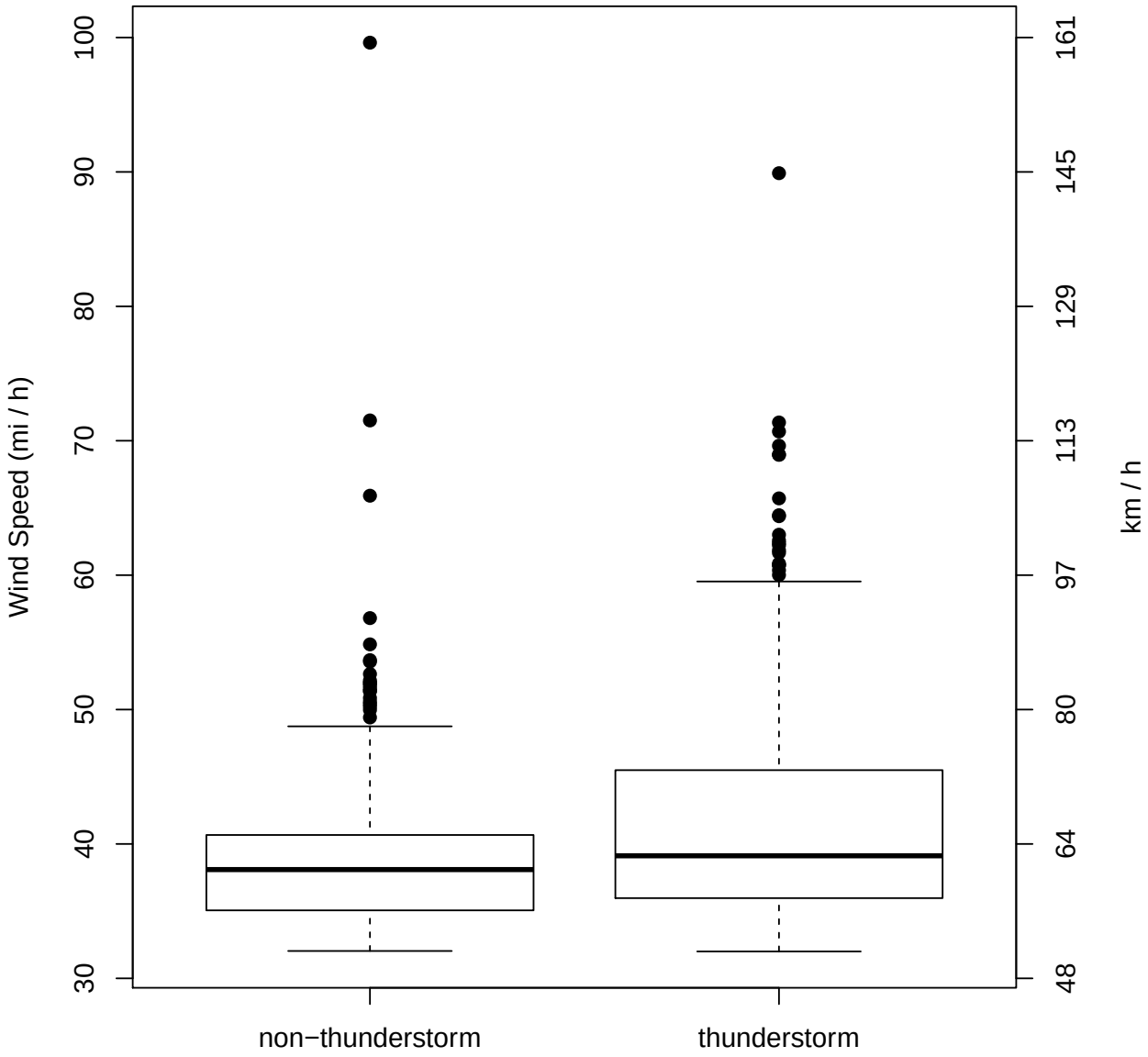


Figure 8: Boxplots of thresholded, declustered observations using the optimal threshold pair for station 722287

both artifacts, time gaps of six months or more are removed from the domain of the Poisson process, $\mathcal{D}$. This will tend to raise return value estimates at individual stations, i.e. be conservative, and comparisons have shown that the changes to the maps are minor.

Figure 9 shows the observations for Station 722320 after thresholding, declustering, and removing all time gaps of six months or more. The left vertical axis is wind speed in mi/h, and the right vertical axis is wind speed in km/h. The horizontal axis is the number of days that have passed since January 21, 1973, the day that the first observation was recorded, after removing all observation gaps from the original observed series of six months or more. For comparison, the length of the time series depicted in Fig. 4 is 13 827 days.

3.1.6 Station Requirements

Not all of the 1180 stations are used in the creation of the final maps. The 575 stations that are used (depicted in Fig. 3 by the green points) all meet the following criteria:

1. There are at least ten thunderstorm and ten non-thunderstorm observations available for fitting after thresholding and declustering. That forces there to be at least five observations per parameter for estimation. A rule-of-thumb for linear models given on page 330 of [23] is six observations or more per parameter.

2. After removing all time gaps of six months or more, the station has been in service for at least 15 years. This matches the recommendation in [24] and the Canadian counterpart of these maps. It is also supported by our own examination of the complete set of stations where we found that stations in service less than 15 years tended to produce lower return value estimates.

3. There is not an abnormally low or high average number of thunderstorms per year relative to neighboring stations. This was accomplished by first counting the number of thunderstorms at each station and computing the average number per year. Then the natural logarithm of those averages were spatially smoothed using the `locfit` [25] package of R [8]. The neighborhood size is chosen by generalized cross validation [26], and the degree of the local polynomial is taken to be one. The `locfit` package is discussed in more detail in the next section, 3.2. The residuals from the smooth are plotted against their fitted values, and abnormal stations are judged by the overall pattern of the points in the plot as well as prediction bounds. If there were no abnormal observations, the points in the plot would form a horizontal band around zero, and all but one or two of the points would fall within the prediction bounds (with an appropriately chosen probability level).

4. There are no abnormal observations after standardization to a 10 m anemometer height and Exposure C as judged by a subject matter expert. For some stations with very tall or short anemometers or with very high global roughness values ($z_0 > 0.3$ m as described in Sec 2), the standardization led to abnormalities.

5. The final estimated return values do not appear anomalous compared to neighboring stations as judged by a subject matter expert.

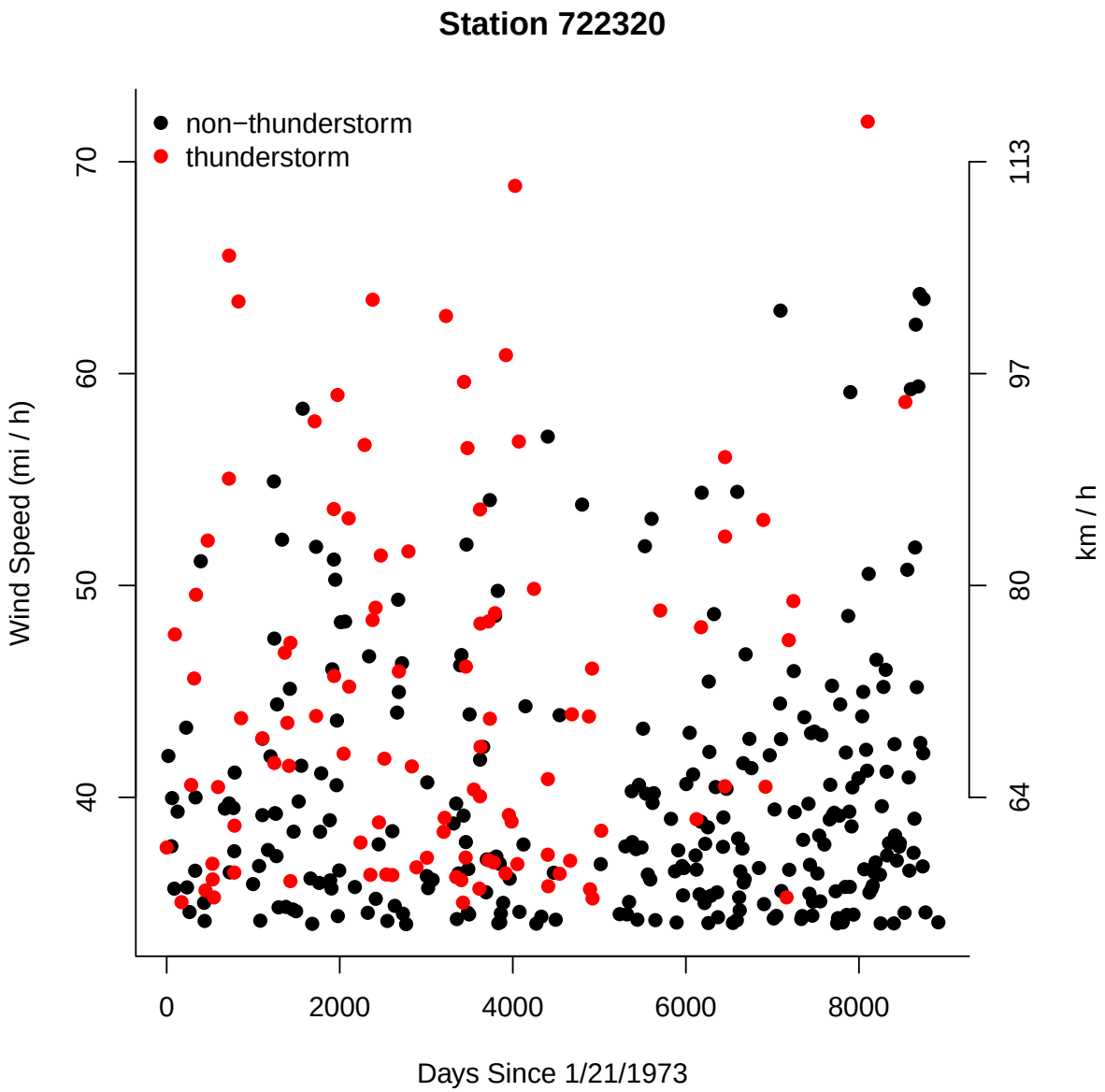


Figure 9: Declustered, thresholded observations for station 722320 ignoring wind type with observation gaps of six months or more removed.

The criteria responsible for not using the overwhelming majority of the 605 stations were 1-4. For criterion 5, stations with potentially anomalous data were identified by review of the magnitudes of the final estimated return values for each station in comparison with neighboring stations. For each of the identified outliers, a subject matter expert reviewed the underlying data and conditions at the station (topography, terrain roughness, anemometer location relative to nearby buildings, etc). Using this procedure, several stations were found to have previously unidentified conditions that significantly impacted the measured wind speeds from certain directions (mainly topographic effects or proximity to buildings). Where physical explanations for the outlying return values were identified, those stations were removed from the dataset.

A list of the meteorological stations used in the creation of the final maps is provided in App. D, along with a map displaying the locations of these 575 stations. This map (Fig. 61) shows that they are spread throughout the contiguous United States. Although stations are more sparsely distributed in some parts of the country, particularly in the Southwestern states and around West Virginia (multiple stations in this area were discarded to topographic effects on the data), no large geographic regions appear to be neglected.

3.2 Stage 2

After all of the return values of interest are calculated at each station, spatial smoothing is used to interpolate them to a regular grid covering the contiguous United States. The grid dimension was set at 200×200, which was deemed appropriate after comparison to a 100×100 grid. The spatial smoothing technique employed is local polynomial regression (LPR) or local regression for short. Some early work in the statistical literature includes [27], where the approach was called locally weighted scatter smoothing (LOWESS), and was limited to a single independent variable (so not useful for spatial smoothing). In [28], the methodology of [27] was extended to multiple independent variables, and the new approach was dubbed LOESS (see [28] for the reason for shortening LOWESS to LOESS). In this work, the implementation of local regression found in the R package LOCFIT [25] is used.

The basic model for local regression, from page 15 of [29] is

$$\ln\{y_N(\theta_j, \phi_j)\} = \mu(\theta_j, \phi_j) + \epsilon(\theta_j, \phi_j), \tag{6}$$

where ln is the natural logarithm function, $y_N(\theta_j, \phi_j)$ is the estimated N-year return value at station j, with coordinates longitude θ_j and latitude ϕ_j, μ is a smooth function of θ and ϕ and ϵ represents a random error that is zero on average and has constant variance (i.e. not a function of θ and ϕ), σ^2. The symbol θ is used for the longitude in place of the more traditional λ since λ refers to the intensity function of the Poisson process here. The natural logarithm of the return values are spatially smoothed instead of the return values themselves so that the assumption of constant residual variance is more tenable. Thus, the uncertainty in the final estimates of the return values will increase with the estimates themselves.

The raw longitudes and latitudes of the stations are not used for smoothing. Their Lambert conformal conic projection (see [30] page 104) is used instead. The parameter and orientation vectors used in the transformation are $(33, 45)'$ and $(90, 0, -98.538)'$, respectively. For the orientation vector, 90 and 0 are the latitude and longitude, respectively, of the north

pole, and -98.583 is about the longitude of the east, west center of the contiguous Unites States. Thus, Eq. (6) is more accurately written as

$$\ln\{y_N(h_x(\theta_j,\phi_j),h_y(\theta_j,\phi_j))\} = \mu(h_x(\theta_j,\phi_j),h_y(\theta_f,\phi_j)) + \epsilon(h_x(\theta_j,\phi_j),h_y(\theta_j,\phi_j)).$$

For notational simplicity, however Eq. (6) is not replaced. The R package `mapproj` [31] is used for the transformations.

In local regression, two tuning parameters must be chosen, which are the degree of the local polynomial and the neighborhood size. While a globally optimal pair of these parameters is possible, a trade-off between them exists. Specifically, a higher degree local polynomial will likely require a larger neighborhood and vice versa. So, the polynomial degree is chosen first, and the neighborhood size is optimized to that choice. A locally linear polynomial is used. Thus, the estimate of $\mu(\theta,\phi)$, say $\widehat{\mu}(\theta,\phi)$, at an arbitrary coordinate (θ,ϕ) is the predicted value from fitting a plane by weighted least squares to the natural logarithm of the return values at neighboring stations. The weight for a station is the tricube function (see page 16 of [29]) of the normalized (to the radius of the neighborhood) euclidean distance from the station to (θ,ϕ) (in the transformed space not longitude, latitude space). This is hypothetically illustrated in Fig. 10. The triangle is the coordinate at which an estimate of $\mu(\theta,\phi)$ is sought. The large circle is the neighborhood. The black plus signs are stations outside of the neighborhood. The black filled dots are stations within the neighborhood, and their size represents the weight placed upon them.

The neighborhood size is adaptive, and it is specified as the proportion of the population to include in the local fit. For example if the neighborhood size were 0.2, the radius of the neighborhood around a given coordinate would be expanded until it included 20% of the population of stations. Specifying the size of the neighborhood in this way ensures that in sparse areas (few stations) the radius of the neighborhood expands so as to sufficiently smooth, and in dense areas (many stations) the radius contracts so as to not over smooth. To choose the neighborhood size, the generalized cross validation score is used as a guide. It is calculated for each value in a sequence of neighborhood sizes using the R function `gcvplot` from the `locfit` package. The generalized cross validation score was introduced for splines in [26] and is defined for local regression on page 31 of [29]. The generalized cross validation score is an approximation to the leave-one-out cross validated error, which would be calculated for a fixed value of the neighborhood size by

1. Deleting a station from the dataset
2. Using the rest of the stations to predict the return value of the deleted station
3. Squaring the difference between the prediction and the observed return value for the deleted station
4. Repeating 1 – 3 for each station in the dataset
5. Summing all of the squared differences.

The neighborhood size with the smallest generalized cross validation score is considered first, but it led to maps that were judged to be under smooth by subject matter experts. The

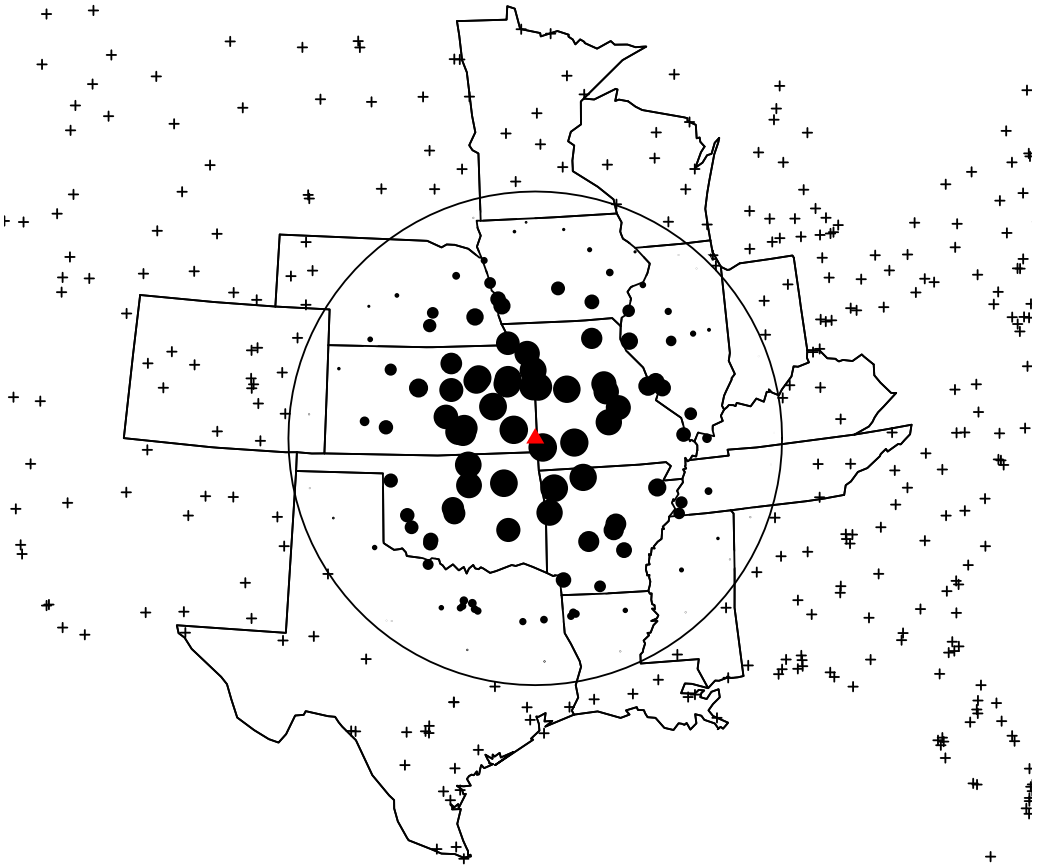


Figure 10: Hypothetical neighborhood around a coordinate in southeast Kansas. The triangle is the coordinate at which an estimate of $\mu(\theta, \phi)$ is sought. The circle is the neighborhood. The plus signs are stations outside of the neighborhood. The black dots are stations within the neighborhood, and their size represents the weight placed upon them.

neighborhood size was incrementally increased from the size suggested by cross validation until the resulting maps were judged to be sufficiently smooth by subject matter experts. The choice was validated by examining the residuals, which is considered further in Sec. 3.2.2. The sequence of potential neighborhood sizes was evenly spaced from $\frac{10}{\text{total stations}}$ to one. So the minimum number of stations to include in a neighborhood was ten. The optimal neighborhood size could depend on the model (intensity function) as well as the mean recurrence interval, but about 20% was generally a good choice and was used almost exclusively. Unless mentioned otherwise, assume all maps in this report were generated using a neighborhood size of about 20%. The neighborhood size depicted in Fig. 10 is consistent with that assumption.

After $\mu(\theta, \phi)$ is estimated at each (θ, ϕ) pair in the regular grid covering the contiguous United States, the estimates are exponentiated to return to the original units (mi/h or km/h). The value $\exp\{\widehat{\mu}(\theta, \phi)\}$ is interpreted as the estimate of the true return value at coordinate (θ, ϕ).

3.2.1 Uncertainty

The error term in Eq. (6), $\epsilon(\theta_j, \phi_j)$ is intended to account for errors due to sampling (white noise). However, if a localized feature has been unknowingly over smoothed, that over smoothing would also be absorbed by the error. The generalized cross validation score tries to minimize the occurrence of this by optimizing the trade off between bias and variance (over smoothing and under smoothing), but in the interest of being conservative, and since the final neighborhood size was chosen to produce a smoother map than the ideal according to cross validation, a measure of uncertainty that attempts to account for both bias due to potential over smoothing and uncertainty due to noise is recommended. The measure of uncertainty is the estimated standard deviation of the prediction of a new hypothetical return value at a desired coordinate, (θ, ϕ). The new hypothetical return value is written as

$$\exp\{\mu(\theta, \phi) + \epsilon(\theta, \phi)\} \tag{7}$$

and it is estimated by replacing $\mu(\theta, \phi)$ and $\epsilon(\theta, \phi)$ by their most likely values $\widehat{\mu}(\theta, \phi)$ and zero, respectively. An estimate of the standard deviation of the prediction of Eq. (7) using the propagation of error formula [32] is

$$\exp\{\widehat{\mu}(\theta, \phi)\} \cdot \sqrt{\widehat{\text{var}}\{\widehat{\mu}(\theta, \phi)\} + \widehat{\sigma}^2}, \tag{8}$$

where $\widehat{\text{var}}\{\widehat{\mu}(\theta, \phi)\}$ is an estimate of the variance of $\widehat{\mu}(\theta, \phi)$ given in Eq. (2.13) of [29] and $\widehat{\sigma}^2$ is an estimate of the error variance, namely the sample variance of the residuals. The words error and residual are used intentionally. The error is a random variable whose true value cannot be known. The residual is an estimate of the error, namely $\ln\{y_N(\theta_j, \phi_j)\} - \widehat{\mu}(\theta_j, \phi_j)$.

3.2.2 Upper Bound

A $(1 - \alpha)100\%$ upper bound at coordinate (θ, ϕ) is given by

$$\exp\{\widehat{\mu}(\theta, \phi)\} + z_{1-\alpha} \cdot \widehat{\text{se}}(\theta, \phi), \tag{9}$$

where $z_{1-\alpha}$ is the $(1-\alpha)$th quantile of the standard normal distribution and $\widehat{\text{se}}(\theta, \phi)$ is Eq. (8) evaluated at (θ, ϕ). This upper bound relies on two conditions:

1. That the average error term in Eq. (6) is zero.
2. That the variance of the error term in Eq. (6) is constant.
3. That the error term in Eq. (6) is normally distributed.

The first two conditions are also necessary for the estimation of $\mu(\theta, \phi)$ by the LOESS procedure. The condition of normality is only necessary for Eq. (9) to maintain its nominal coverage. The zero mean and constant variance of the error terms was examined by residuals versus predicted values (map values) plots. Such a plot for the 700-year return value and Gumbel type tail length is found in Fig. 11. The solid red line marks zero, and the dashed red lines mark a 95% confidence interval for the mean of the residuals, which encompasses zero. Since the points form a horizontal band, constant variance is reasonable too. For the Gumbel type tail length, for all return values of interest, the assumptions of zero mean and constant variance errors are tenable.

The normality of the error terms was examined through normal quantile-quantile plots of the residuals, which are discussed on page 106 of [23]. Normality was formally tested by imposing simultaneous confidence bounds on the quantile-quantile plots. These bounds were calculated using a slight modification of the `envelope` function of the boot package [33] for R [8], which is based on Eq. 4.17 of [34]. For the Gumbel type tail length parameter and the 700 year mean recurrence interval, the quantile-quantile plot with bounds is found in Fig. 12. There do not appear to be any substantial divergences from normality, and the tails (extremes) appear to be quite well modeled by the normal probability distribution. The normal assumption for the error terms is reasonable for all return values of interest for the Gumbel type tail length parameter. Figure 13, which depicts a histogram of the residuals with a best fit normal distribution overlaid for the 700 year mean recurrence interval and Gumbel type tail length further confirms this.

4 Intensity Functions

The intensity functions considered here are all special cases of the intensity function found in Eq. (1.19) of [11],

$$\frac{1}{\psi_t}\left(1+\zeta_t\frac{y-\omega_t}{\psi_t}\right)_+^{-1/\zeta_t-1},$$

where the notation $(\cdot)_+$ is defined as

$$(\cdot)_+ = \begin{cases} \cdot \text{ for } \cdot > 0 \\ 0 \text{ otherwise} \end{cases}.$$

The ζ_t parameters control the tail length of the intensity function at a given t, but for wind speed data they can be difficult to estimate precisely. This is discussed further in App. B

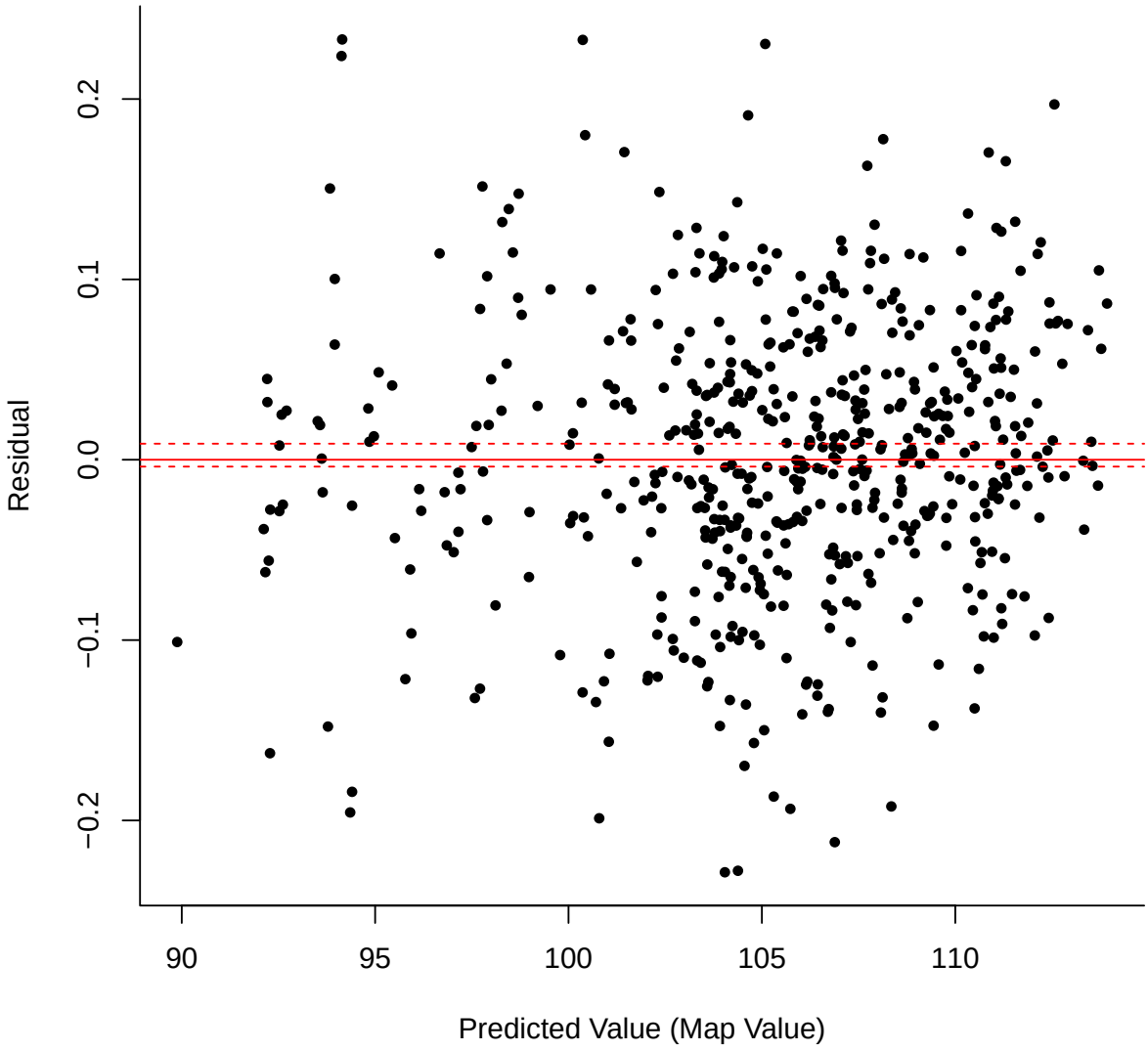


Figure 11: A plot of the residuals versus the predicted values (map values) for $N = 700$ and the Gumble type tail length

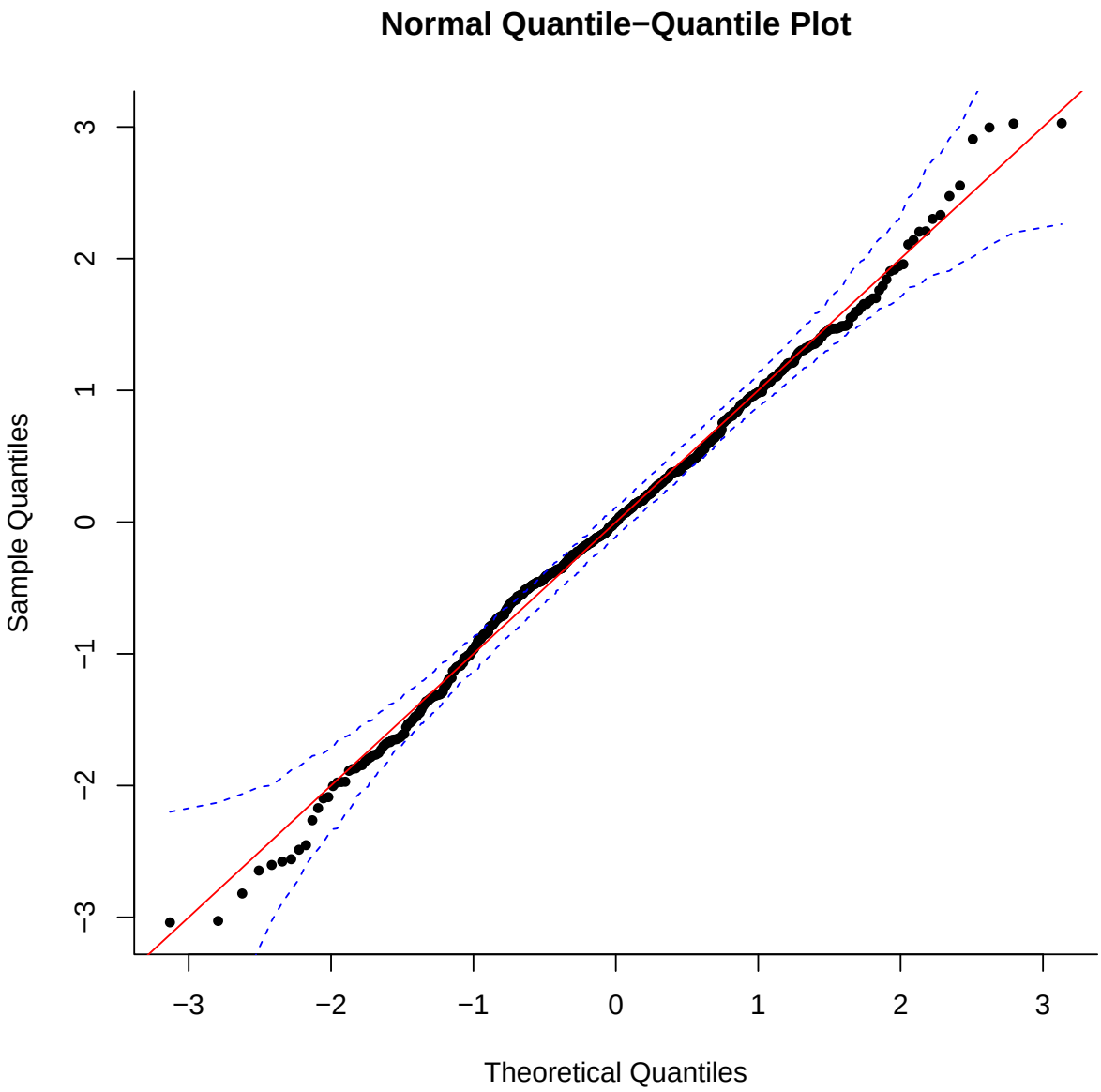


Figure 12: A normal quantile-quantile plot of the residuals with 95% simultaneous confidence bounds for the 700 year mean recurrence interval and the Gumbel type tail length parameter.

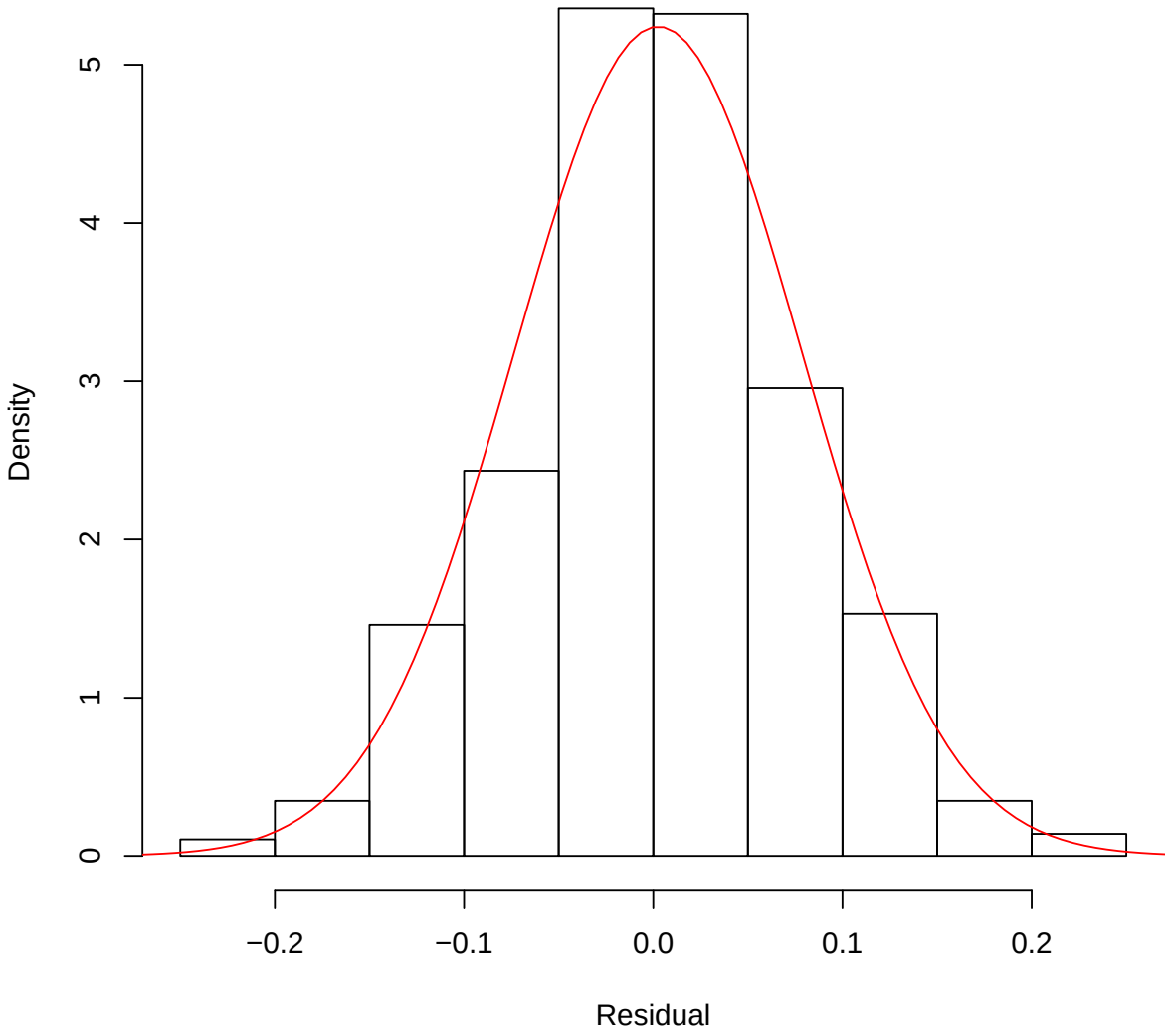


Figure 13: A histogram of the residuals, with overlaid best fitting normal distribution for the map with the Gumbel like tail length parameter and 700 year mean recurrence interval.

where the tail length parameters are estimated from the data, and there is a notable increase in the standard errors of the estimated return values. A reasonable value is typically chosen, and the remaining two parameters are estimated from data. The Gumbel like tail length is taken to be $\zeta_t = 0$ (an obvious connection to the type 1 or Gumbel variant of the Generalized Extreme Value Distribution ([35] page 75)). This is of course implemented by taking a limit to avoid division by zero, and the result is

$$\frac{1}{\psi_t} \exp\left\{\frac{-(y-\omega_t)}{\psi_t}\right\}.$$

Thus, Eq. (1) becomes

$$\lambda(t,y) = \begin{cases} \frac{1}{\psi_s} \exp\left\{\frac{-(y-\omega_s)}{\psi_s}\right\} & \text{for } t \text{ in a thunderstorm period} \\ \frac{1}{\psi_{ns}} \exp\left\{\frac{-(y-\omega_{ns})}{\psi_{ns}}\right\} & \text{for } t \text{ in a non-thunderstorm period.} \end{cases} \tag{10}$$

Gumbel like tail length parameters are well accepted in the wind engineering community and are generally the standard [9].

Results which suggested that an intensity function with a finite tail best fits wind speed data have prompted the authors of [18] to specify a bound that wind speeds are not expected to exceed. This is done by selecting values for ζ_t that are negative and again estimating the other two parameters with data. In [18] $\zeta_t = -0.1$ was proposed. In this report, we also use the more conservative value (higher estimated return values) $\zeta_t = -0.05$. Thus, the intensity functions

$$\lambda(t,y) = \begin{cases} \frac{1}{\psi_s}\left(1 + -0.05 \cdot \frac{y-\omega_s}{\psi_s}\right)_+^{19} & \text{for } t \text{ in a thunderstorm period} \\ \frac{1}{\psi_{ns}}\left(1 + -0.05 \cdot \frac{y-\omega_{ns}}{\psi_{ns}}\right)_+^{19} & \text{for } t \text{ in a non-thunderstorm period.} \end{cases} \tag{11}$$

and

$$\lambda(t,y) = \begin{cases} \frac{1}{\psi_s}\left(1 + -0.1 \cdot \frac{y-\omega_s}{\psi_s}\right)_+^{9} & \text{for } t \text{ in a thunderstorm period} \\ \frac{1}{\psi_{ns}}\left(1 + -0.1 \cdot \frac{y-\omega_{ns}}{\psi_{ns}}\right)_+^{9} & \text{for } t \text{ in a non-thunderstorm period.} \end{cases} \tag{12}$$

are also considered here. The maps constructed using the procedure described in Sec. 3 with the intensity functions in Eqs. (10), (11) and (12) for several mean recurrence intervals between 10 and 100 000 years are found in App. C.

5 R Package for Plotting Maps and Extracting Values

A wind speed map may be thought of as a regular grid of longitude, latitude coordinates, (θ, ϕ), each with a corresponding wind speed value. Thus, a map is well described as a table with three columns, the approach used in this work. However, such a format is not human friendly, so software in the form of an R package named `WindMap` was created to

plot the maps and overlay common geographic boundaries as well as extract values for user specified coordinates. Per the R core development team, "R is a language and environment for statistical computing and graphics." It is an open source project, and the source code or a Windows or Mac binary installer may be freely downloaded from http://www.r-project.org. For Linux systems, R is available from most package managers, or it can be compiled from source. Detailed instructions on compiling R from source on Linux systems (and other systems) can be found at http://www.r-project.org. The R package for creating the maps described in this report is available from the authors upon request at which time installation instructions will also be provided.

To plot a map of 50-year return values over the contiguous United States using the Gumbel like tail length parameters, the package should be installed and loaded into an R session and the following code executed:

```
WindMap(N = 50,
        type = "gumbel",
        uncert = FALSE,
        region = ".",
        long = NULL,
        lat = NULL,
        pi.alpha = 0.05,
        mph.increment = 5,
        dir = "./test_plots",
        file.pt.est = "est.png",
        file.se = "se.png",
        file.cv = "cv.png",
        file.ub = "ub.png")
```

The function `WindMap` is the only user visible function in the package. The definition of each argument to the function `WindMap` is as follows:

- `N` - (integer) The mean recurrence interval in years. Possible values are 10, 25, 50, 100, 300, 700, 1200, 1700, 2000, 2500, 3000, 5000, 10 000, 50 000 and 100 000
- `type` - (character) The value of the tail length parameter. Possible values are `"gumbel"`, `"k.005"` and `"k.01"`
- `uncert` - (boolean) Whether or not to calculate the standard error, coefficient of variation and upper confidence bound
- `region` - (character or `NULL`) `"."` for the contiguous United States, the state name for a specific state and `NULL` if the user wishes to extract values instead of plot a map. So far only `"."` and `NULL` are implemented
- `long` - (numeric or `NULL`) The longitudes at which the user is requesting values or `NULL` if the user wishes to plot a map instead of extracting values
- `lat` - (numeric or `NULL`) The latitudes at which the user is requesting values or `NULL` if the user wishes to plot a map instead of extracting values

- `pi.alpha` - (numeric) The confidence level of the upper bound is (1 - `pi.alpha`)100%
- `mph.increment` - (integer) The difference in mph between contour lines
- `dir` - (character) The directory in which the file containing the plotted map should be saved
- `file.pt.est` - (character) The name of the file in which the map of the estimated expected return values should be saved
- `file.se` - (character) The name of the file in which the map of the standard errors of the estimated expected return values should be saved
- `file.cv` - (character) The name of the file in which the map of the coefficients of variation should be saved
- `file.ub` - (character) The name of the file in which the map of the upper confidence bounds should be saved

To extract values and uncertainties from the maps with a 50 year mean recurrence interval and Gumbel like tail length parameters at the coordinates (-77.20, 39.15) (Gaithersburg, MD) and (-105.26, 40.03) (Boulder, CO), execute the following code:

```
WindMap(N = 50,
        type = "gumbel",
        uncert = TRUE,
        region = NULL,
        long = c(-77.20, -105.26),
        lat = c(39.15, 40.03),
        pi.alpha = 0.05,
        mph.increment = NA,
        dir = "./test_values",
        file.pt.est = "est.png",
        file.se = "se.png",
        file.cv = "cv.png",
        file.ub = "ub.png")
```

To extract values at *any* user specified coordinate, a further interpolation step after the spatial smoothing with local regression is performed within the `WindMap` package (the spatial smoothing is not done within the `WindMap` package, instead the results of it are stored as tables in the package). Any user specified coordinate should be encompassed by a square made of four points from the regular grid that defines the map (boundary cases are handled appropriately but should occur with low frequency). The linear model

$$\widehat{\mu}(\theta, \phi) = c_0 + c_1 h_x(\theta, \phi) + c_2 h_y(\theta, \phi) + c_3 h_x(\theta, \phi) h_y(\theta, \phi)$$

is fitted to those four points using ordinary least squares. Recall from Sec. 3.2 that h_x and h_y represent a Lambert conformal conic projection of the longitude and latitude coordinates. This creates a surface that passes directly through the values associated with the corners of the square, and which may be used to interpolate to any coordinate within the square.

6 Conclusion

In this report, an approach to creating maps of N-year return values for N between 10 and 100 000 years was described. The approach proceeded in two stages. In the first stage, a two-dimensional non-homogeneous Poisson process model was fitted to the data from each station using maximum likelihood. The raw data from each station was not directly used. The data were thresholded and declustered before fitting, and time gaps between observations of six months or more were removed from the time dimension of the Poisson process domain. The fitted Poisson process models were then used to estimate the return values of interest.

In the second stage, the natural logarithms of the estimated return values were spatially smoothed using local regression to create maps over the contiguous United States. The neighborhood size of the local regression was dynamic and guided by generalized cross validation. The final choice of the neighborhood size was based on the appropriate smoothness of the maps and it was validated by examining residuals. The degree of the local polynomial was set to unity. The standard uncertainty was defined as the standard deviation of the prediction of a new hypothetical return value, and an upper confidence bound was calculated by leveraging the normal distribution. So that it is convenient to plot the maps or extract values from them, an R package with those goals was developed.

Maps of return values as well as their corresponding maps of standard errors are found in App. C. In all of the maps in App. C, the same general trend is apparent, smaller values on the coasts and larger values in the middle. Another trend (which is expected) is that as the tail length parameters change from Gumbel to -0.05, then to -0.1, the return values decrease. For the large mean recurrence intervals, e.g. 100 000 years, the decrease is considerable.

Acknowledgements

This work is part of a cooperative effort by the National Windstorm Impact Reduction Program, Engineering Laboratory, NIST and the Statistical Engineering Division, Information Technology Laboratory, NIST, aimed at developing estimates of design wind speeds for the contiguous United States.

Bibliography

[1] R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24:180–190, 1928.

[2] B. V. Gnedenko. Sur la distribution limite du terme maximum d'une série aléatoire. *The Annals of Mathematics*, 44:423–453, 1943.

[3] III Pickands, J. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3:119–131, 1975.

[4] A. C. Davidson and R. L. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society (Series B Methodological)*, 52:393–442, 1990.

[5] III Pickands, J. The two-dimensional poisson process and extremal processes. *Journal of Applied Probability*, 8:745–756, 1971.

[6] R. L. Smith. Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science*, 4:367–393, 1989.

[7] S. Coles. The use and misuse of extreme value models in practice. In B. Finkenstädt and H. Rootzén, editors, *Extreme Values in Finance, Telecommunications, and the Environment*, chapter 2, pages 79–100. Chapman & Hall/CRC, 2004.

[8] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. URL http://www.R-project.org/.

[9] ASCE. *Minimum Design Loads for Buildings and Other Structures ASCE/SEI 7-10*. American Society of Civil Engineers, Reston, Virginia, 2010.

[10] F. J. Masters, P. J. Vickery, P. Bacon, and E. N. Rappaport. Toward objective, standardized intensity estimates from surface wind speed observations. *Bulletin of the American Meteorological Society*, 91(12):1665–1681, 2010.

[11] R. L. Smith. Statistics of extremes, with applications in environment, insurance, and finance. In B. Finkenstädt and H. Rootzén, editors, *Extreme Values in Finance, Telecommunications, and the Environment*, chapter 1, pages 1–78. Chapman & Hall/CRC, 2004.

[12] J.R. Norris. *Markov chains*. Cambridge Univ Pr, 1998.

[13] S.I. Resnick. *Adventures in stochastic processes*. Birkhauser, 1992.

[14] R.D. Reiss. *A course on point processes*. Springer-Verlag New York, 1993.

[15] P. J. Bickel and K. A. Doksum. *Mathematical Statistics Basic Ideas and Selected Topics*. Prentice Hall, second edition, 2001.

[16] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7:308–313, 1965.

[17] E. C. Mannshardt-Shamseldin, R. L. Smith, S. R. Sain, L. O. Mearns, and D. Cooley. Downscaling extremes: A comparison of extreme value distributions in point-source and gridded precipitation data. *The Annals of Applied Statistics*, 4:484–502, 2010.

[18] F. T. Lombardo, J. A. Main, and E. Simiu. Automated extraction and classification of thunderstorm and non-thunderstorm wind data for extreme-value analysis. *Journal of wind Engineering and Industrial Aerodynamics*, 97:120–131, 2009.

[19] A. L. Pintar and F. T. Lombardo. Mapping return values of extreme wind speeds. In M.-L. T. Lee, M. Gail, R. Pfeiffer, G. Satten, T. Cai, and A. Gandy, editors, *Risk Assessment and Evaluation of Predictions*. Springer, in press 2014.

[20] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury, second edition, 2002.

[21] Jun Shao. *Mathematical Statistics*. Springer, 2003.

[22] III Pickands, J. Bayes quantile estimation and threshold selection for the generalized pareto family. In J. Galambos, J. Lechner, and E. Simiu, editors, *Extreme Value Theory and Applications*, chapter 2, pages 123–138. Kluwer Academic Publishers, 1994.

[23] J. Neter, M. H. Kutner, C. J. Nachtsheim, and W. Wasserman. *Applied Linear Statistical Models*. WCB/McGraw-Hill, Boston, Massachusetts, fourth edition, 1996.

[24] H. C. Shellard. The estimation of design wind speeds. In *Wind Effects on Buildings and Structures Vol. 1*, pages 29–52. London: Her Majesty's Stationary Office, 1965.

[25] C. Loader. *locfit: Local Regression, Likelihood and Density Estimation.*, 2013. URL http://CRAN.R-project.org/package=locfit. R package version 1.5-9.1.

[26] P. Craven and T. Wahba. Smoothing noisy data with spline functions. *Numerische Mathematik*, 31:377–403, 1979.

[27] W. S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74:829–836, 1979.

[28] W. S. Cleveland and S. J. Devlin. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83:596–610, 1988.

[29] C. Loader. *Local regression and likelihood*. Springer Verlag, 1999.

[30] J.P. Snyder. *Map projections–a working manual*. United States Government Printing Office, 1987.

[31] Doug McIlroy. Packaged for R by Ray Brownrigg, Thomas P Minka, and transition to Plan 9 codebase by Roger Bivand. *mapproj: Map Projections*, 2011. URL http://CRAN.R-project.org/package=mapproj. R package version 1.1-8.3.

[32] NIST/SEMATECH. e-handbook of statistical methods. http://www.itl.nist.gov/div898/handbook/mpc/section5/mpc55.htm. accessed 2013-10-25.

[33] Angelo Canty and B. D. Ripley. *boot: Bootstrap R (S-Plus) Functions*, 2015. R package version 1.3-16.

[34] A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge, 1997. URL http://statwww.epfl.ch/davison/BMA/. ISBN 0-521-57391-2.

[35] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions: Volume 2*. Wiley, second edition, 1995.

[36] E. Simiu, J. Bietry, and J. J. Filliben. Sampling errors in estimation of extreme winds. *Journal of The Structural Division, ASCE*, 104(3):491–501, March 1978.

Appendices

A Poisson Process Generalized Pareto Equivalency

The Poisson process approach to extreme values is closely connected to the generalized Pareto approach, which may be more familiar to some. In fact, under certain restrictions, they are equivalent. The Poisson Process approach was chosen for this work because it is more flexible in general than the generalized Pareto approach. See [7], page 91 for several advantages of the Poisson process approach over the generalized Pareto approach. In this appendix, the N-year return value is derived for both approaches under a specified set of assumptions for each to highlight a set of assumptions necessary for equivalence.

A.1 Generalized Pareto Return Value

For reference, the cumulative distribution function for the generalized Pareto distribution is $F(y) = 1 - (1 + \xi\frac{y-b}{\sigma})_+^{-1/\xi}$, where b is the threshold (see [11] page 9). The generalized Pareto approach makes two assumptions. First, the magnitude of threshold crossings are assumed to be independent and identically distributed generalized Pareto random variables, with scale parameter σ and tail length parameter ξ. Second, the times of the threshold crossings are assumed to be governed by a one dimensional homogeneous Poisson process with parameter γ. Note that none of σ, ξ, γ or the threshold b are indexed by time, and so could not differentiate between thunderstorm and non-thunderstorm wind types. To derive the return value, the probability of observing one or more values larger than y_N is calculated.

$$\begin{aligned}
&\text{Prob(1 or more values larger than } y_N\text{)} \\
&\quad = \text{ Prob(1 crossing and } Y_1 > y_N\text{) +} \\
&\qquad + \text{ Prob(2 crossings and } Y_1 > y_N \text{ or } Y_2 > y_N\text{) } + \ \ldots \qquad (13) \\
&\quad = \gamma e^{-\gamma}\left(1 + \xi\frac{y_N - b}{\sigma}\right)_+^{-1/\xi} + \\
&\qquad \sum_{n=2}^{\infty}\frac{\gamma^n e^{\gamma}}{n!}\left\{1 - \left[1 - \left(1 + \xi\frac{y_N - b}{\sigma}\right)_+^{-1/\xi}\right]^n\right\}.
\end{aligned}$$

However, Eq. (13) simplifies to

$$1 - \exp\left\{-\gamma\left(1 + \xi\frac{y_N - b}{\sigma}\right)_+^{-1/\xi}\right\}. \qquad (14)$$

So the return value is

$$y_n = \frac{\sigma\gamma^{\xi}}{\xi}\left[-\log\left(\frac{N-1}{N}\right)\right]^{-\xi} - \frac{\sigma}{\xi} + b \tag{15}$$

A.2 Poisson Process Return Value

Assuming the intensity function

$$\lambda(y,t) \;=\; \frac{1}{\psi}\left(1+\zeta\frac{y-\omega}{\psi}\right)_{+}^{-1/\zeta-1}, \tag{16}$$

Section 3.1.2 states that the return value is the solution to the equation

$$1-\exp\left\{-\left[1+\zeta\frac{y_N-\omega}{\psi}\right]_{+}^{-1/\zeta}\right\} = \frac{1}{N}. \tag{17}$$

Note that the right side of Eq. (16) is not actually a function of t, which is a deviation from the intensity functions described in Sec. 4, where the intensity functions differentiate between thunderstorm and non-thunderstorm wind types. From Eq. (17), the return value is derived to be

$$y_N = \frac{\psi}{\zeta}\left[-\log\left(\frac{N-1}{N}\right)\right]^{-\zeta} - \frac{\psi}{\zeta} + \omega. \tag{18}$$

Thus, if

$$\xi = \zeta,$$

$$\sigma = -\zeta\left[\frac{-\psi}{\zeta} + \omega - b\right]$$

and

$$\gamma = \left[\frac{-\psi}{\zeta\left(\frac{-\psi}{\zeta}+\omega-b\right)}\right]^{1/\zeta}$$

the return values are equal.

B Full Estimation

The intensity functions considered in Sec. 4 fixed ζ_s and ζ_{ns} at three different values, 0 (Gumbel), -0.05 and -0.1. It was noted in Sec. 3.1.4 that there was not good agreement between the data and fitted Gumbel model for Station 722287 because an outlier was present.

The agreement when the tail length parameters, ζ_s and ζ_{ns} are fixed at -0.05 or -0.1 is worse. To improve the agreement between the data and the Poisson process model, the tail length parameters, ζ_s and ζ_{ns} may be included as parameters to estimate instead of fixed at specific values. Figure 14 shows far better agreement between the data for Station 722287 and the Poisson process model. The estimates of the tail length parameters ζ_s and ζ_{ns} are -0.05 and 0.307, respectively. Note the positive estimate of ζ_{ns} to better accommodate the outlier.

Figure 15 shows a map of estimated expected 50-year return values and a map of their standard errors when the tail length parameters of the Poisson process model are estimated using the data instead of being specified. The general spatial trend of higher wind speeds in the center of the country and lower wind speeds on the coasts is found in the map of expected return values in Fig. 15, just as it is in Figs. 18, 33 and 48. The biggest difference in spatial trends is a clear upward trend in Florida in Fig. 15 whereas Figs. 18, 33, and 48 depict no trend in the estimated expected return values in Florida. Further, the magnitudes of the estimated expected return values in Fig. 15 are slightly lower than those in Fig. 18, but they are quite similar to those in Fig. 33 and slightly higher than those in 48. The magnitudes of the standard errors of the expected return values in Fig. 15 on the other hand are considerably higher (more than 20%) than those in Figs. 18, 33 and 48. This is sensible because estimating the tail length parameters adds an additional source of uncertainty, and the tail length parameters tend to be the most difficult of the parameters to estimate well.

In keeping with the practice of the wind engineering community, no maps that include estimation of the tail length parameters are recommended for use. This recommendation is consistent with the results reported in [36]. The approach is discussed here only to show that estimating the tail length parameters can result in better agreement between the model and the data for some stations, and that the resulting maps may not be radically different than those created using the traditional approaches. Another interesting observation from this exercise is the proportion of tail length parameters estimated to be negative (implying the existence of finite maximum wind speed). For the thunderstorm wind type, the proportion is 0.84, and for the non-thunderstorm wind type, it is 0.76. The facts that those proportions are appreciably larger than 0.5 and that the estimates in Fig. 15 are slightly lower in most places than those in Fig. 18 suggest that the Gumbel like tail length is conservative (higher estimated return values compared to other models), a reasonable position to support in this setting.

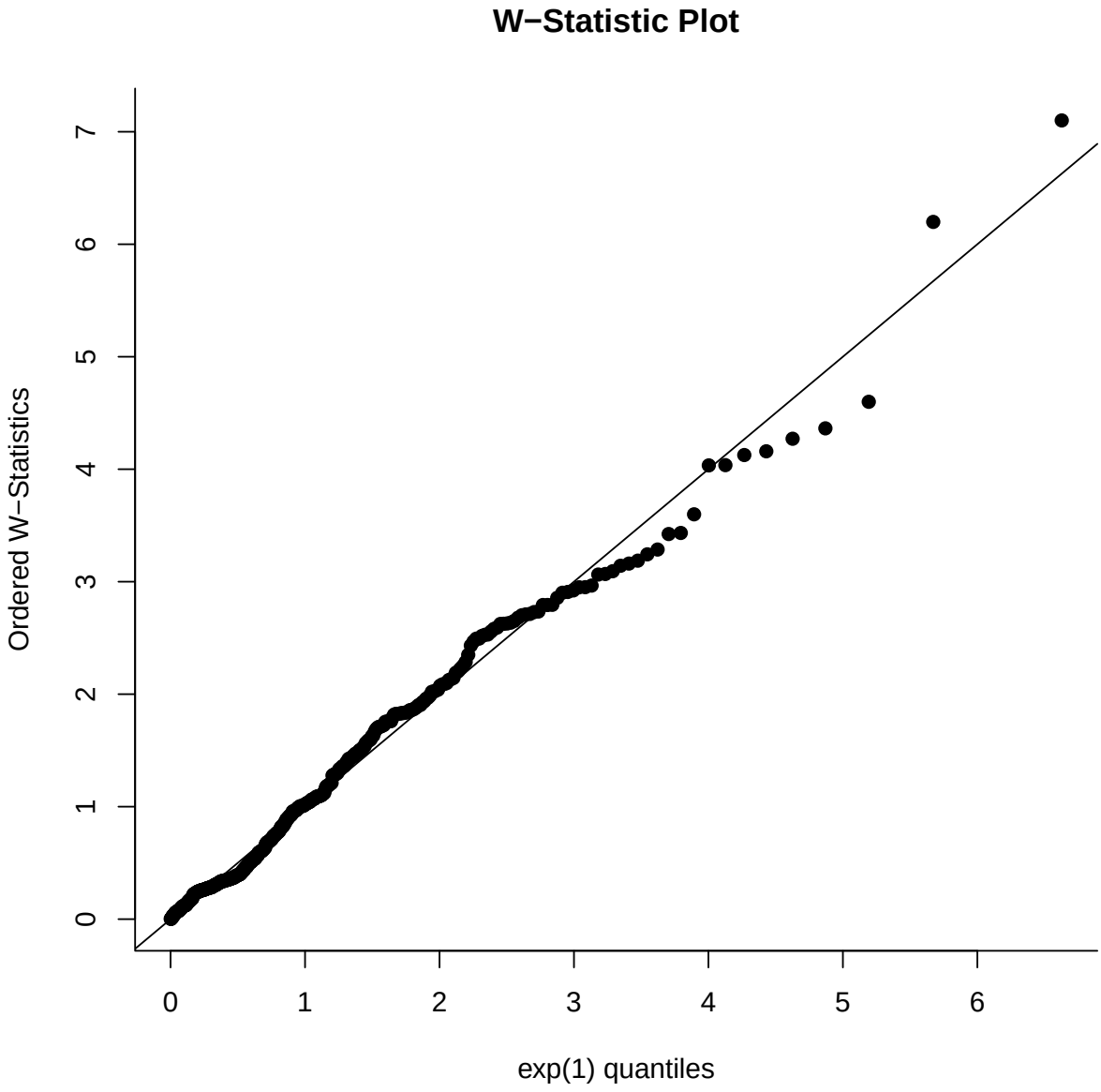
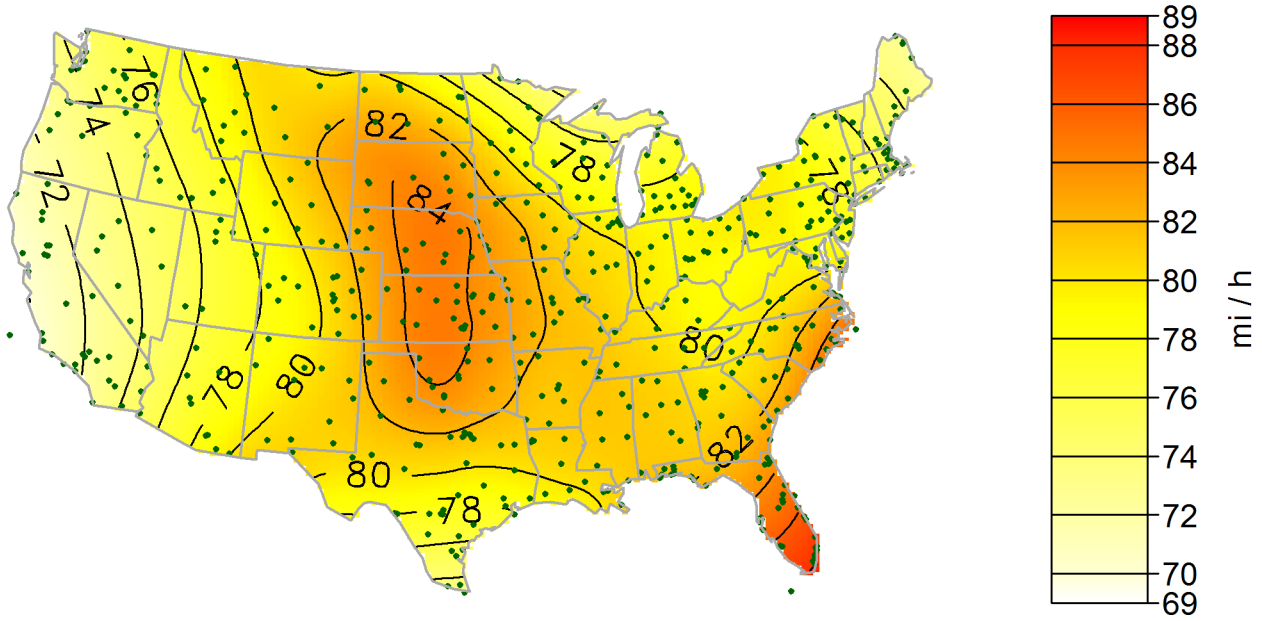
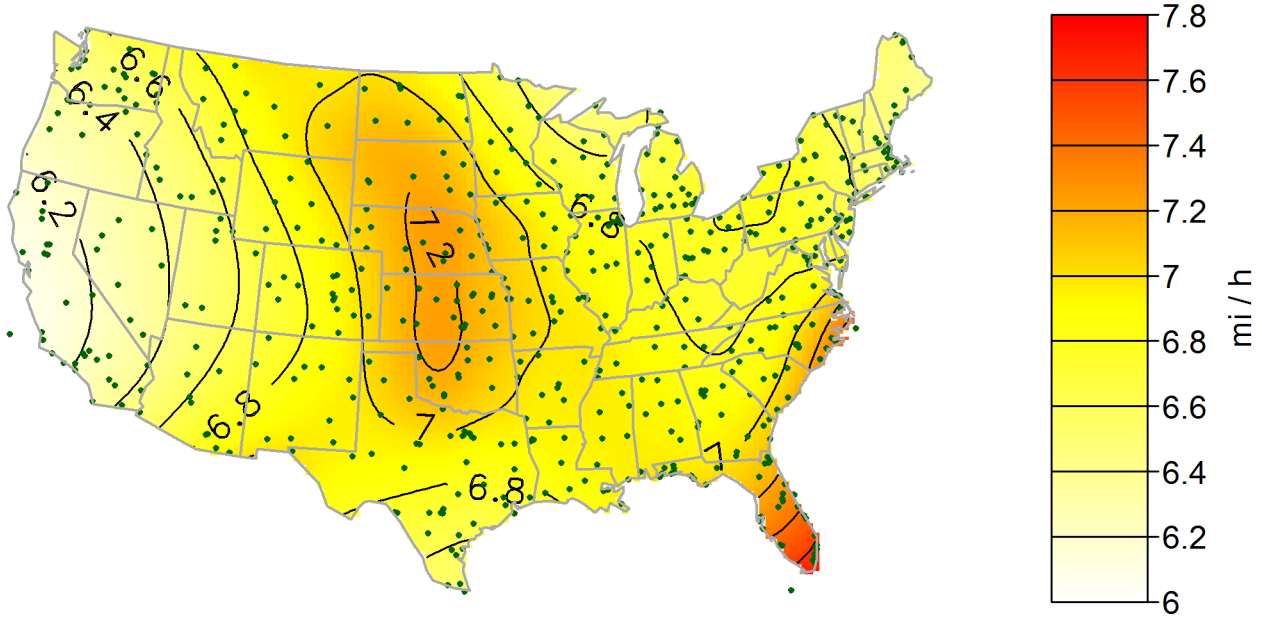


Figure 14: Plot of ordered W-statistics versus Exp(1) percentiles for Station 722287 with full estimation.



(a) Map of estimated expected 50-year return values interval.



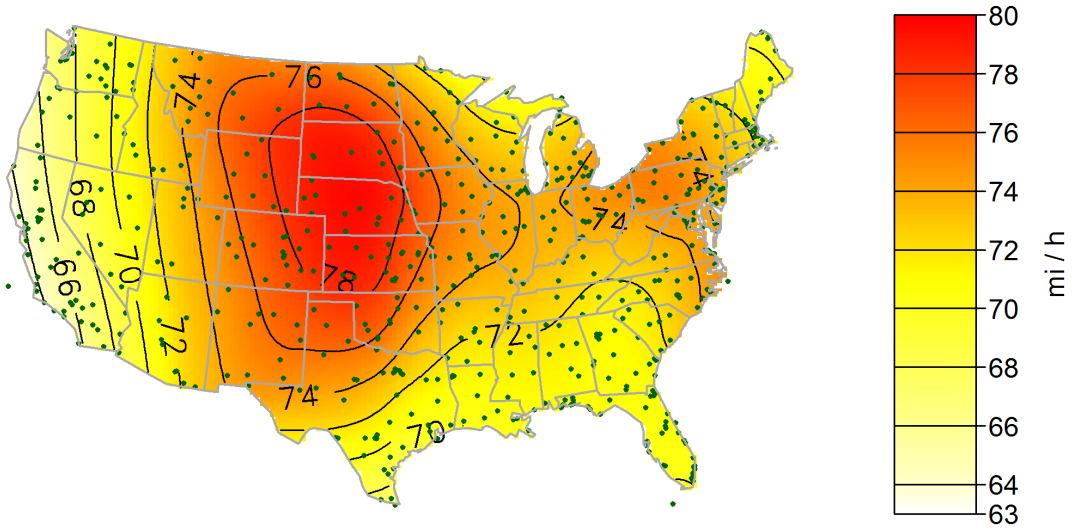
(b) Map of standard errors of estimated expected 50-year return values.

Figure 15: Maps for a 50 year mean recurrence interval with full estimation, i.e. estimating the tail length parameters, ζ_s and ζ_{ns}.

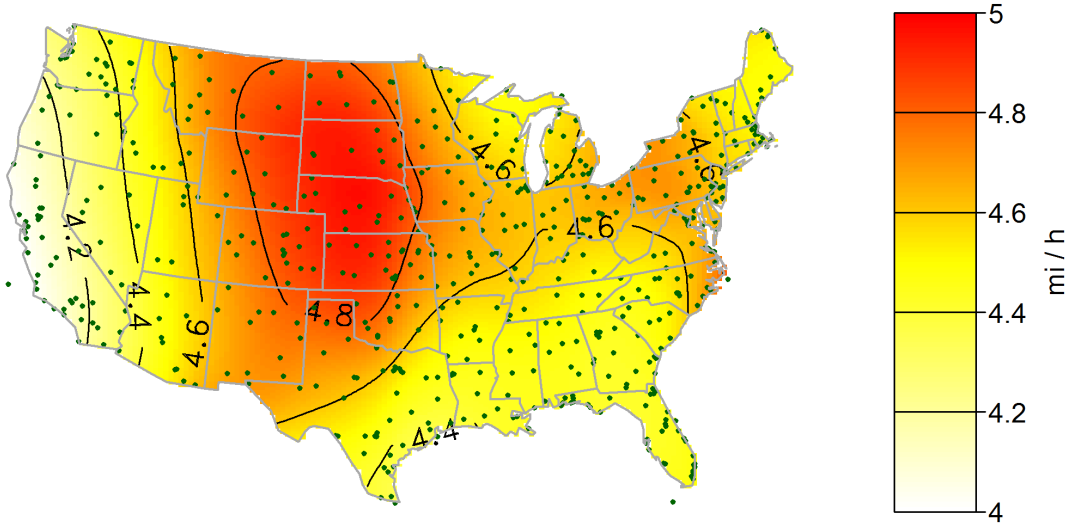
C All Maps

The units in all maps are mi/h. Multiply by 1.6 to get km/h

C.1 Gumbel Tail Length

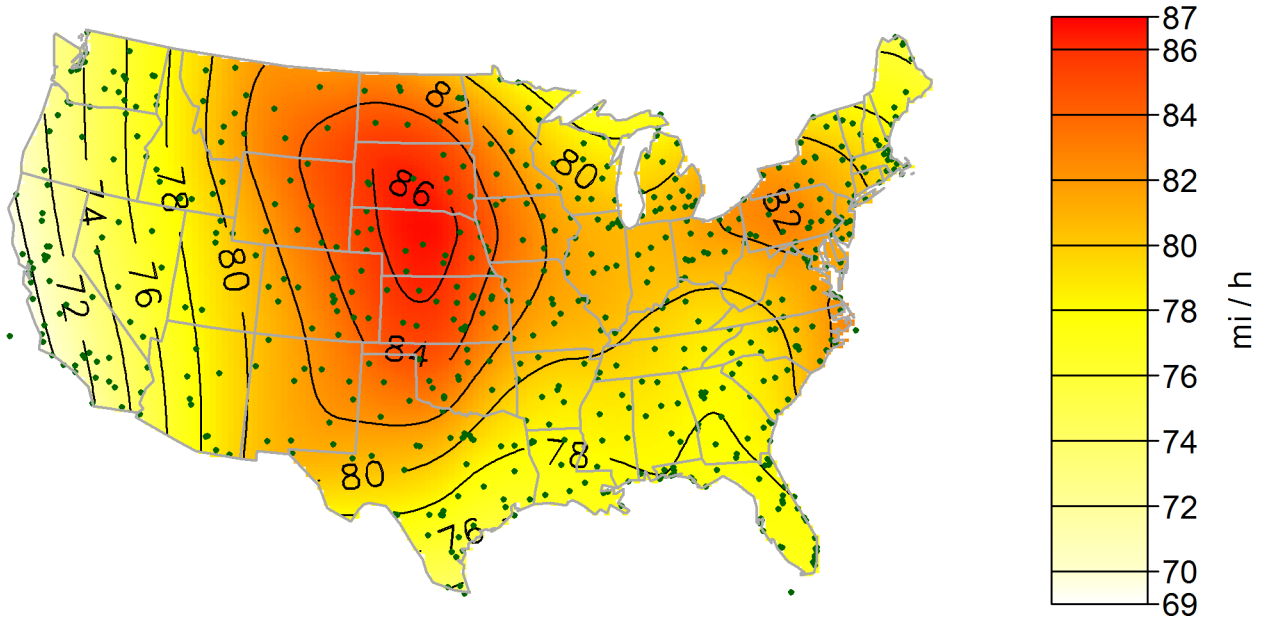


(a) Map of estimated expected 10-year return values.

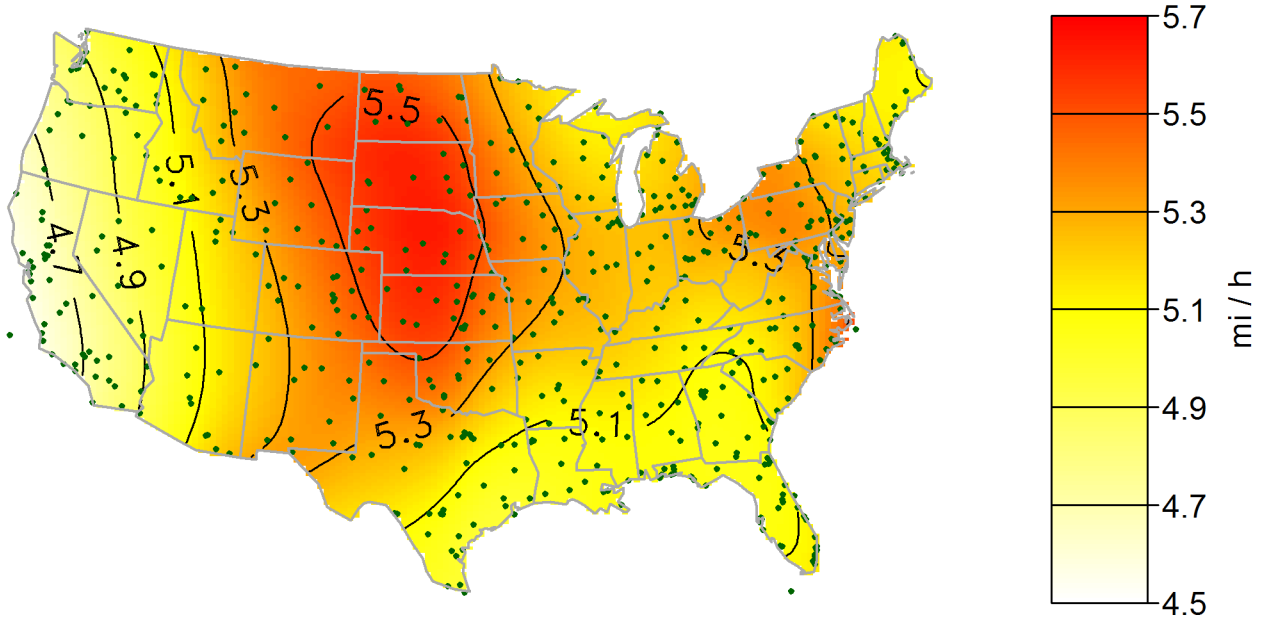


(b) Map of standard errors of estimated expected 10-year return values.

Figure 16: Gumbel maps for a 10 year mean recurrence interval.

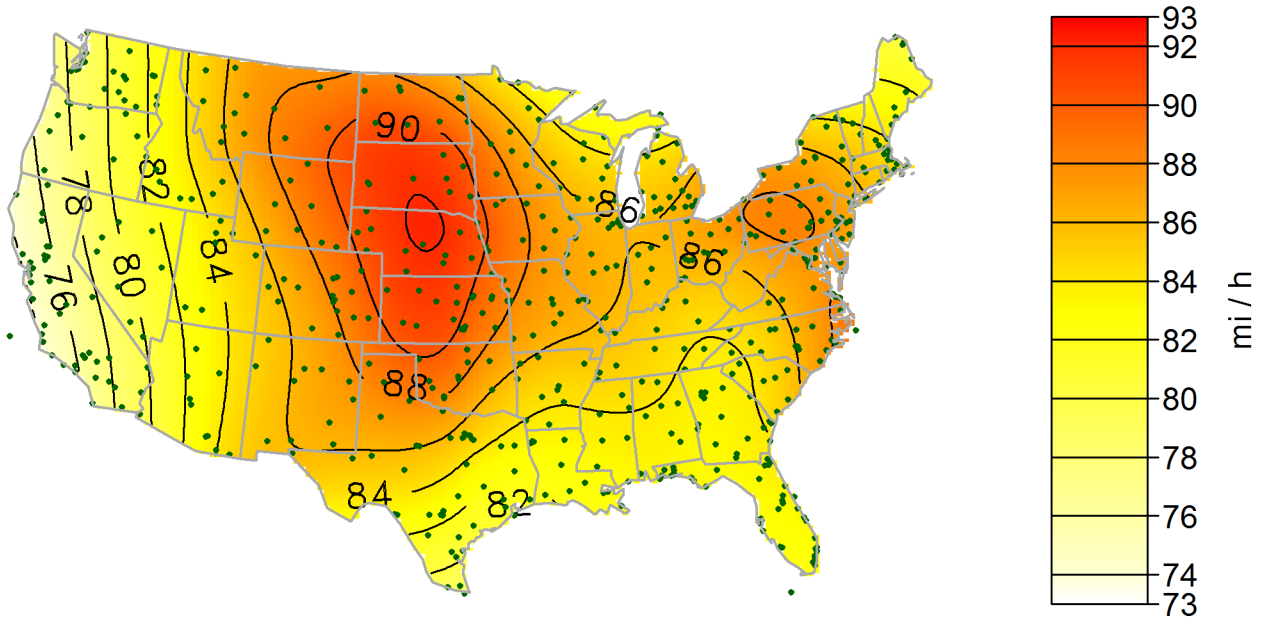


(a) Map of estimated expected 25-year return values.

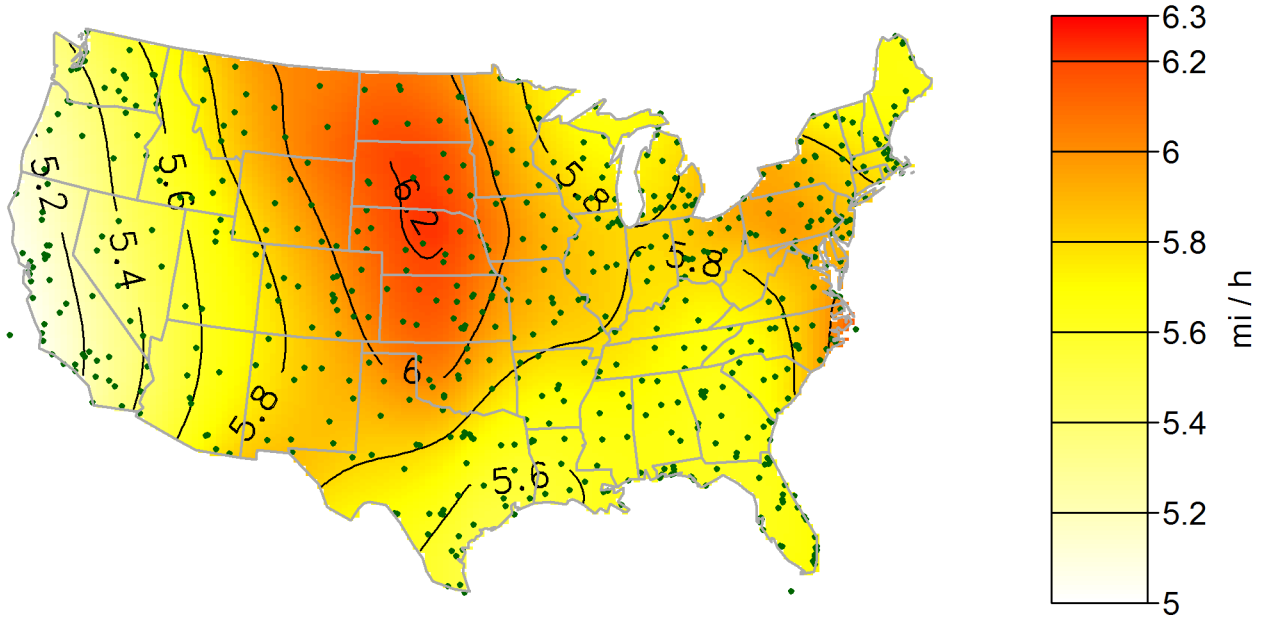


(b) Map of standard errors of estimated expected 25-year return values.

Figure 17: Gumbel maps for a 25 year mean recurrence interval.

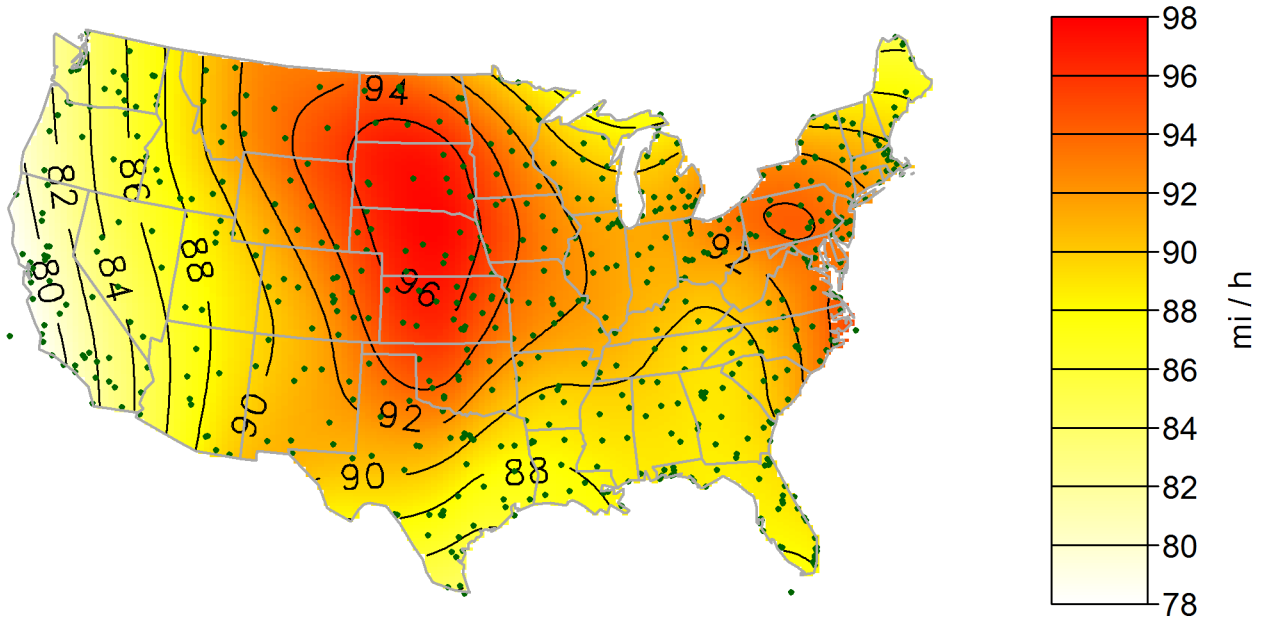


(a) Map of estimated expected 50-year return values.

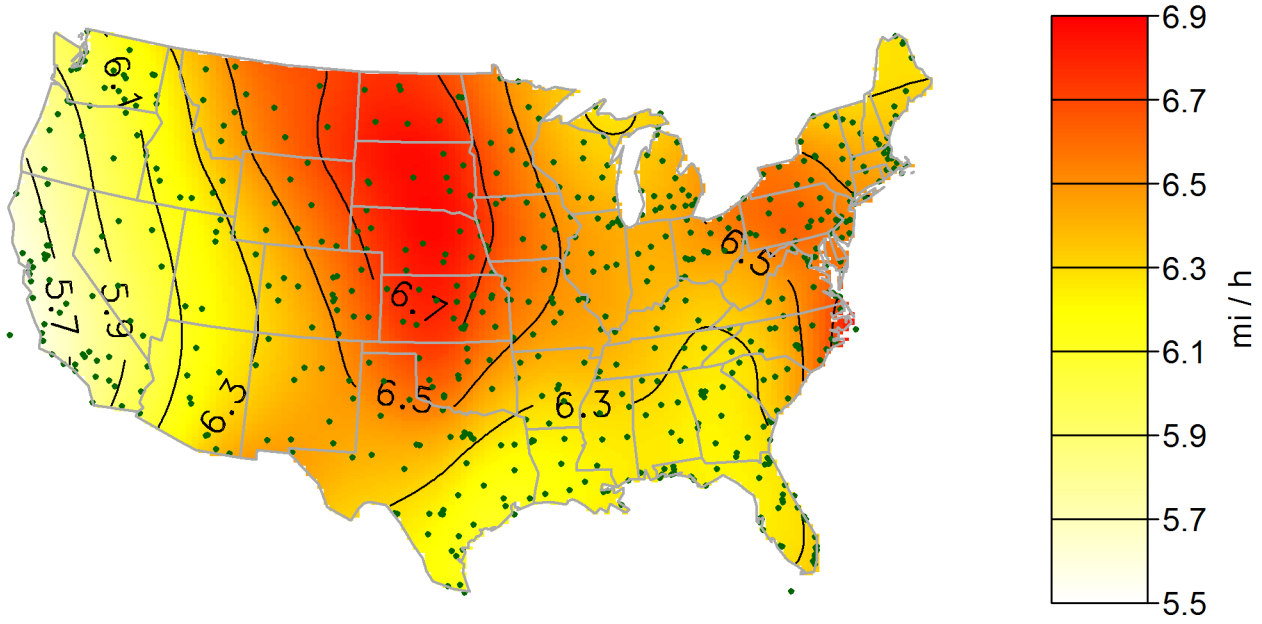


(b) Map of standard errors of estimated expected 50-year return values.

Figure 18: Gumbel maps for a 50 year mean recurrence interval.

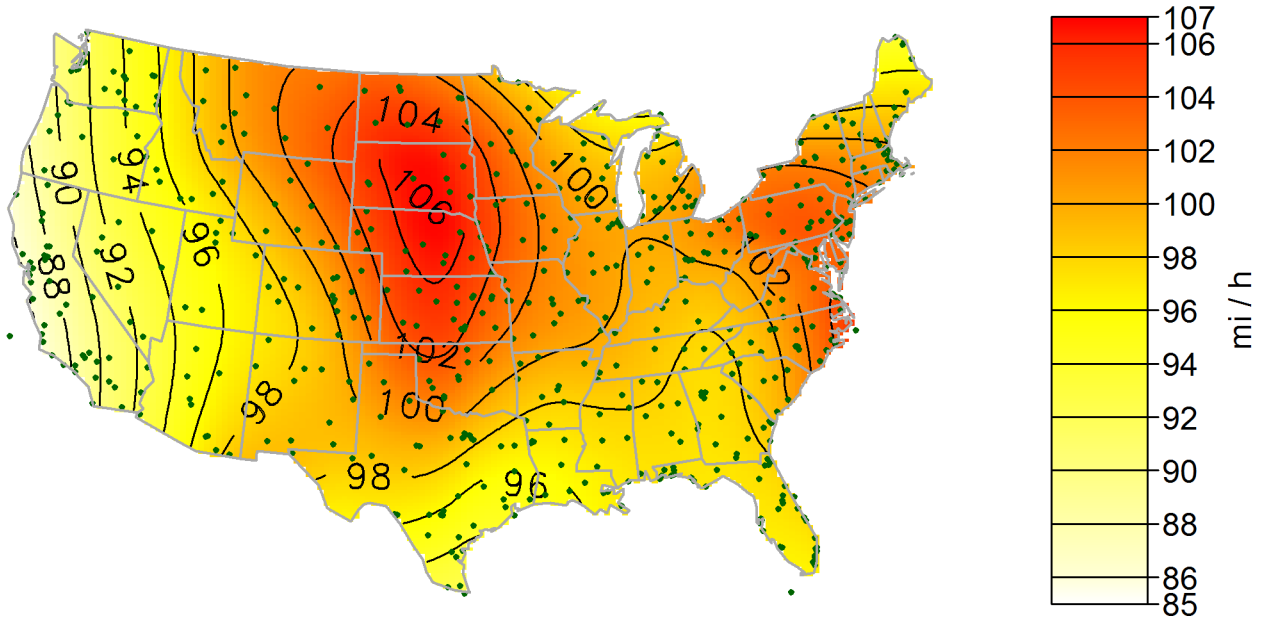


(a) Map of estimated expected 100-year return values.

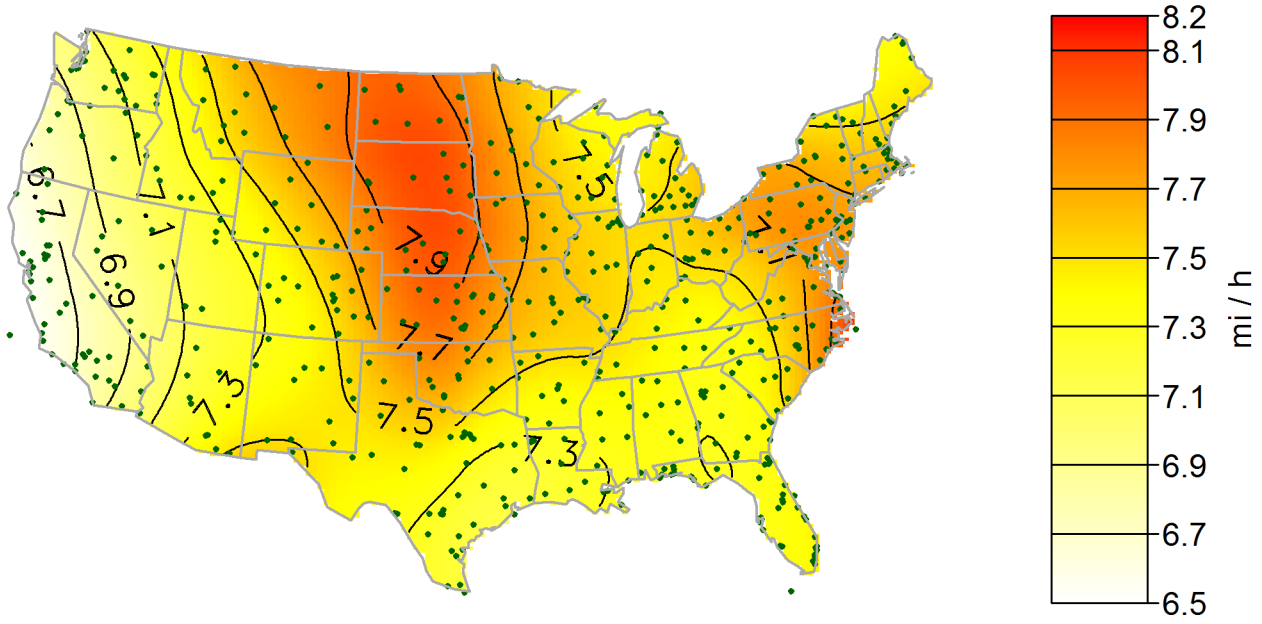


(b) Map of standard errors of estimated expected 100-year return values.

Figure 19: Gumbel maps for a 100 year mean recurrence interval.

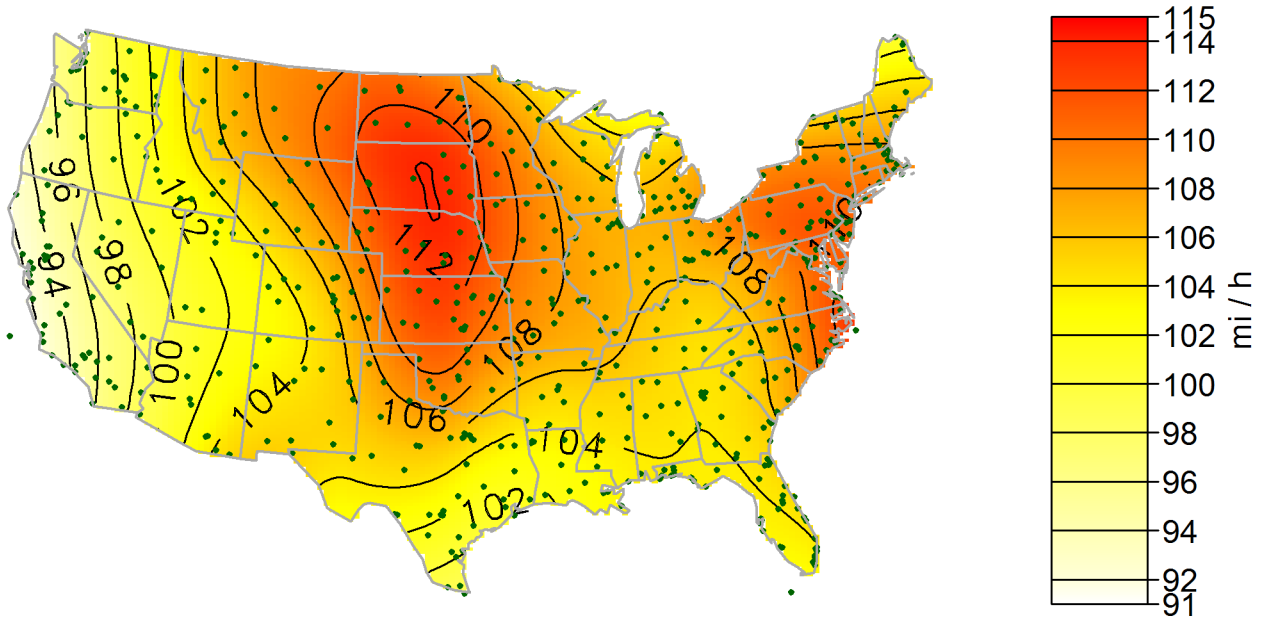


(a) Map of estimated expected 300-year return values.

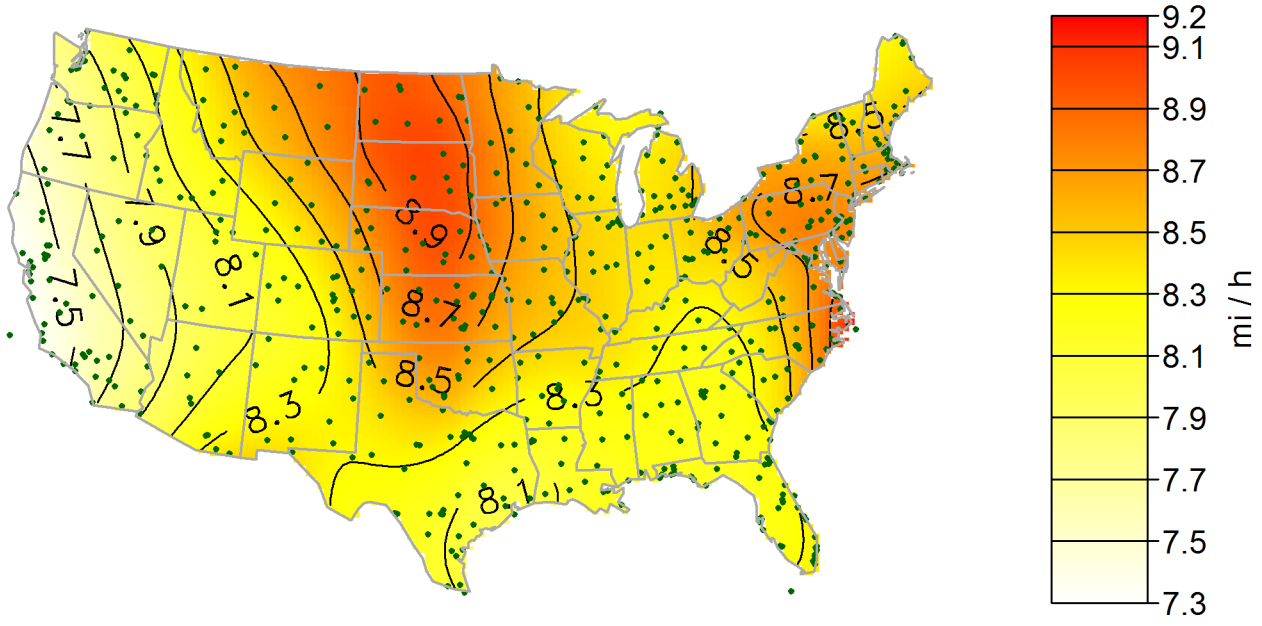


(b) Map of standard errors of estimated expected 300-year return values.

Figure 20: Gumbel maps for a 300 year mean recurrence interval.

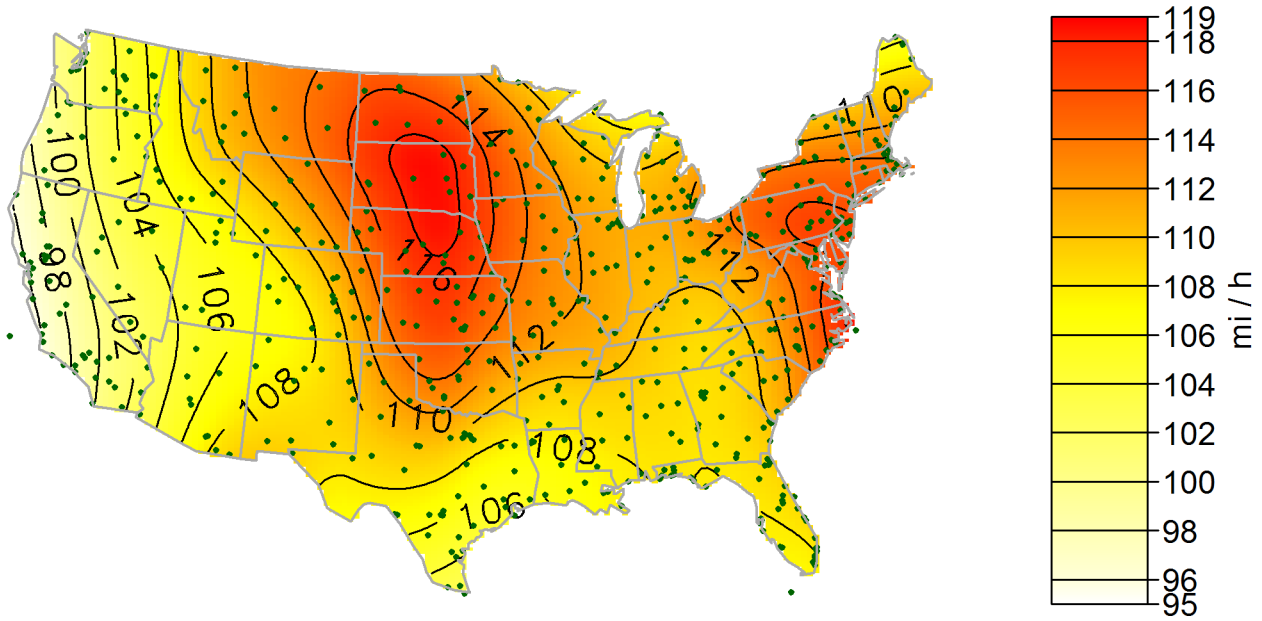


(a) Map of estimated expected 700-year return values.

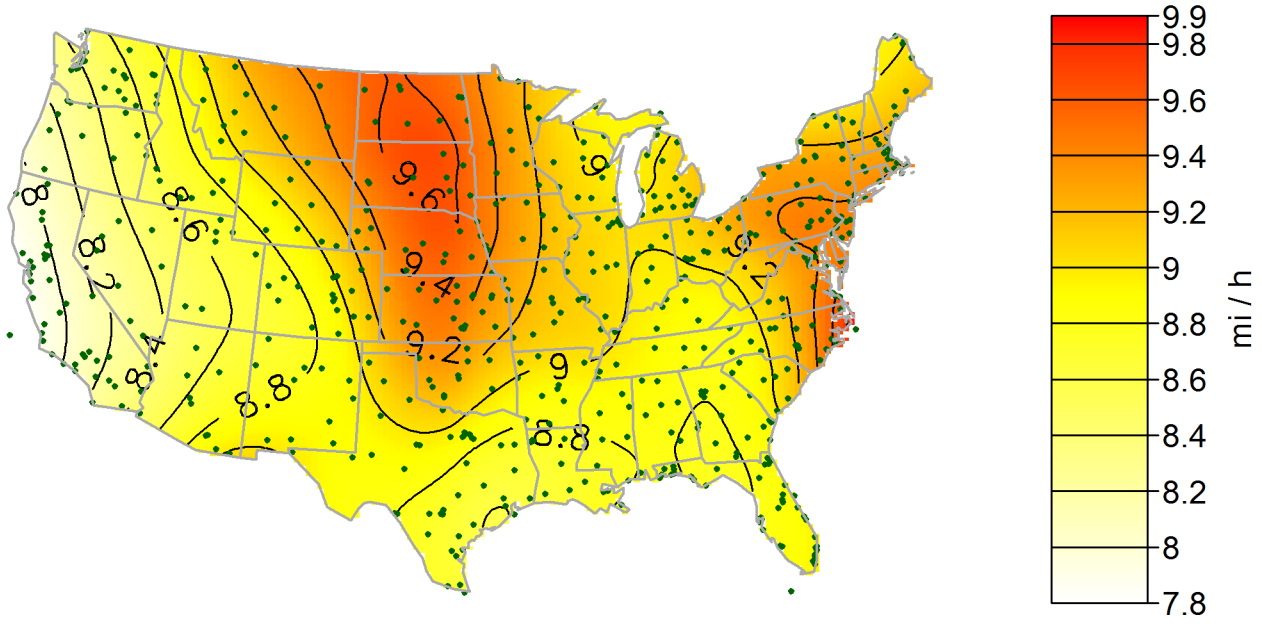


(b) Map of standard errors of estimated expected 700-year return values.

Figure 21: Gumbel maps for a 700 year mean recurrence interval.

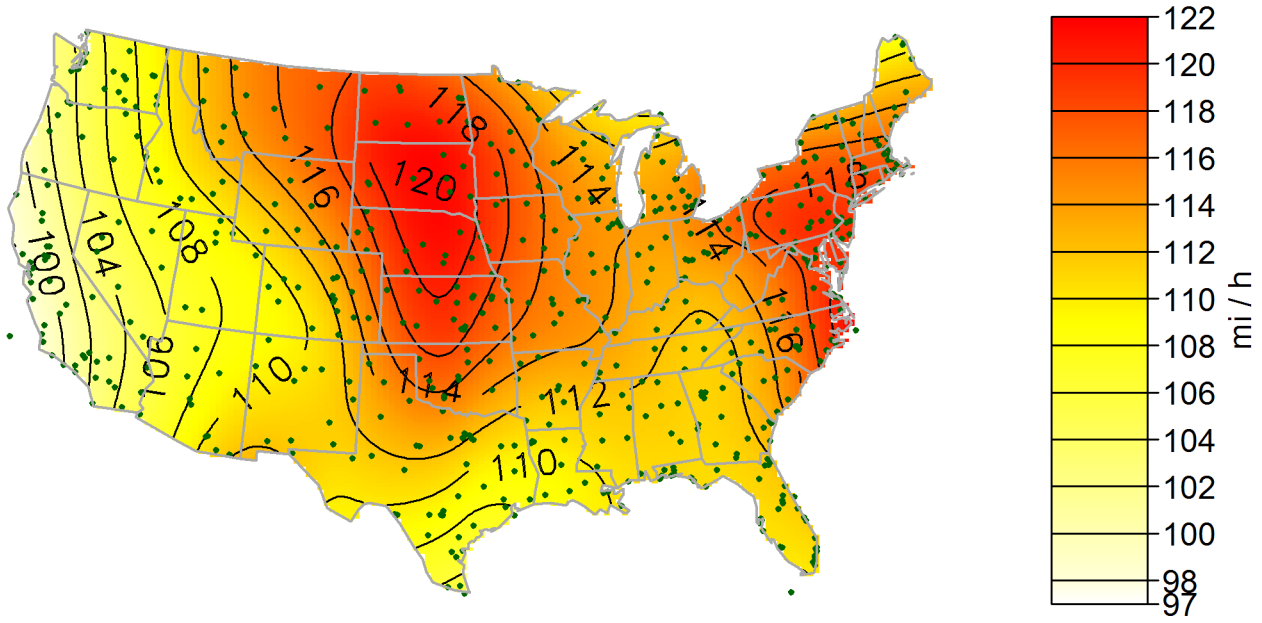


(a) Map of estimated expected 1200-year return values.

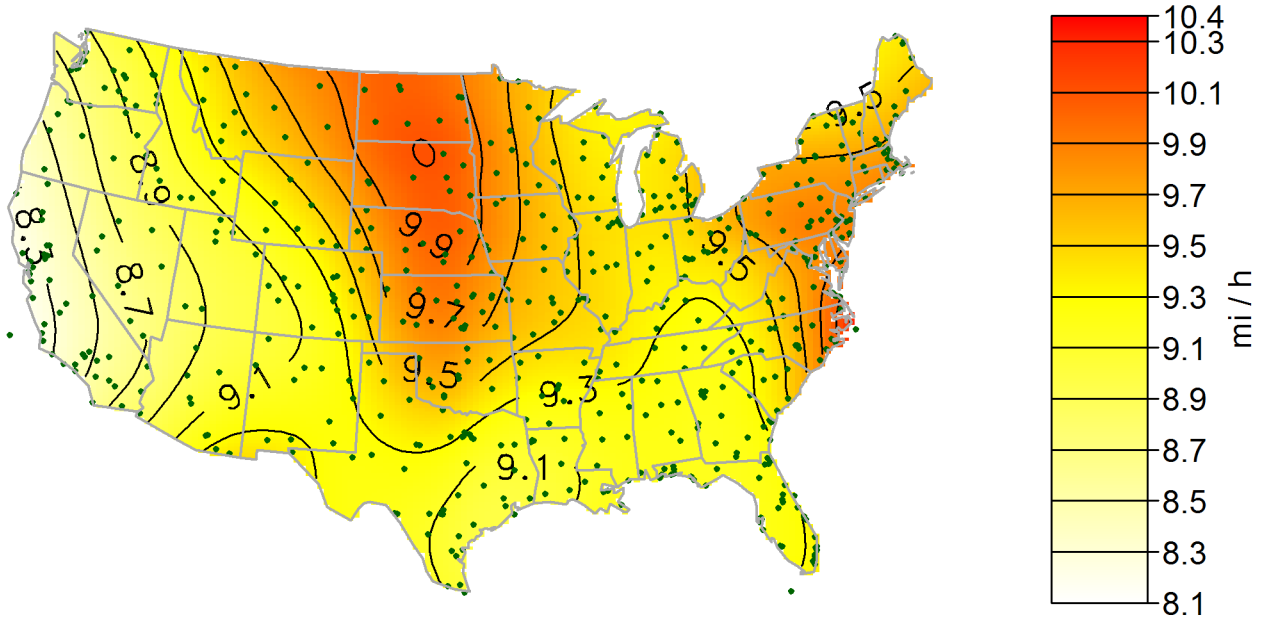


(b) Map of standard errors of estimated expected 1200-year return values.

Figure 22: Gumbel maps for a 1200 year mean recurrence interval.

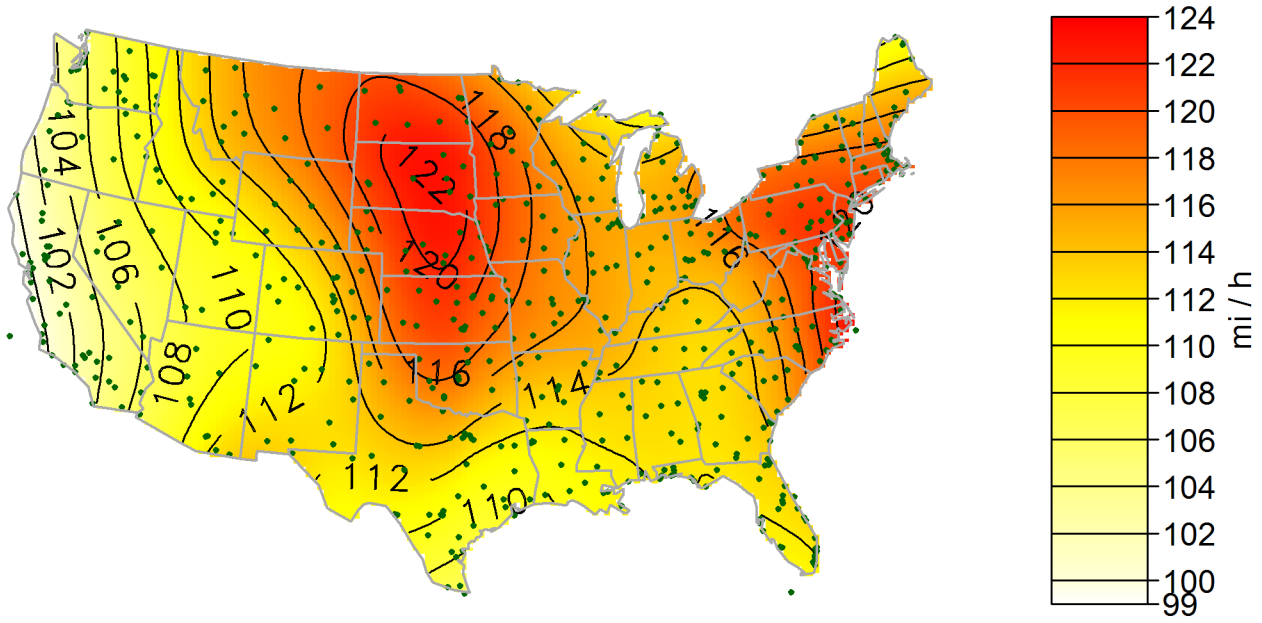


(a) Map of estimated expected 1700-year return values.

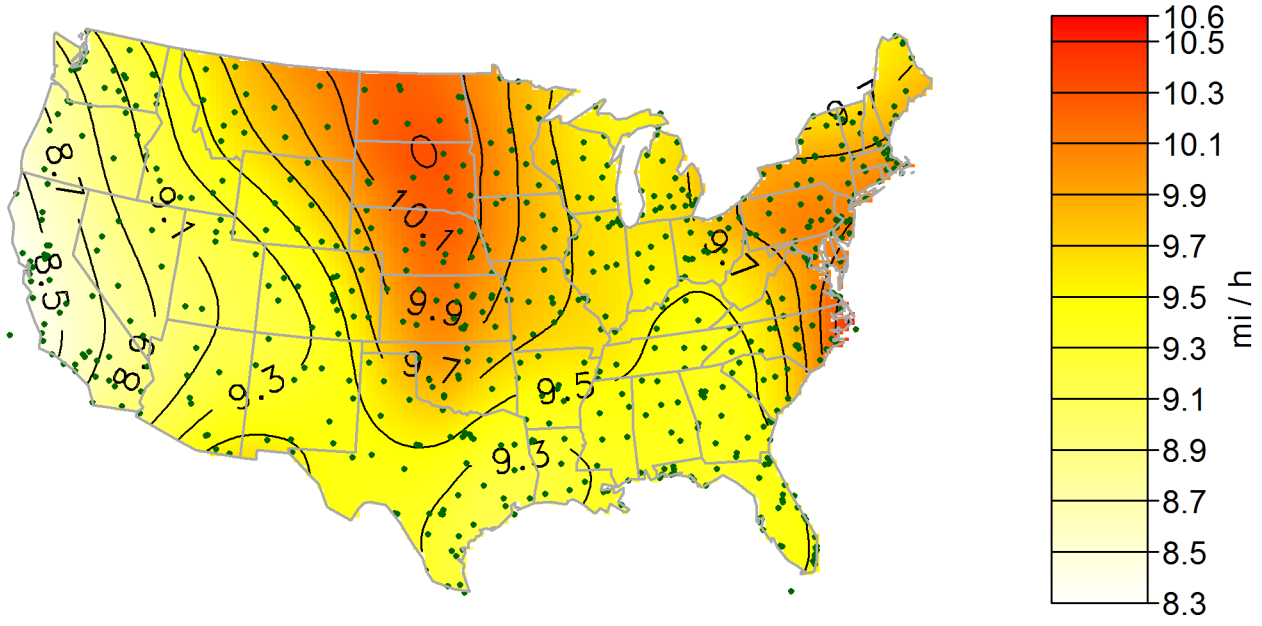


(b) Map of standard errors of estimated expected 1700-year return values.

Figure 23: Gumbel maps for a 1700 year mean recurrence interval.

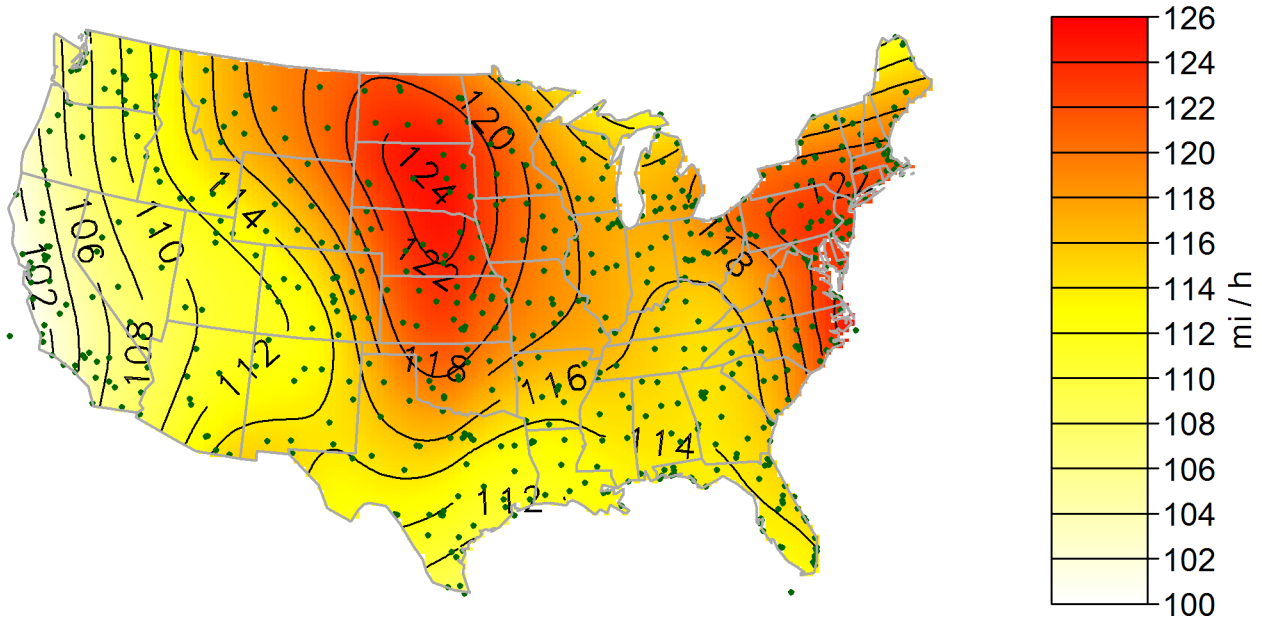


(a) Map of estimated expected 2000-year return values.

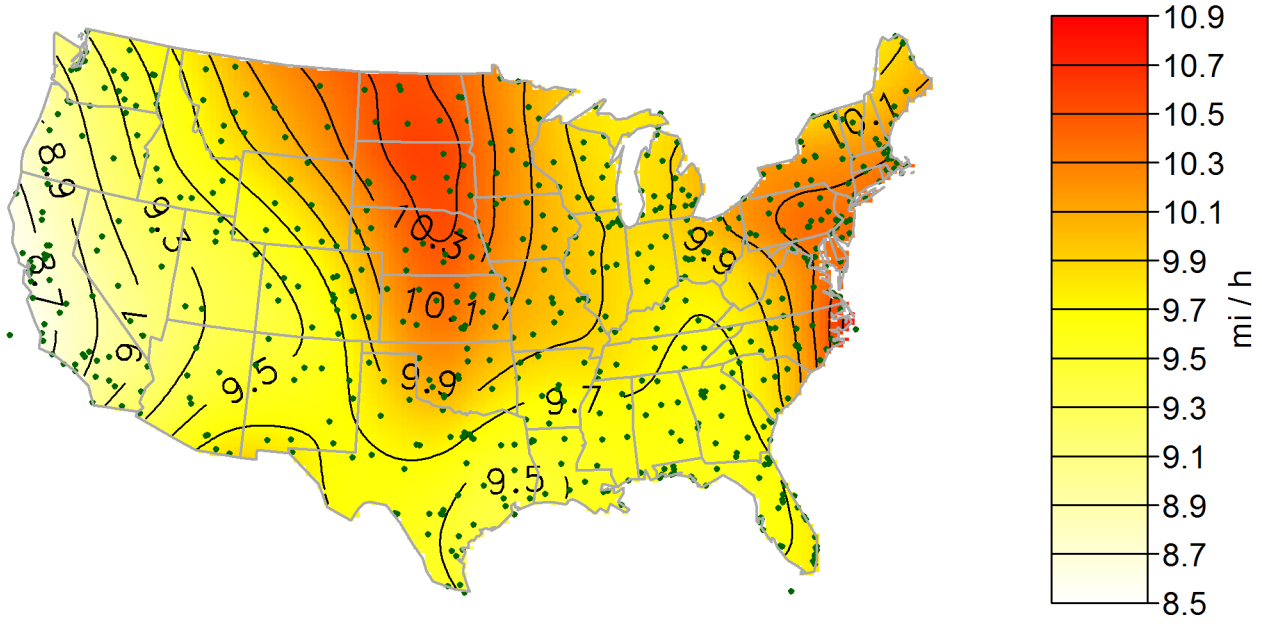


(b) Map of standard errors of estimated expected 2000 return values.

Figure 24: Gumbel maps for a 2000 year mean recurrence interval.

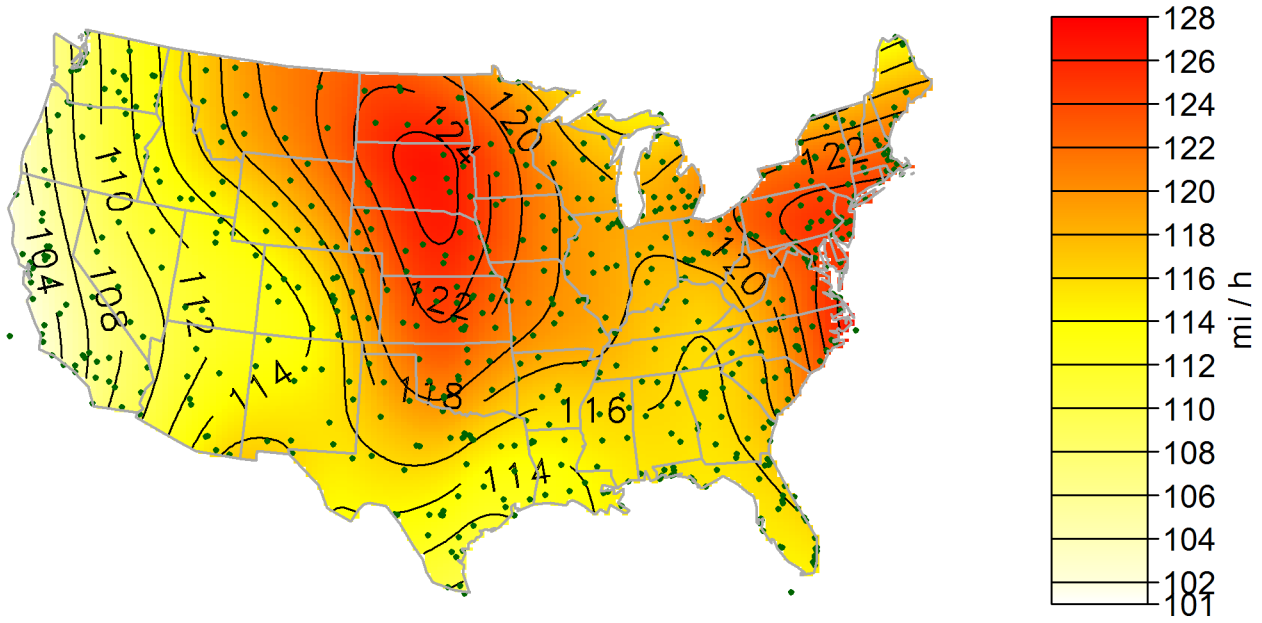


(a) Map of estimated expected 2500-year return values.

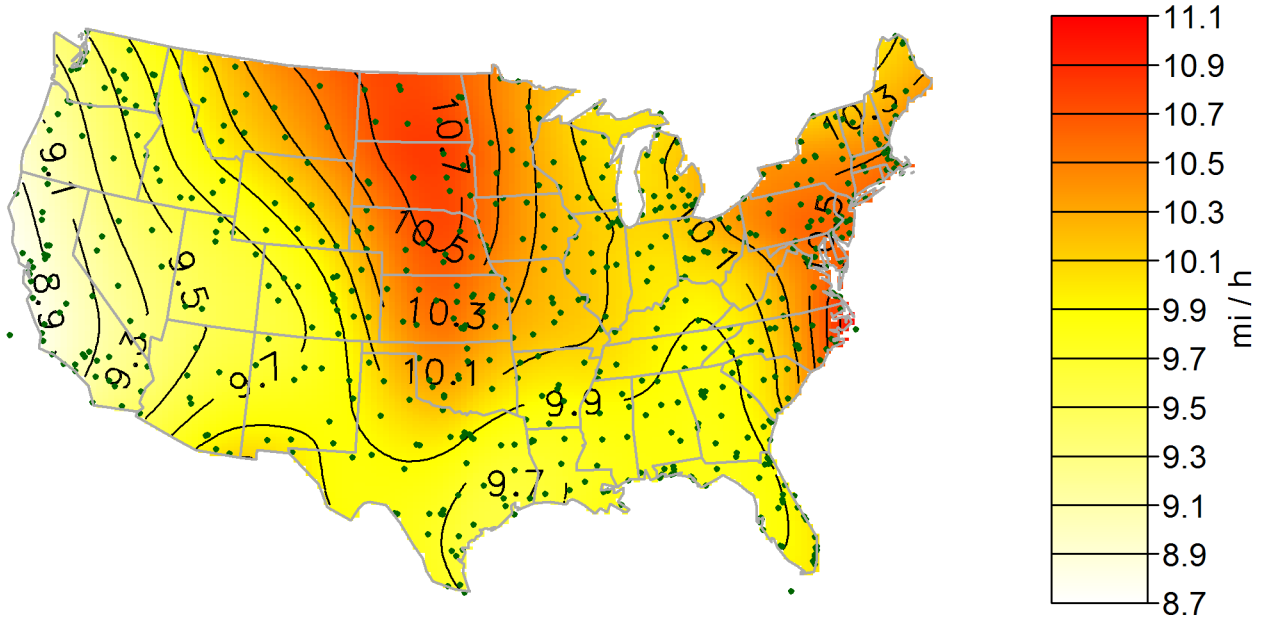


(b) Map of standard errors of estimated expected 2500-year return values.

Figure 25: Gumbel maps for a 2500 year mean recurrence interval.

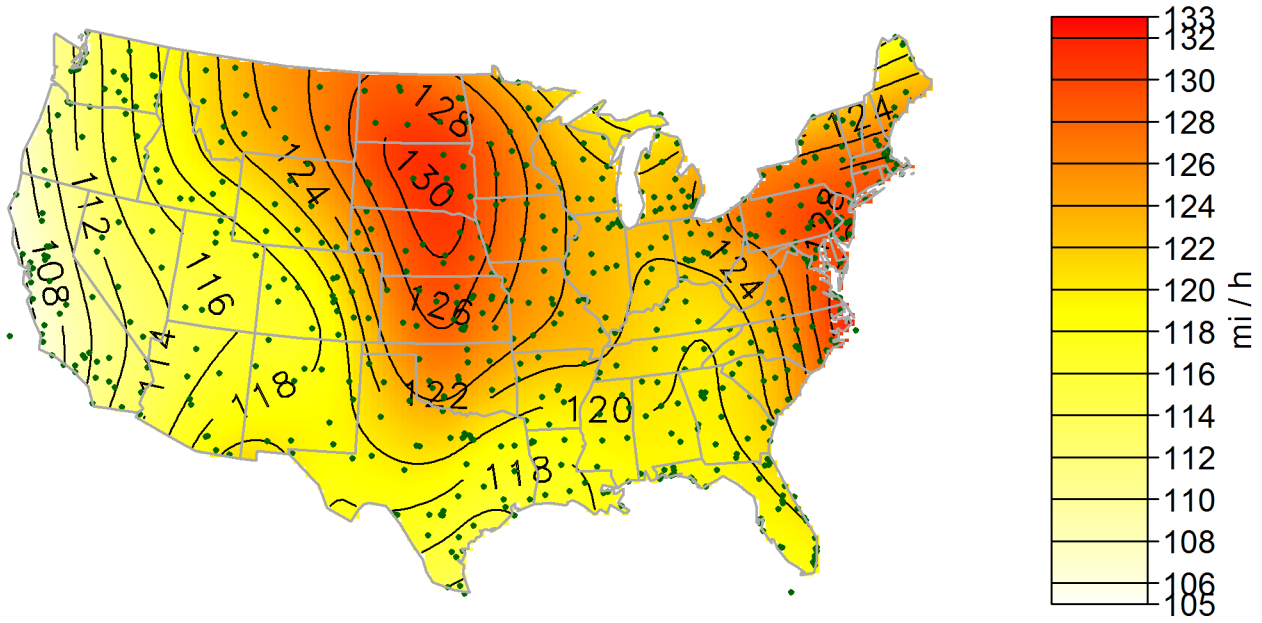


(a) Map of estimated expected 3000-year return values.

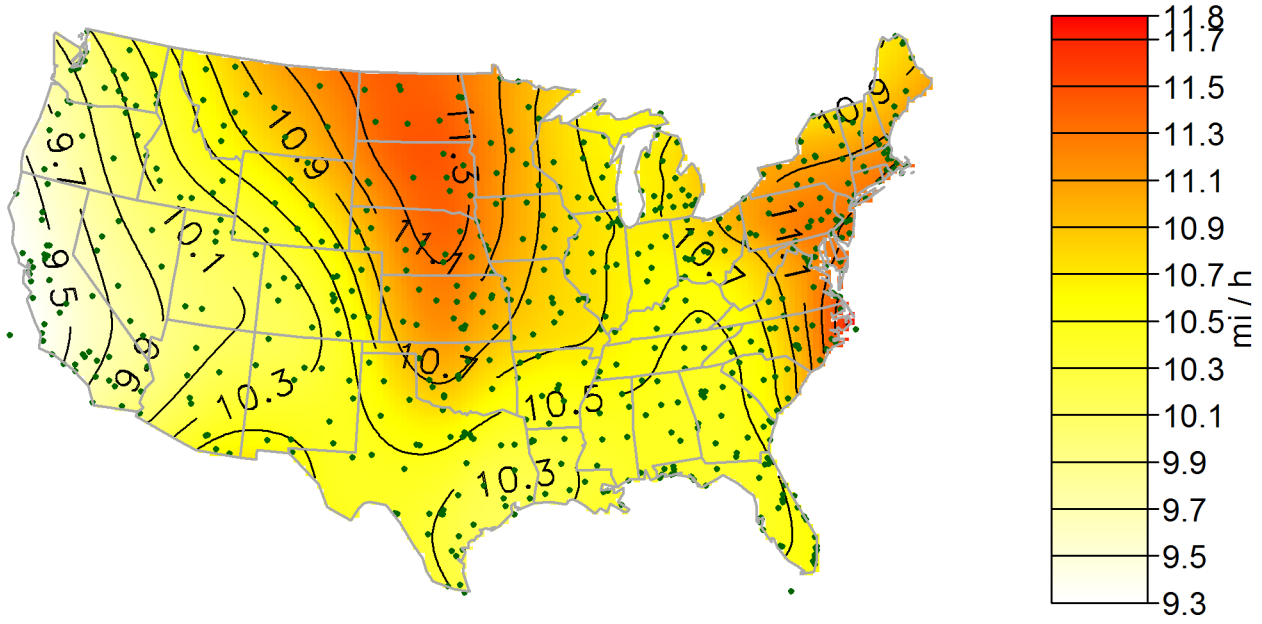


(b) Map of standard errors of estimated expected 3000-year return values.

Figure 26: Gumbel maps for a 3000 year mean recurrence interval.

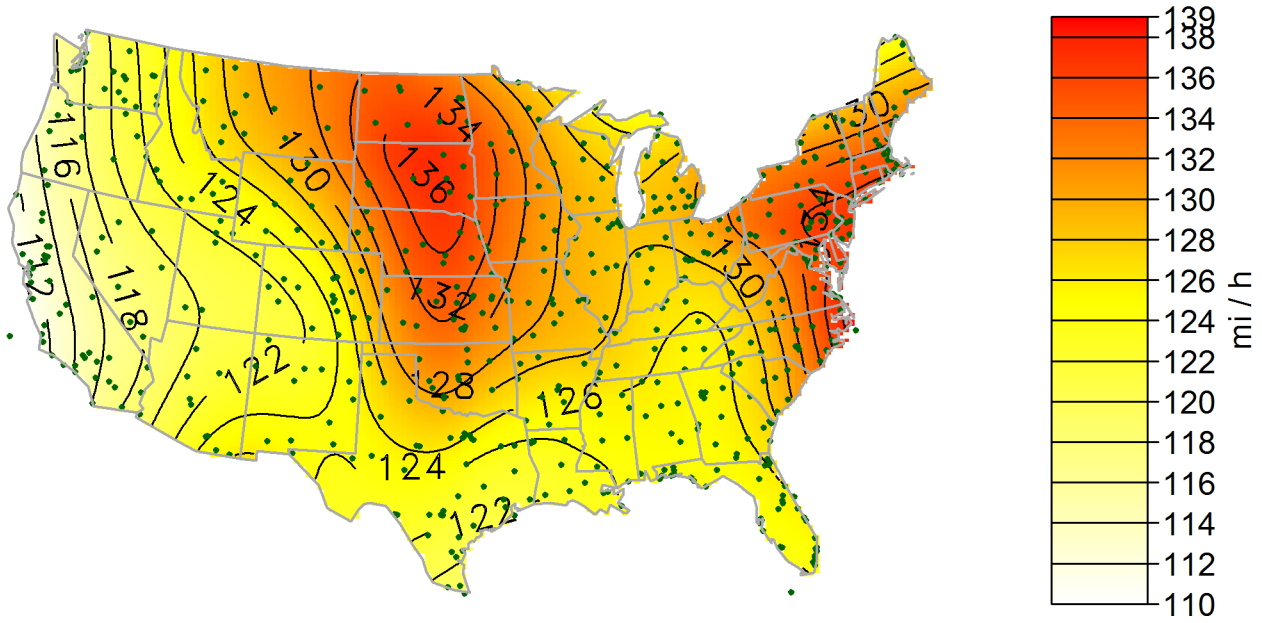


(a) Map of estimated expected 5000-year return values.

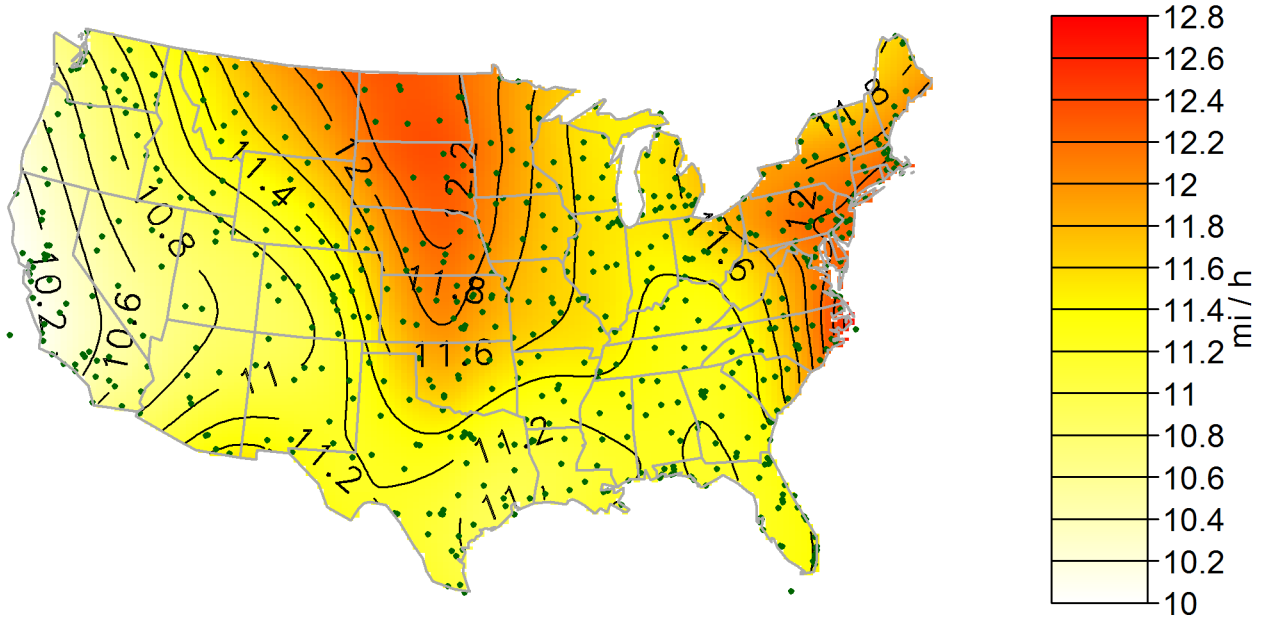


(b) Map of standard errors of estimated expected 5000-year return values.

Figure 27: Gumbel maps for a 5000 year mean recurrence interval.

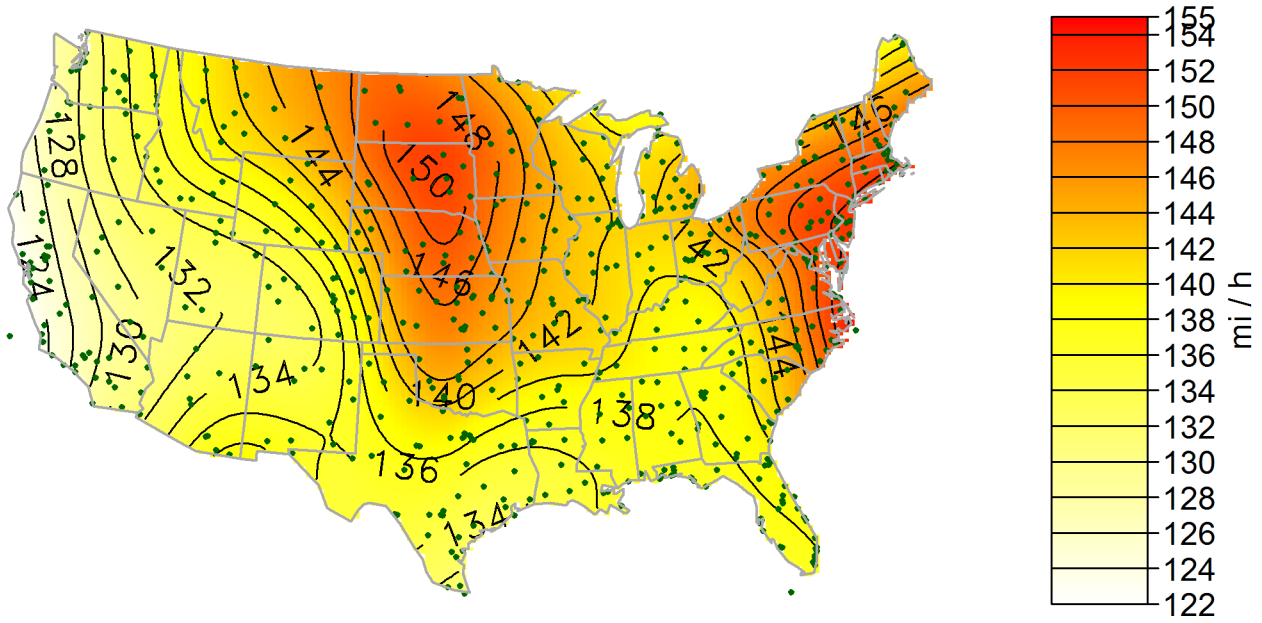


(a) Map of estimated expected 10 000-year return values.

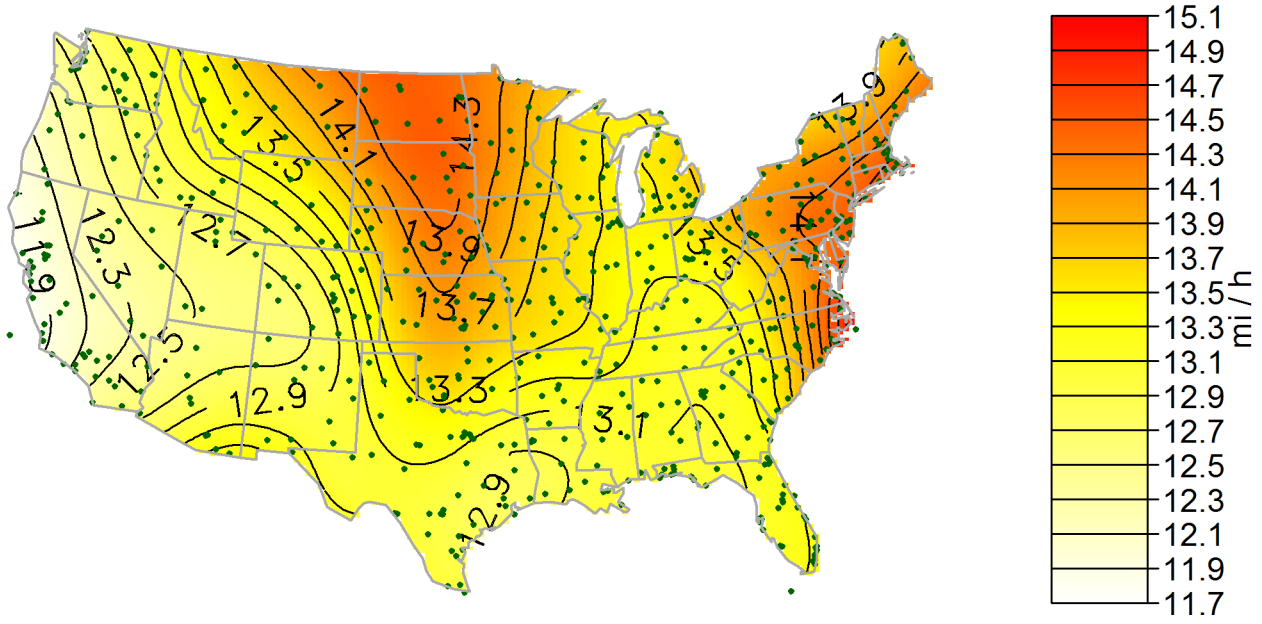


(b) Map of standard errors of estimated expected 10 000-year return values.

Figure 28: Gumbel maps for a 10 000 year mean recurrence interval.

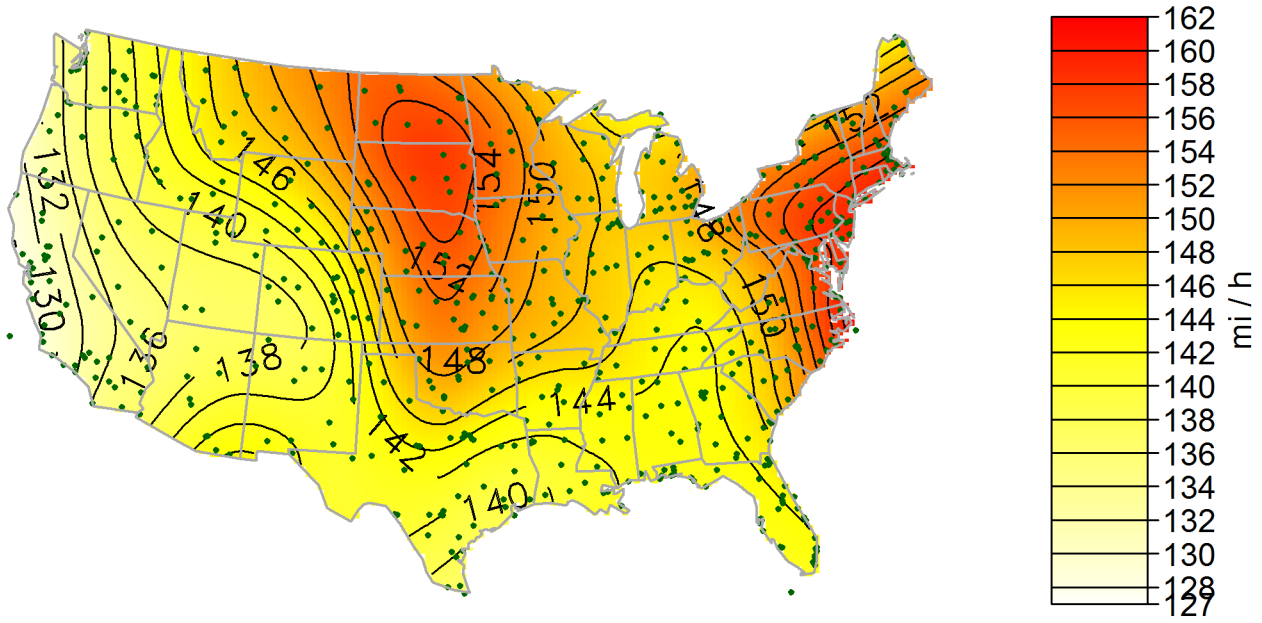


(a) Map of estimated expected 50 000-year return values.

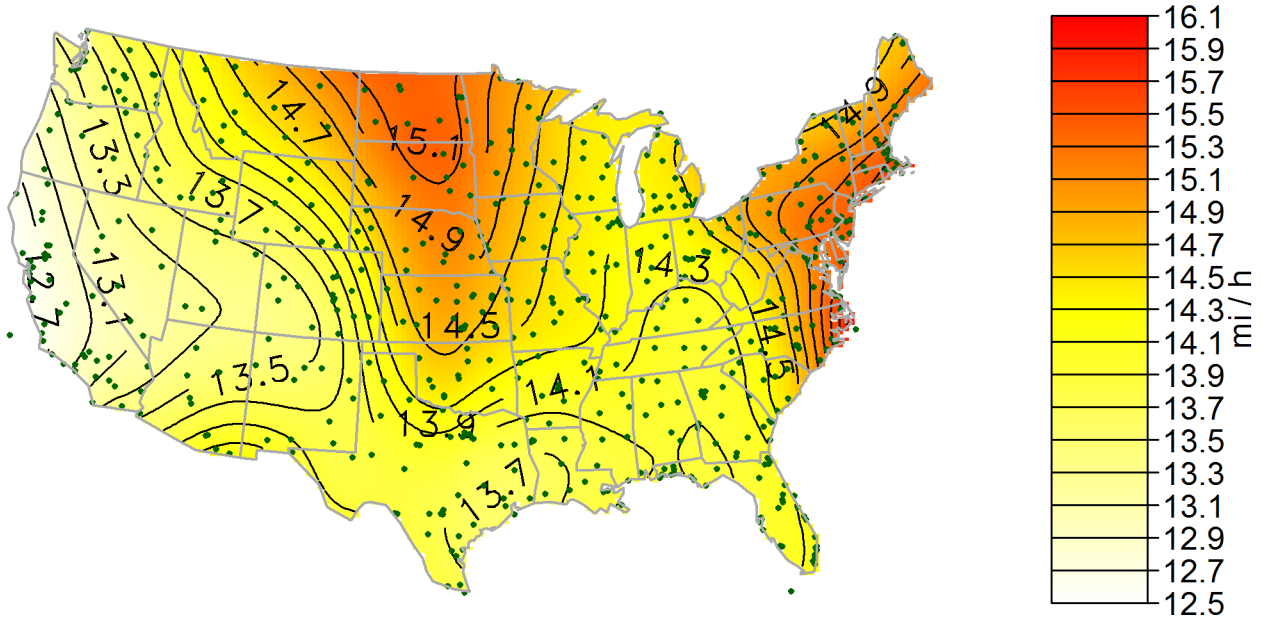


(b) Map of standard errors of estimated expected 50 000-year return values.

Figure 29: Gumbel maps for a 50 000 year mean recurrence interval.



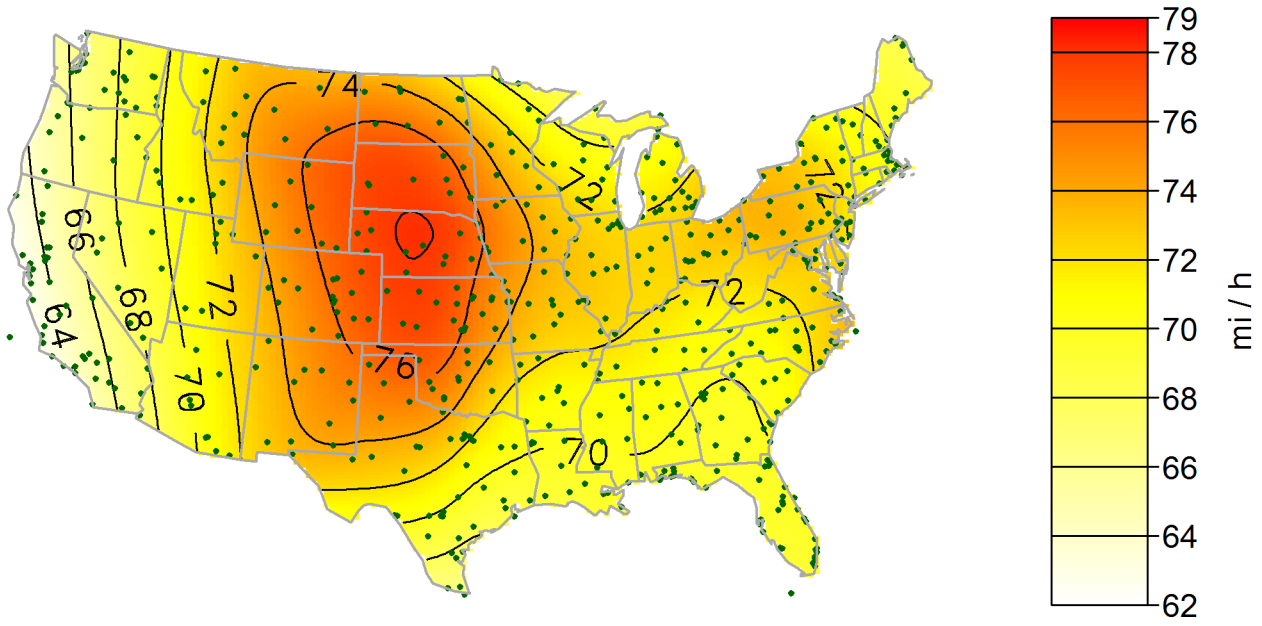
(a) Map of estimated expected 100 000-year return values.



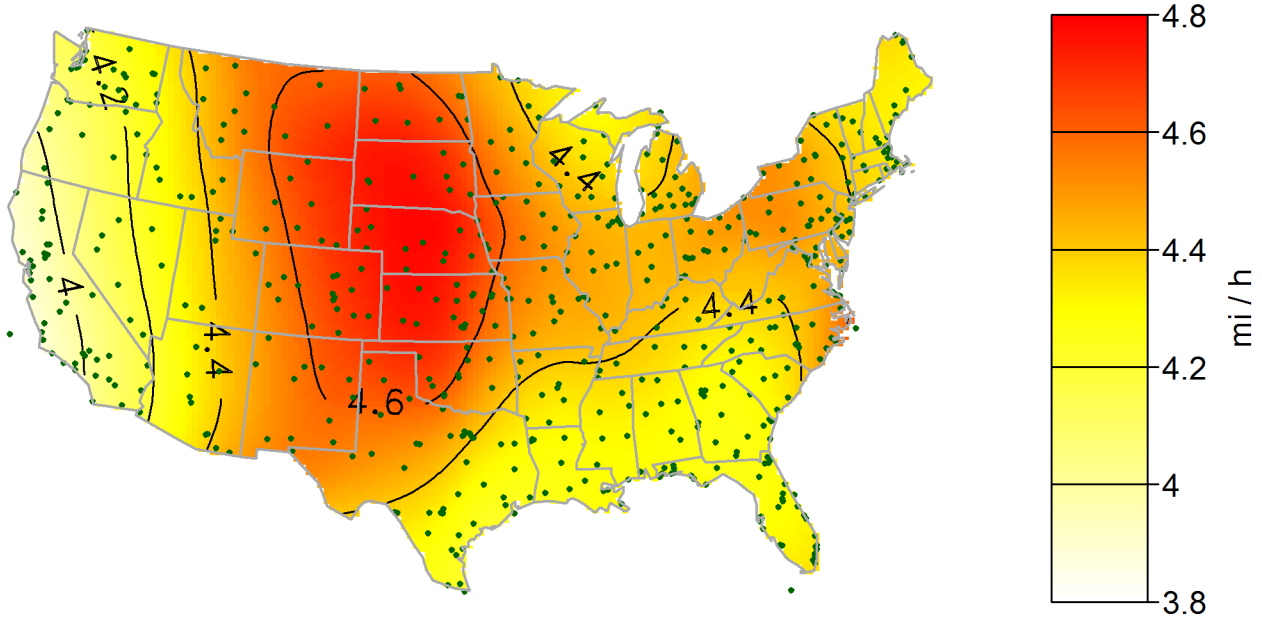
(b) Map of standard errors of estimated expected 100 000-year return values.

Figure 30: Gumbel maps for a 100 000 year mean recurrence interval.

C.2 -0.05 Tail Length

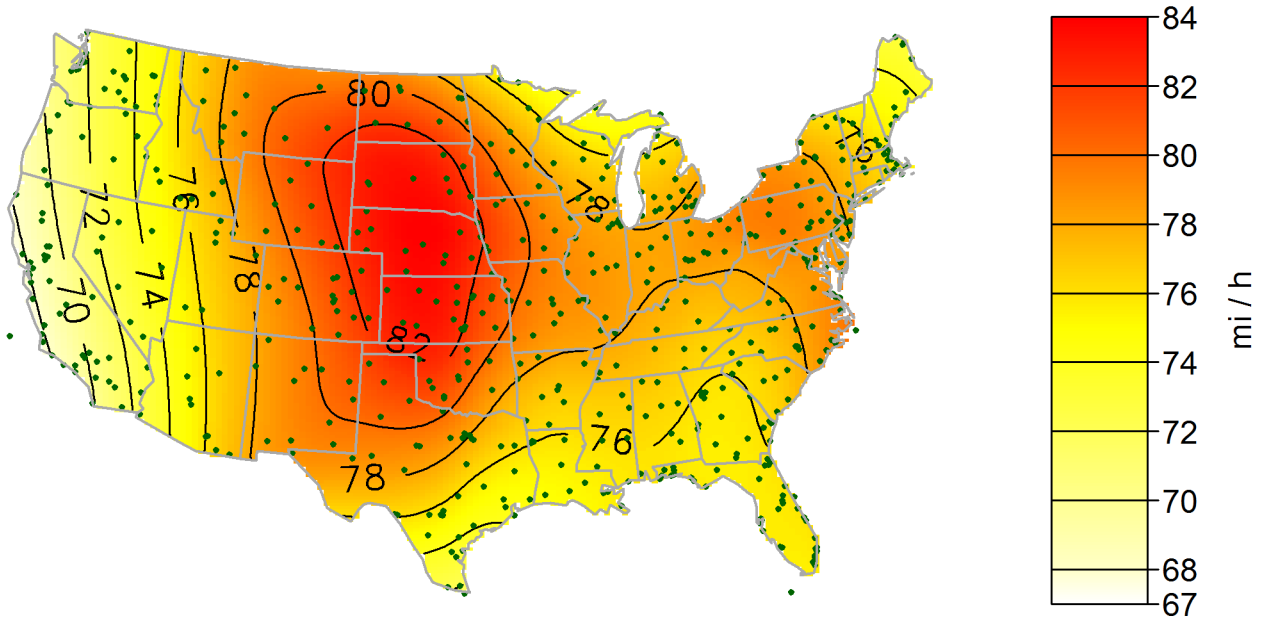


(a) Map of estimated expected 10-year return values.

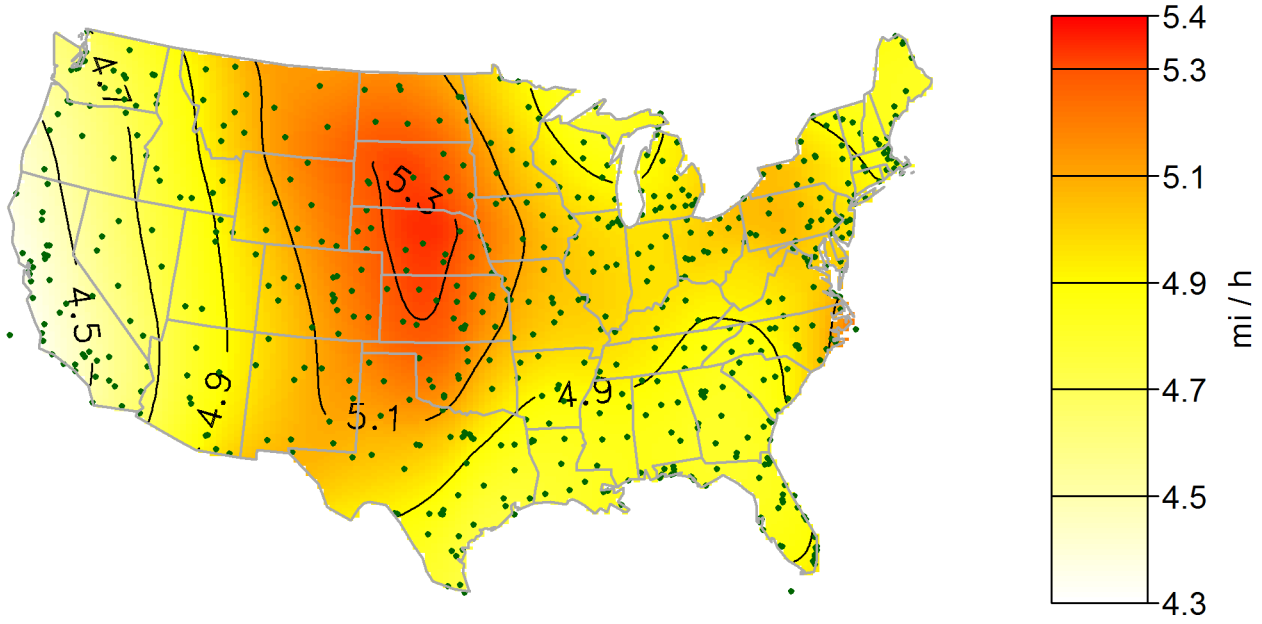


(b) Map of standard errors of estimated expected 10-year return values.

Figure 31: -0.05 tail length maps for a 10 year mean recurrence interval.

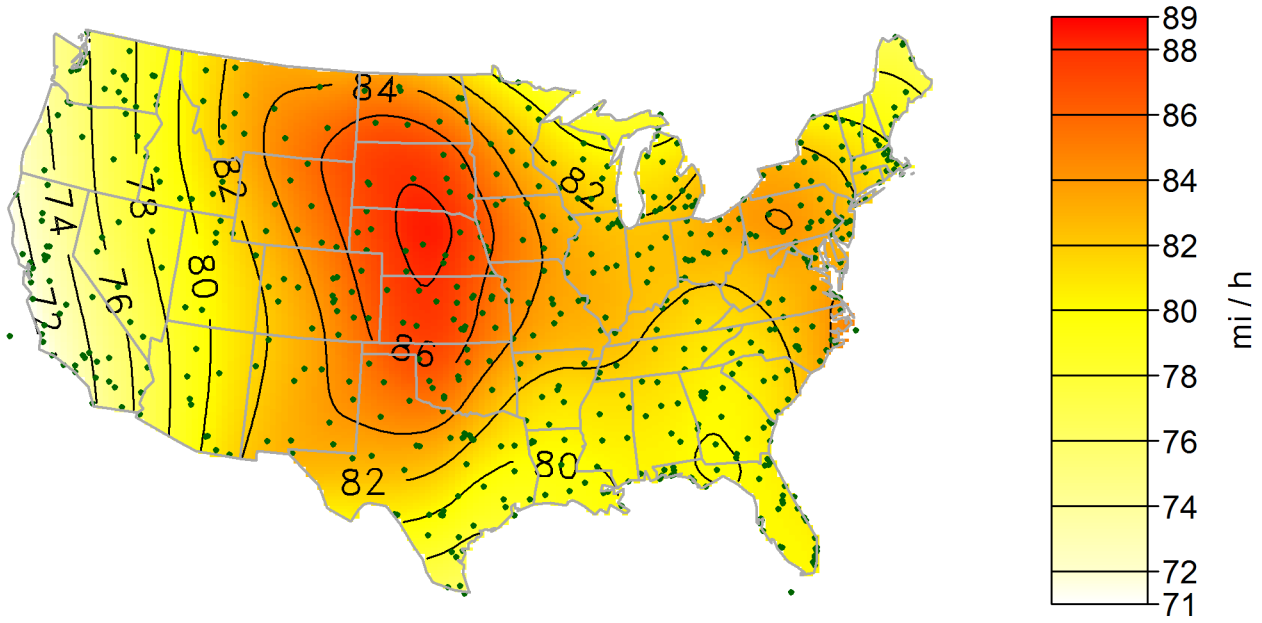


(a) Map of estimated expected 25-year return values.

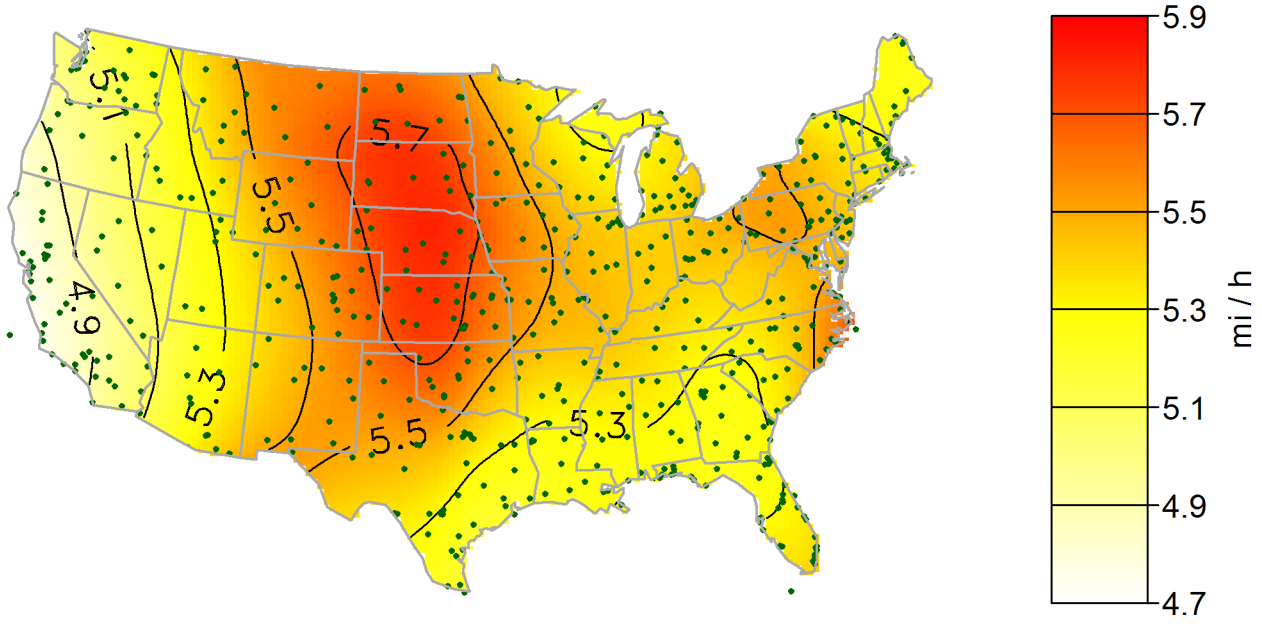


(b) Map of standard errors of estimated expected 25-year return values.

Figure 32: -0.05 tail length maps for a 25 year mean recurrence interval.

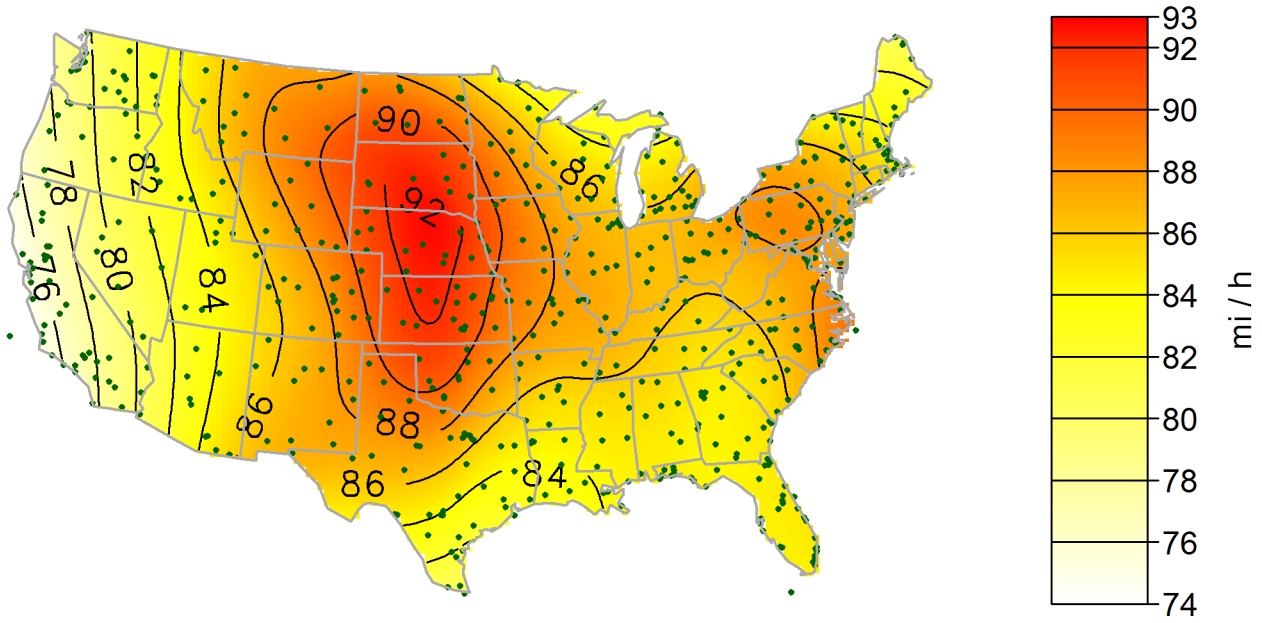


(a) Map of estimated expected 50-year return values.

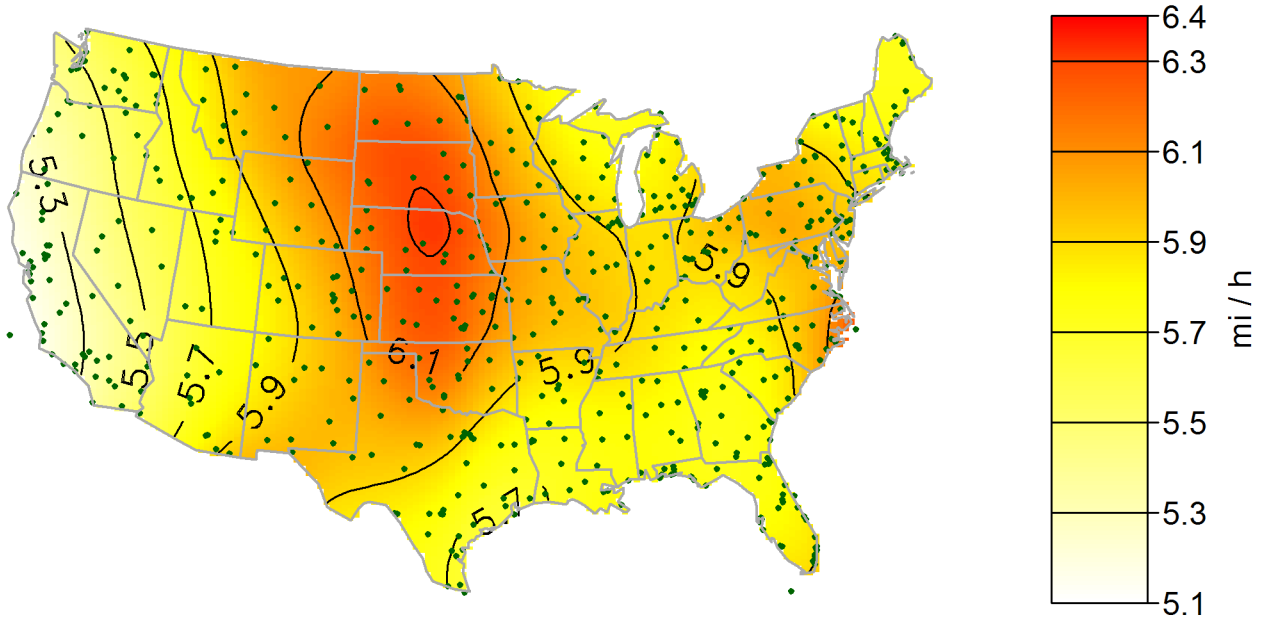


(b) Map of standard errors of estimated expected 50-year return values.

Figure 33: -0.05 tail length maps for a 50 year mean recurrence interval.

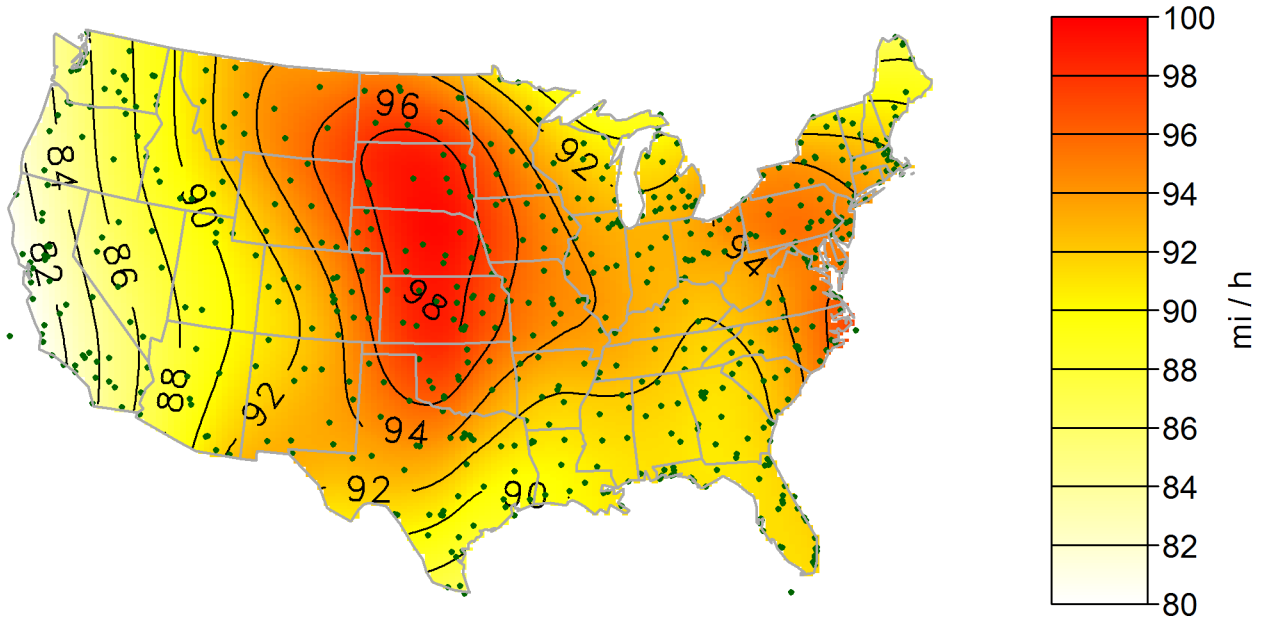


(a) Map of estimated expected 100-year return values.

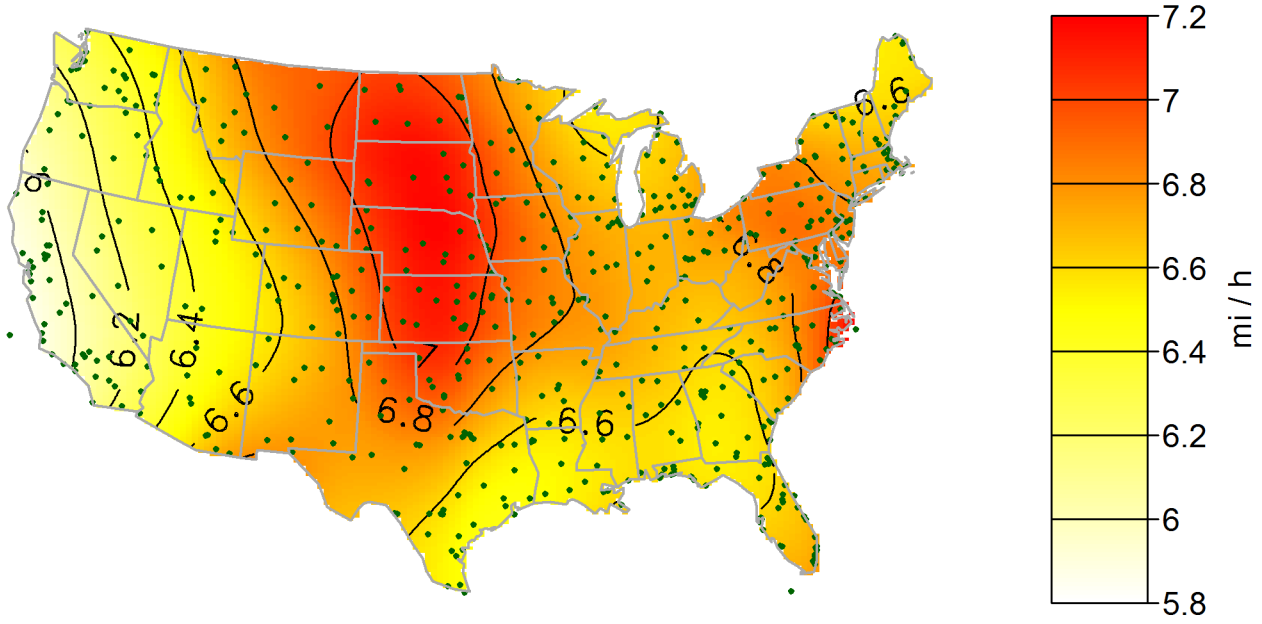


(b) Map of standard errors of estimated expected 100-year return values.

Figure 34: -0.05 tail length maps for a 100 year mean recurrence interval.

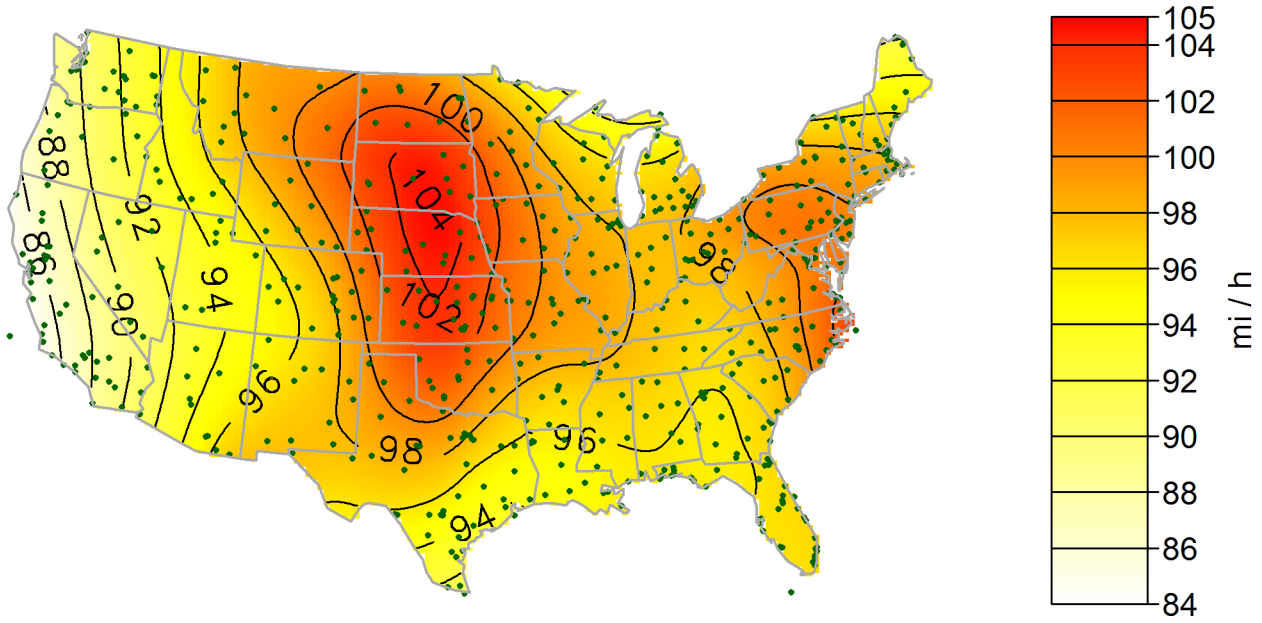


(a) Map of estimated expected 300-year return values.

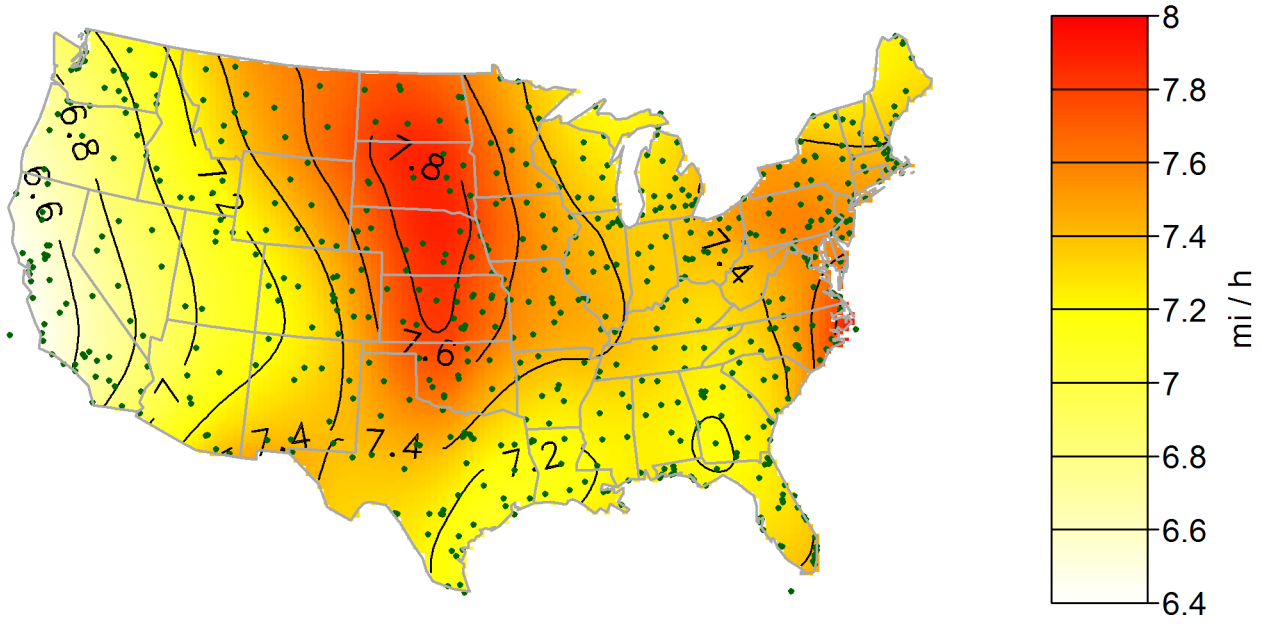


(b) Map of standard errors of estimated expected 300-year return values.

Figure 35: -0.05 tail length maps for a 300 year mean recurrence interval.

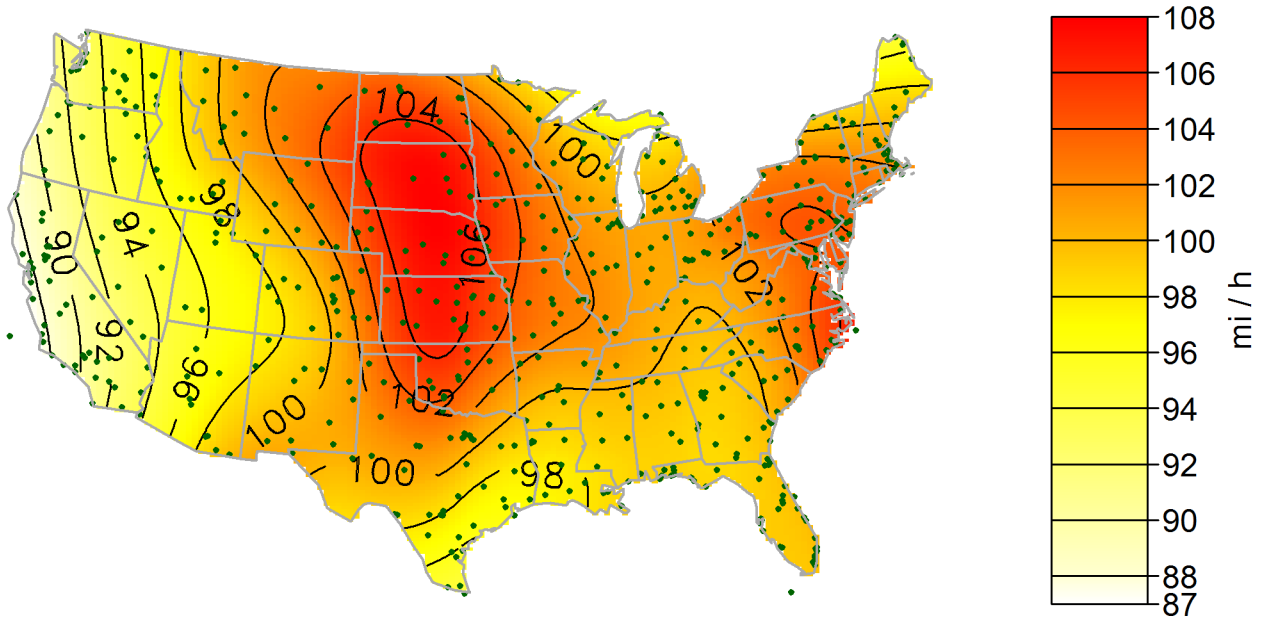


(a) Map of estimated expected 700-year return values.

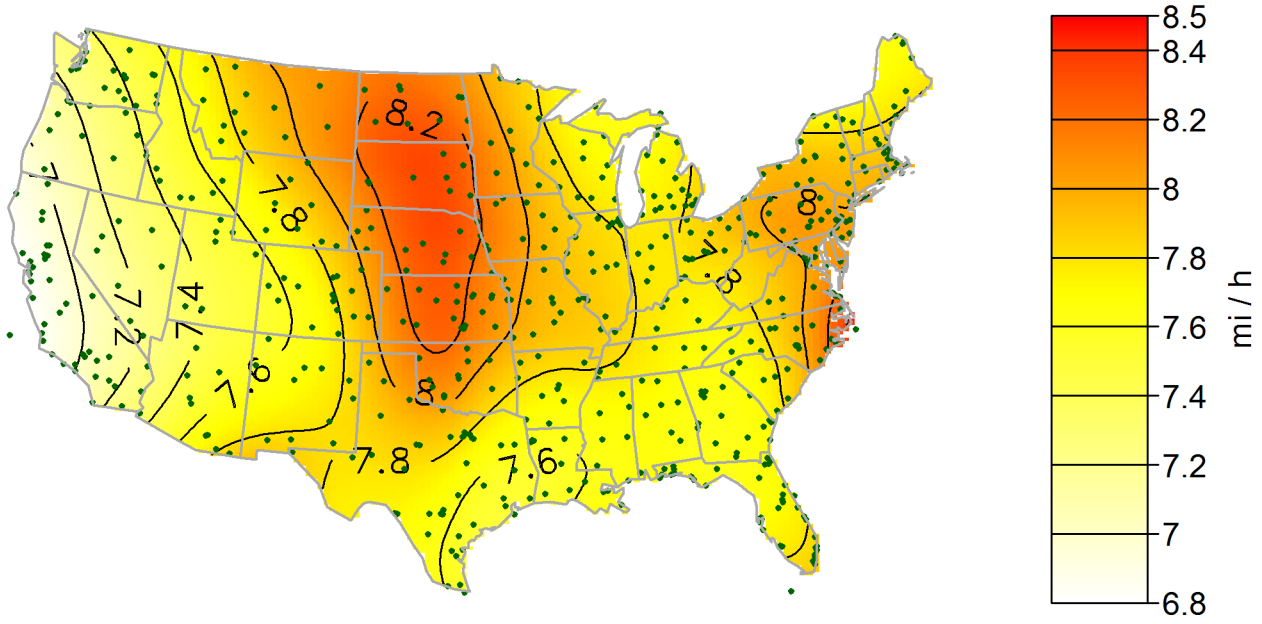


(b) Map of standard errors of estimated expected 700-year return values.

Figure 36: -0.05 tail length maps for a 700 year mean recurrence interval.

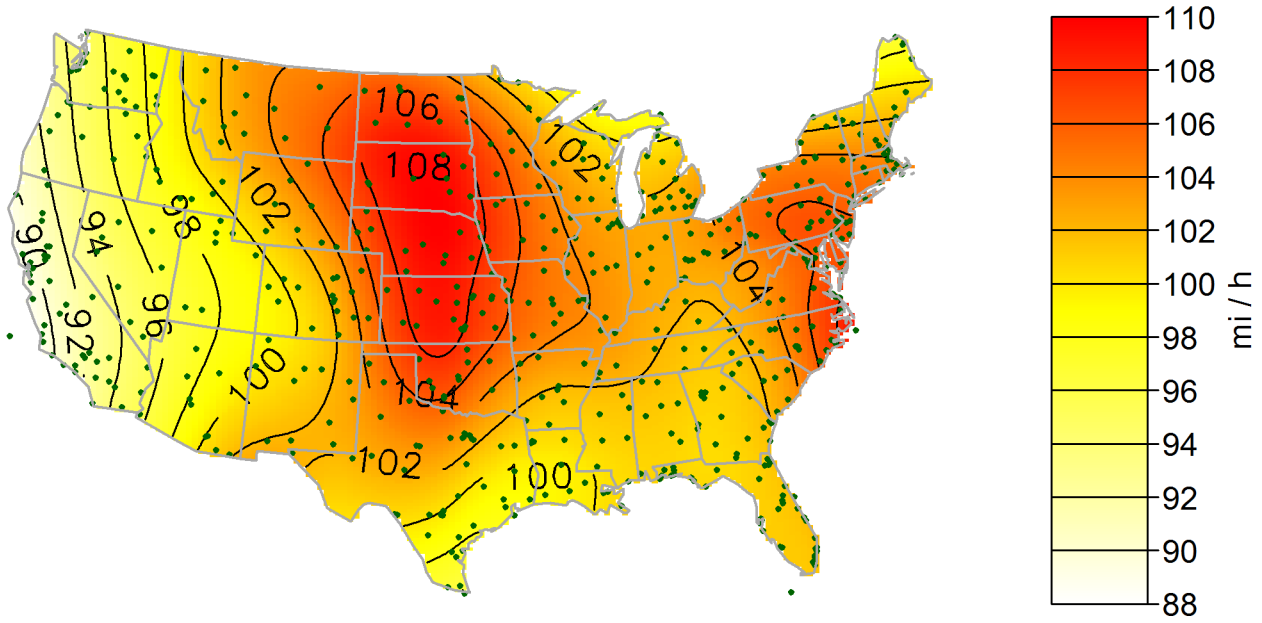


(a) Map of estimated expected 1200-year return values.

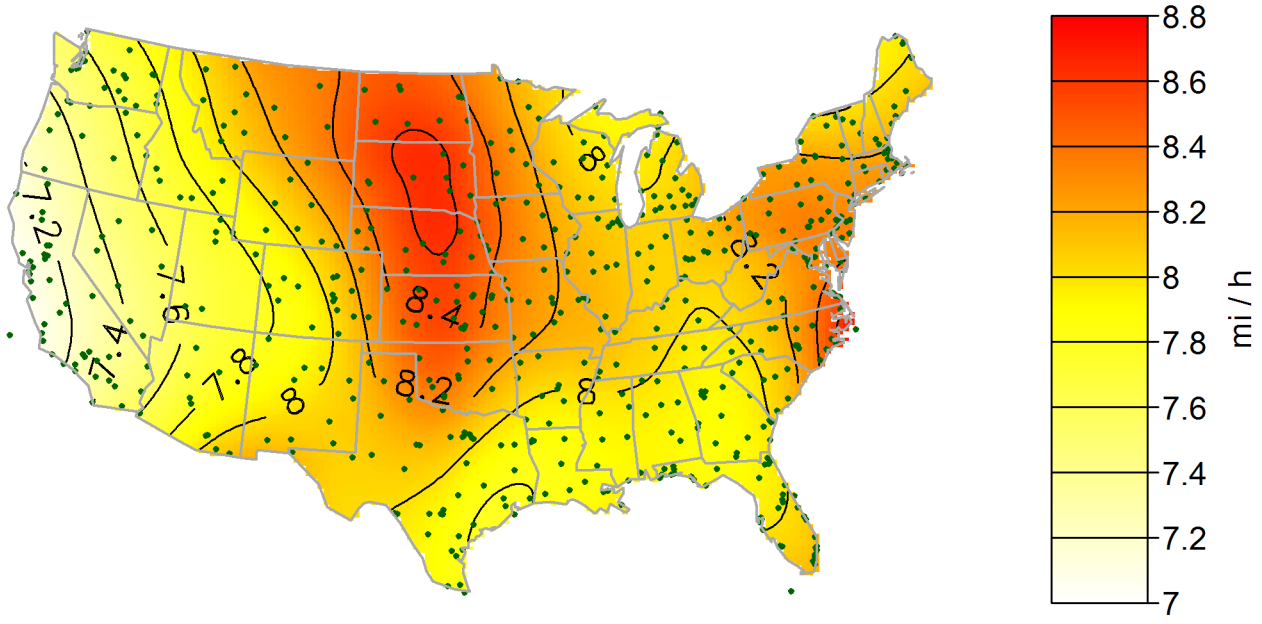


(b) Map of standard errors of estimated expected 1200-year return values.

Figure 37: -0.05 tail length maps for a 1200 year mean recurrence interval.

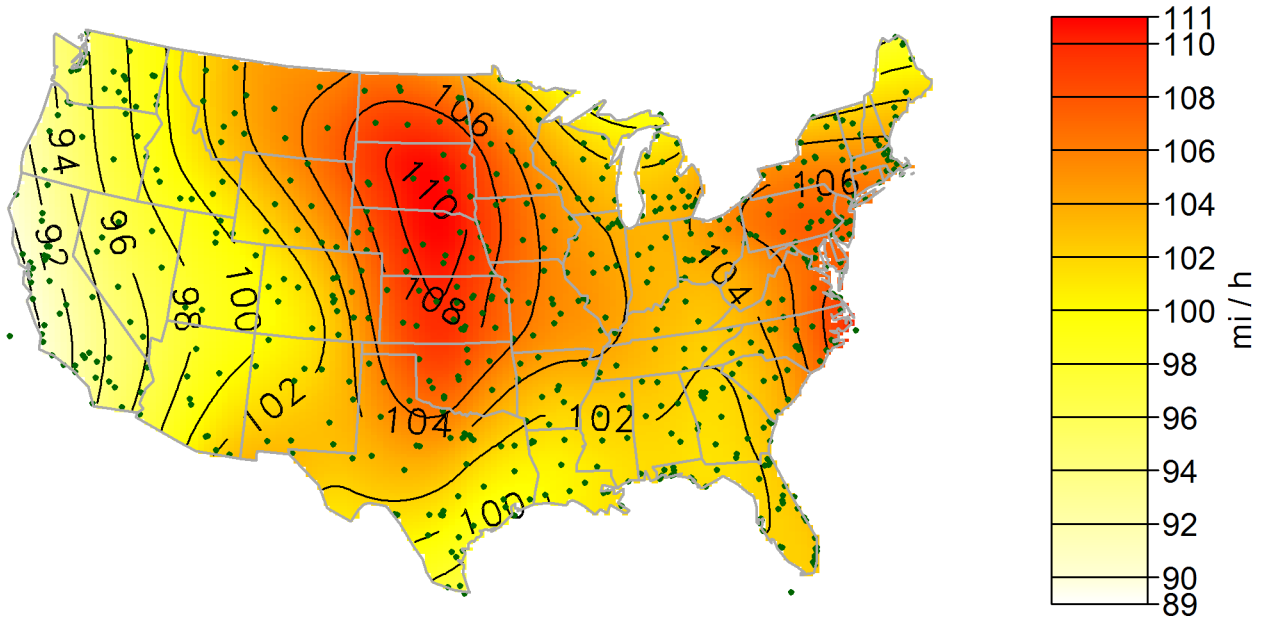


(a) Map of estimated expected 1700-year return values.

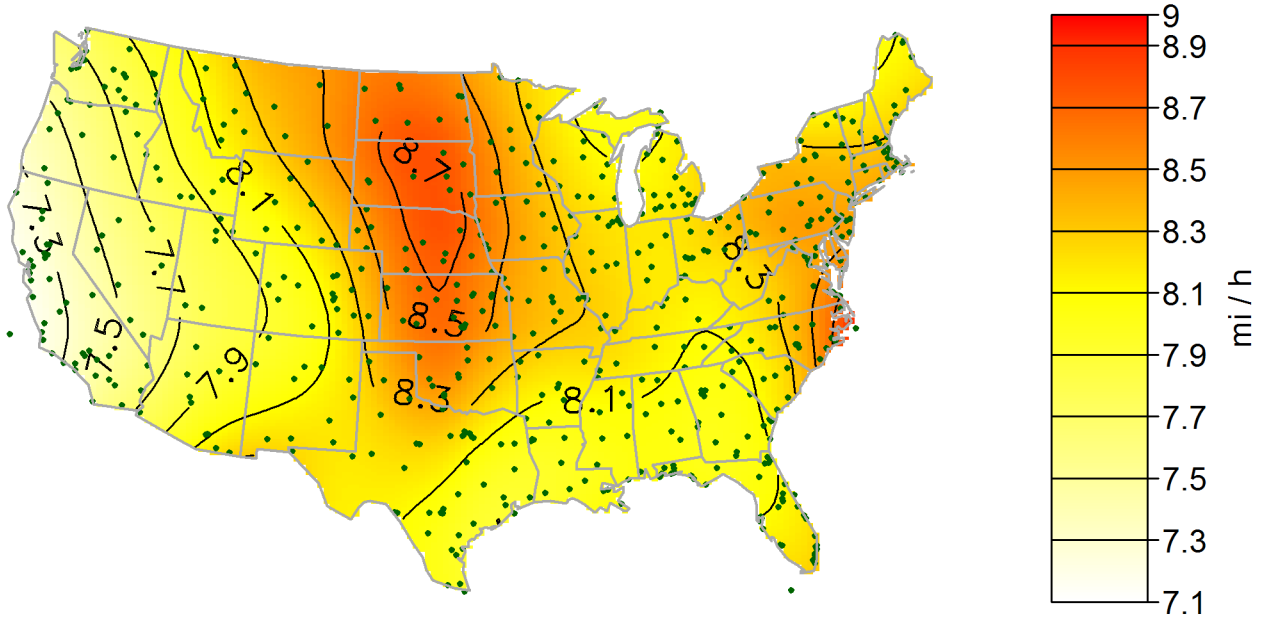


(b) Map of standard errors of estimated expected 1700-year return values.

Figure 38: -0.05 tail length maps for a 1700 year mean recurrence interval.

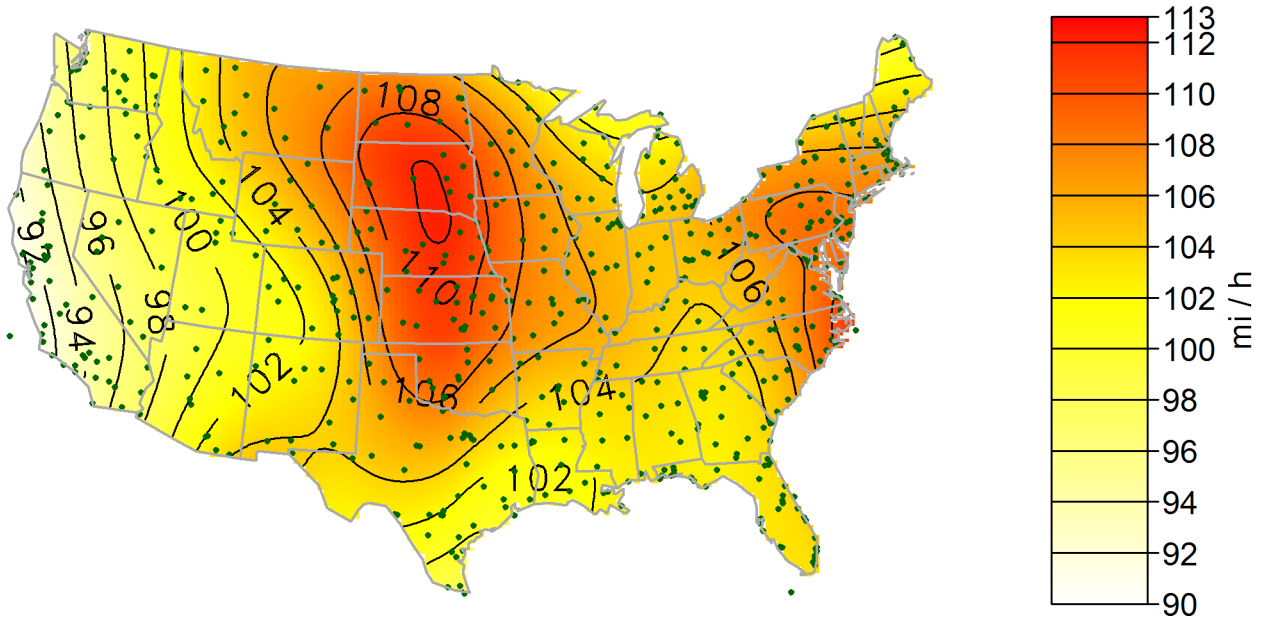


(a) Map of estimated expected 2000-year return values.

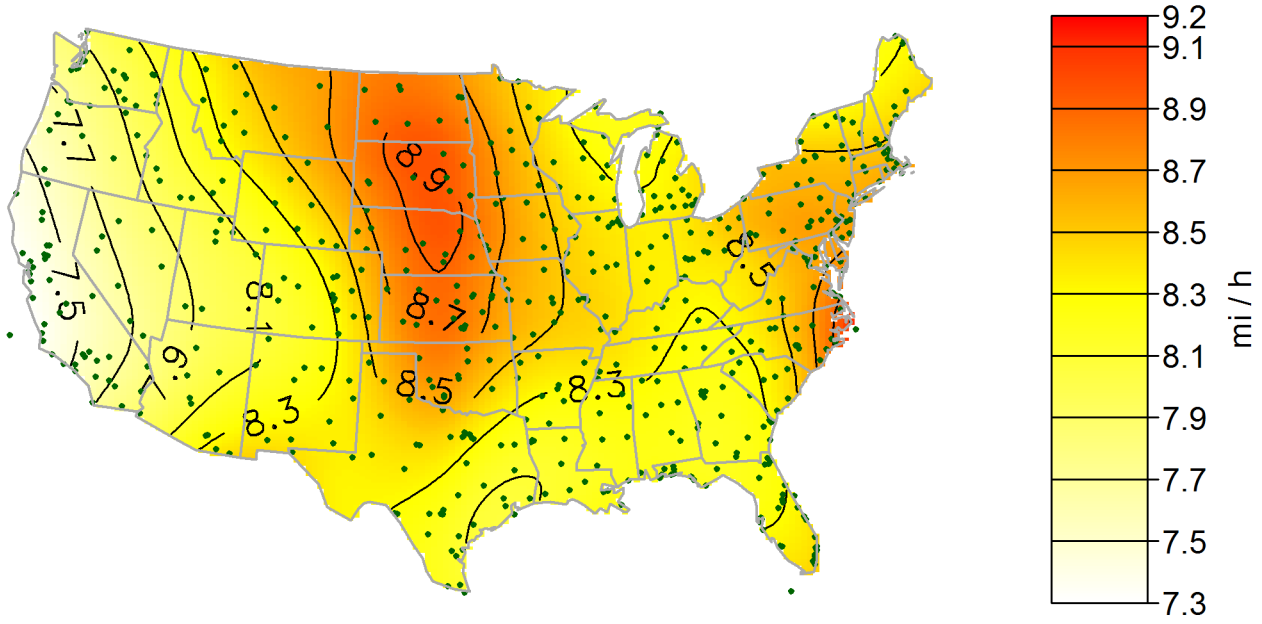


(b) Map of standard errors of estimated expected 2000-year return values.

Figure 39: -0.05 tail length maps for a 2000 year mean recurrence interval.

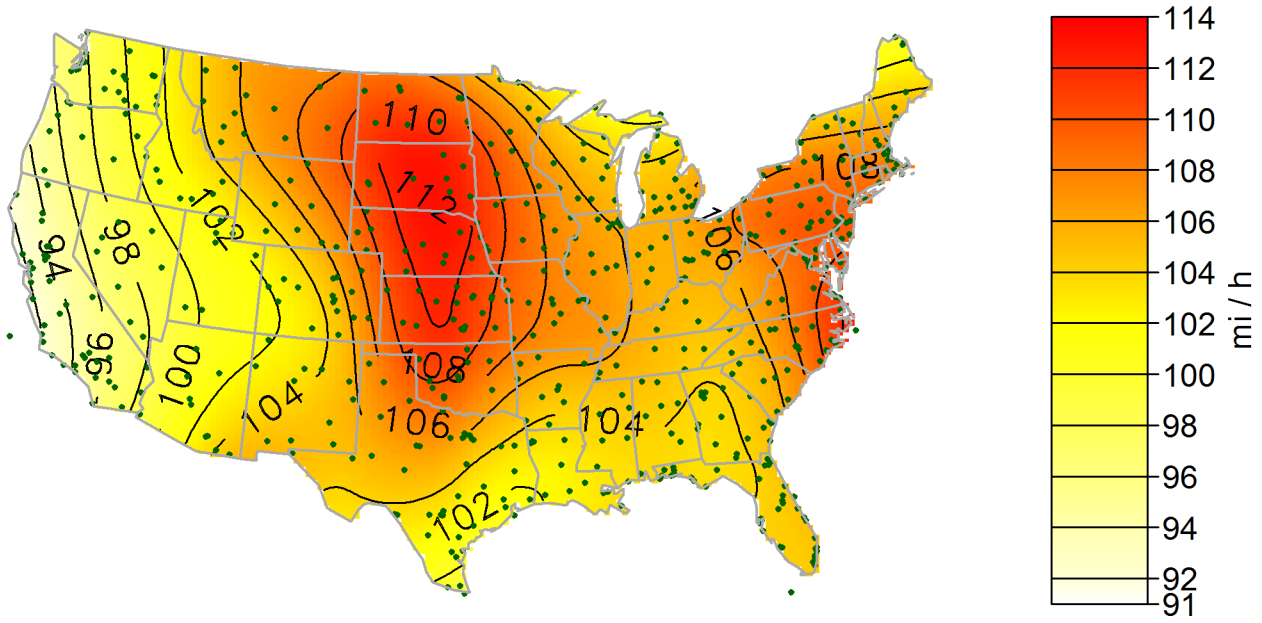


(a) Map of estimated expected 2500-year return values.

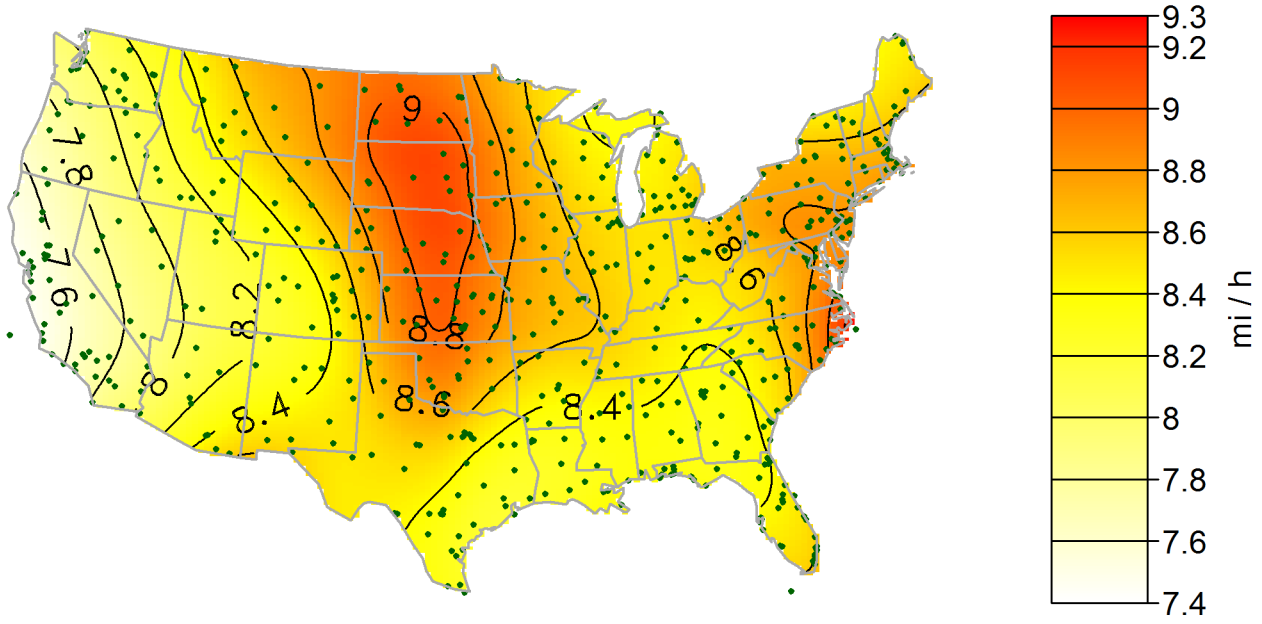


(b) Map of standard errors of estimated expected 2500-year return values.

Figure 40: -0.05 tail length maps for a 2500 year mean recurrence interval.

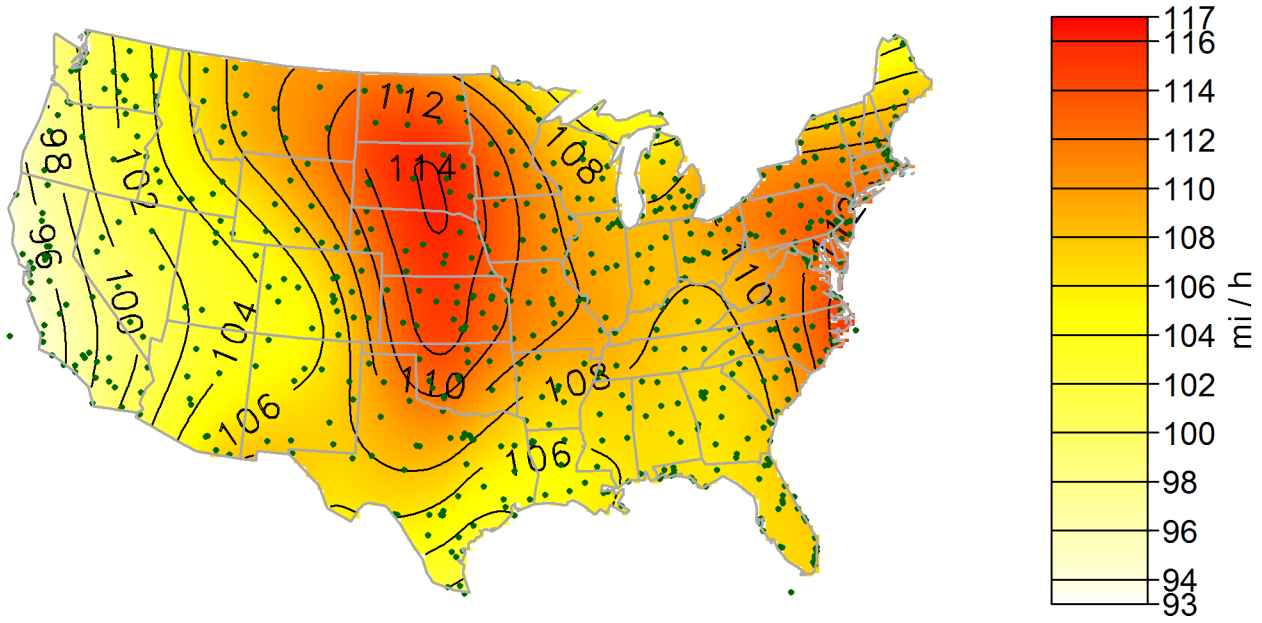


(a) Map of estimated expected 3000-year return values.

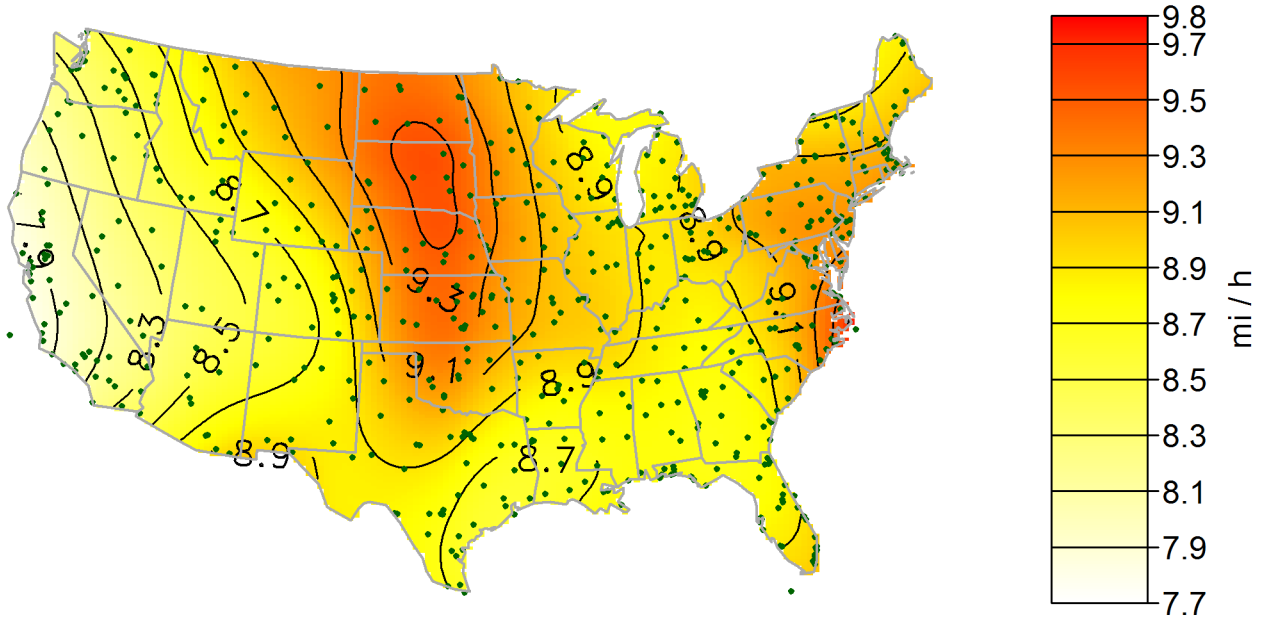


(b) Map of standard errors of estimated expected 3000-year return values.

Figure 41: -0.05 tail length maps for a 3000 year mean recurrence interval.

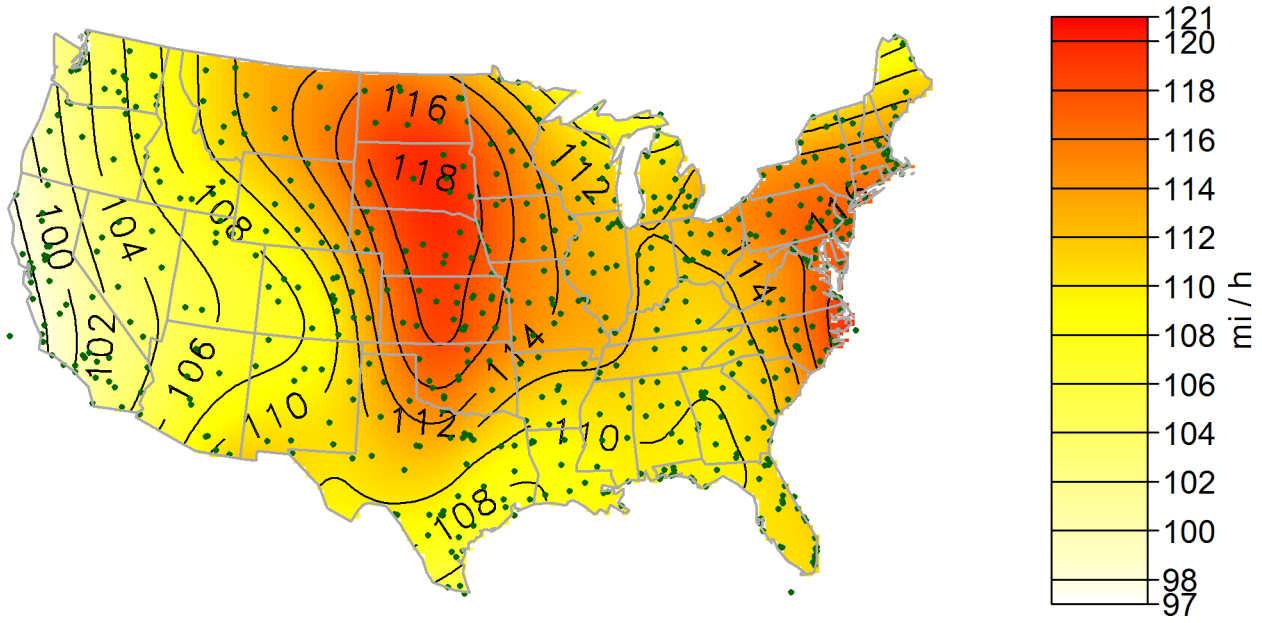


(a) Map of estimated expected 5000-year return values.

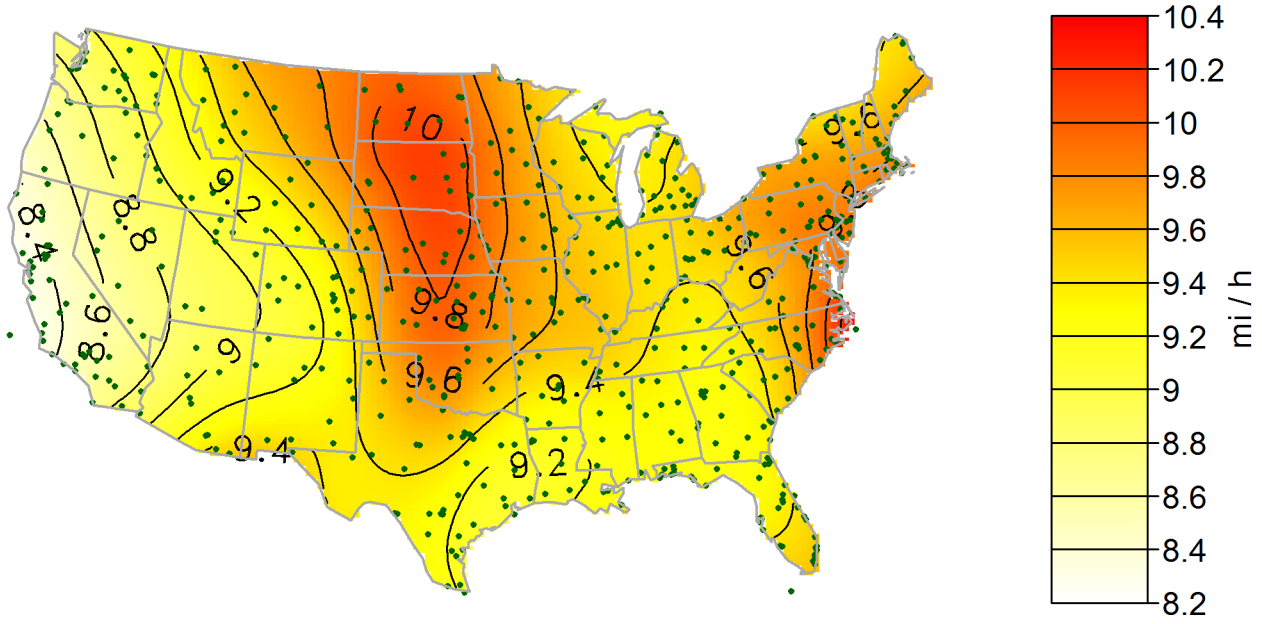


(b) Map of standard errors of estimated expected 5000-year return values.

Figure 42: -0.05 tail length maps for a 5000 year mean recurrence interval.

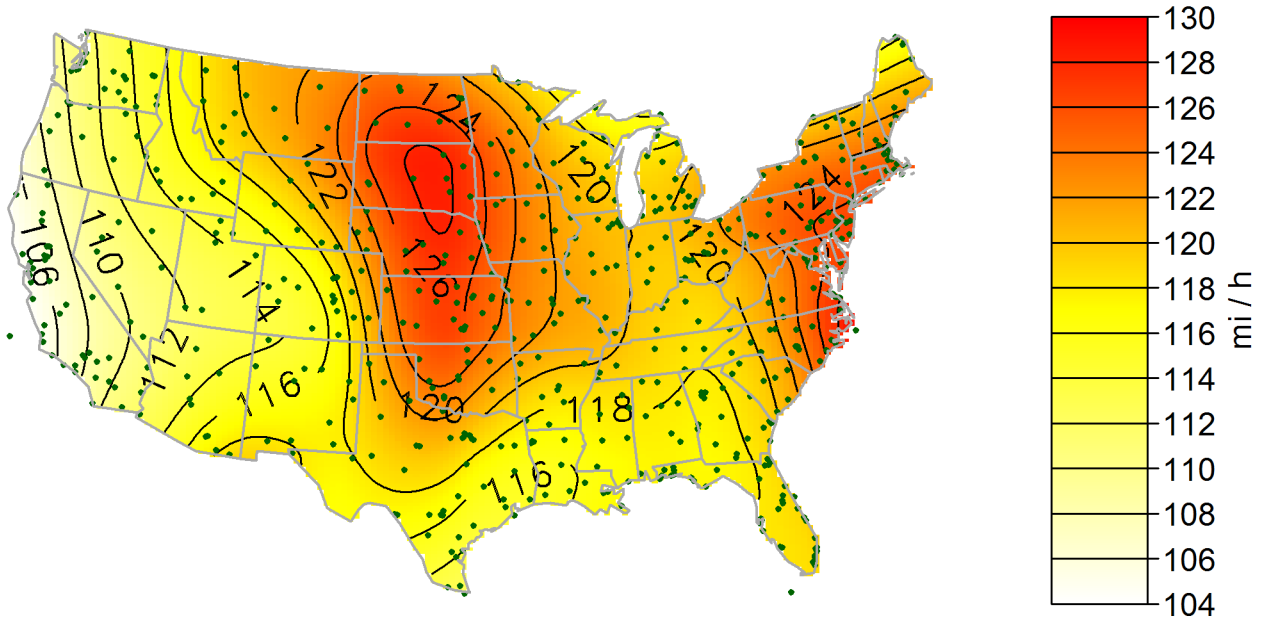


(a) Map of estimated expected 10 000-year return values.

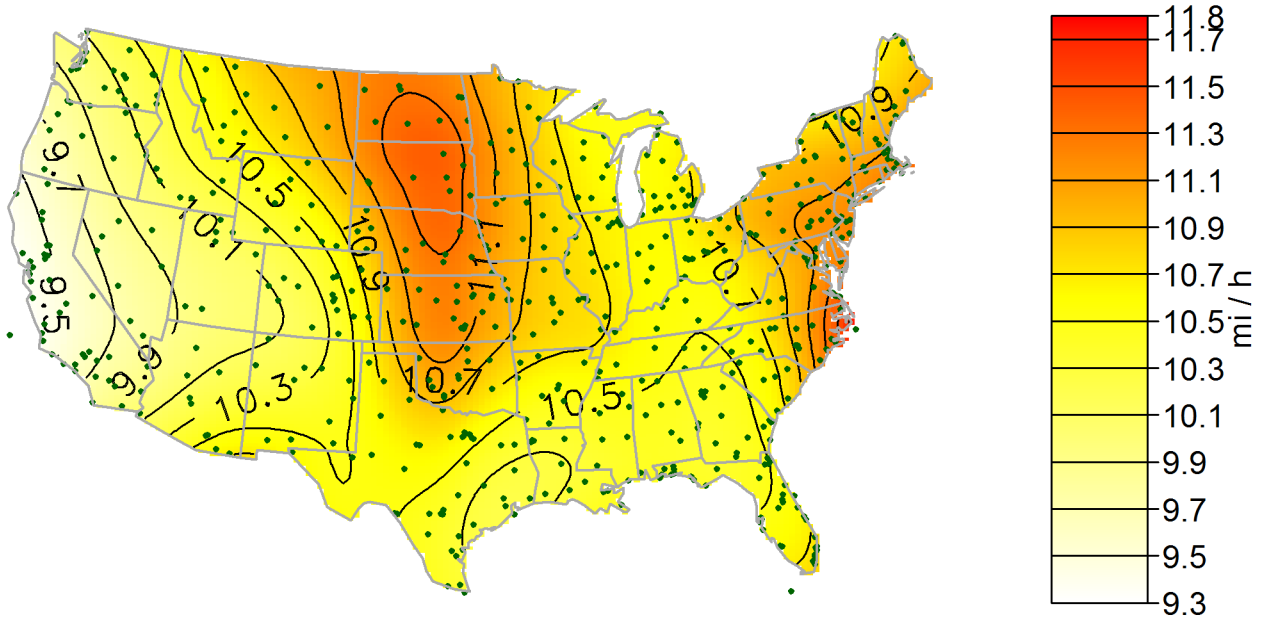


(b) Map of standard errors of estimated expected 10 000-year return values.

Figure 43: -0.05 tail length maps for a 10 000 year mean recurrence interval.

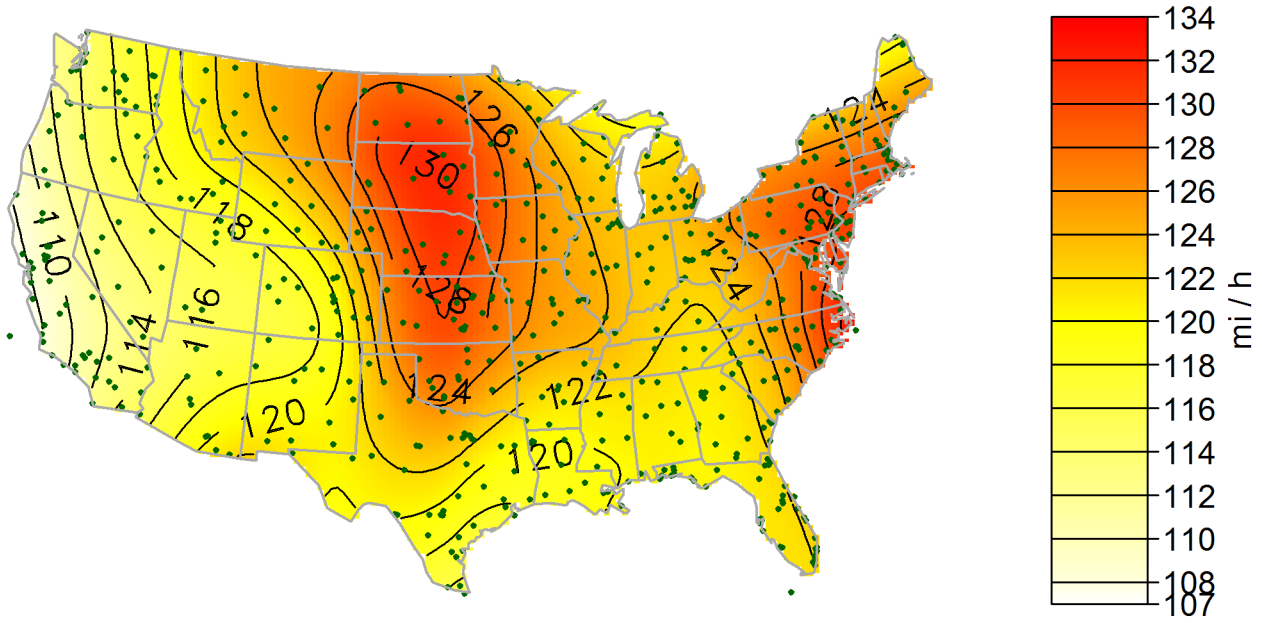


(a) Map of estimated expected 50 000-year return values.

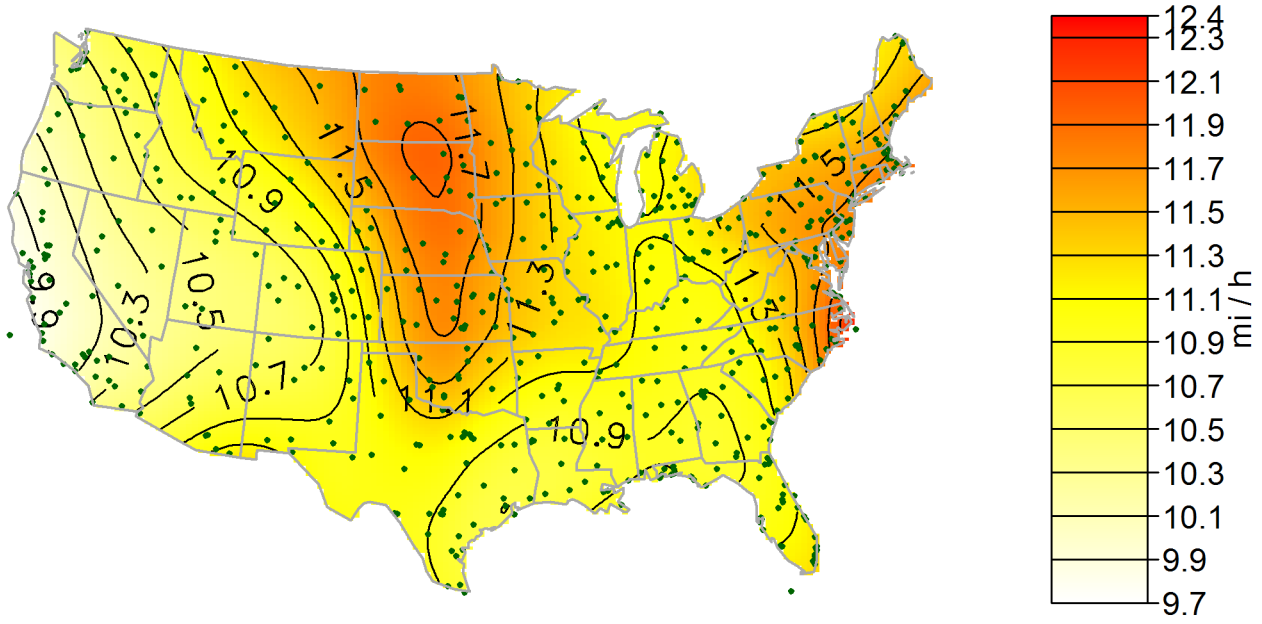


(b) Map of standard errors of estimated expected 50 000-year return values.

Figure 44: -0.05 tail length maps for a 50 000 year mean recurrence interval.



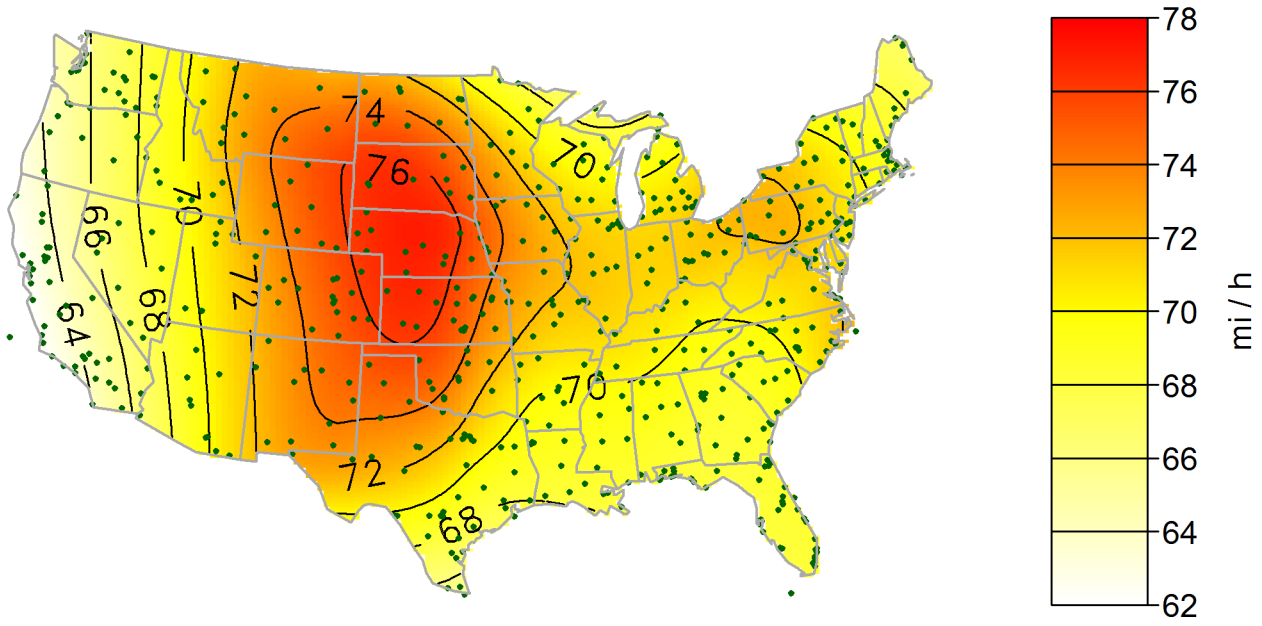
(a) Map of estimated expected 100 000-year return values.



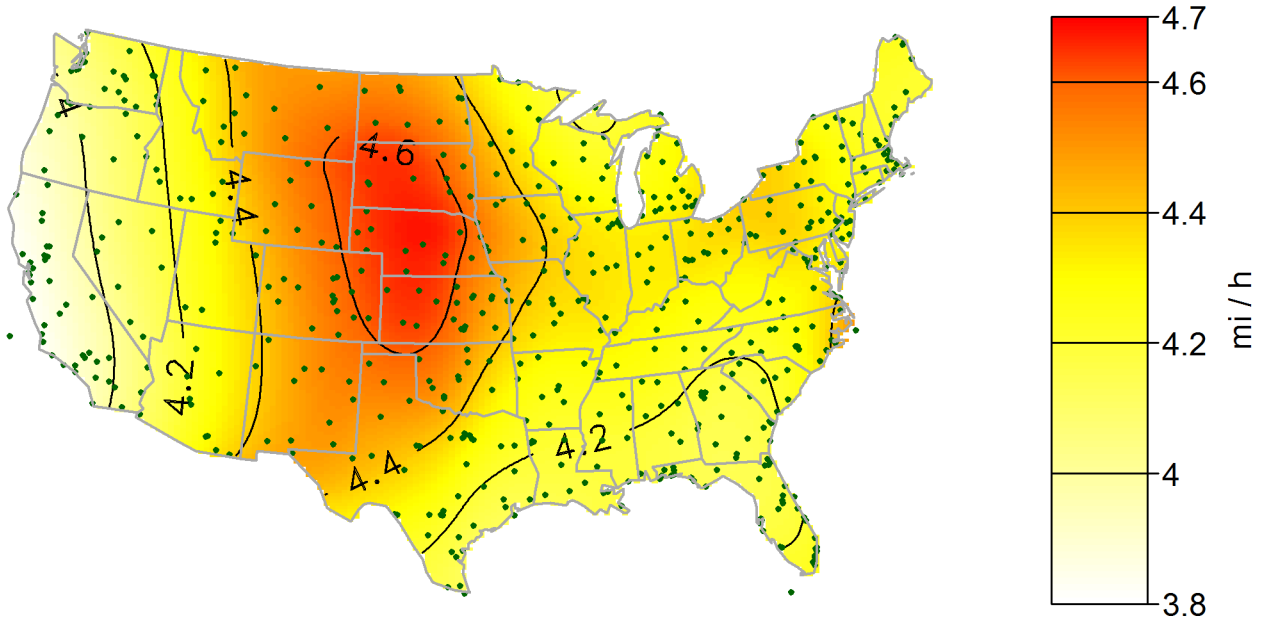
(b) Map of standard errors of estimated expected 100 000-year return values.

Figure 45: -0.05 tail length maps for a 100 000 year mean recurrence interval.

C.3 -0.1 Tail Length

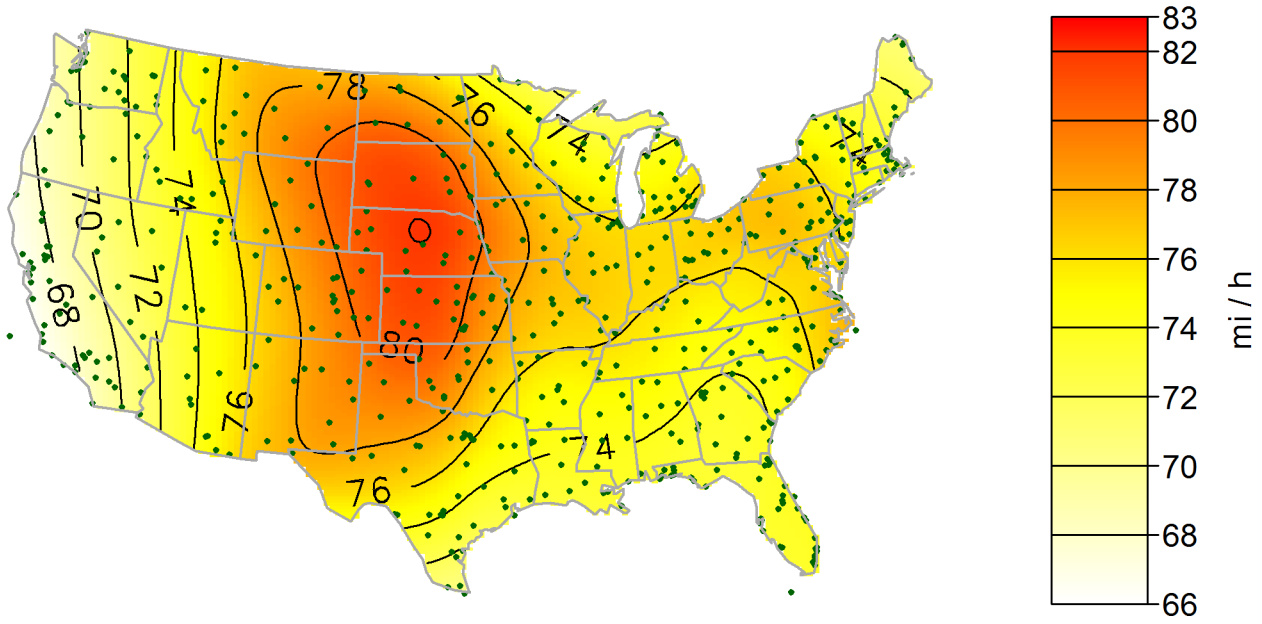


(a) Map of estimated expected 10-year return values.

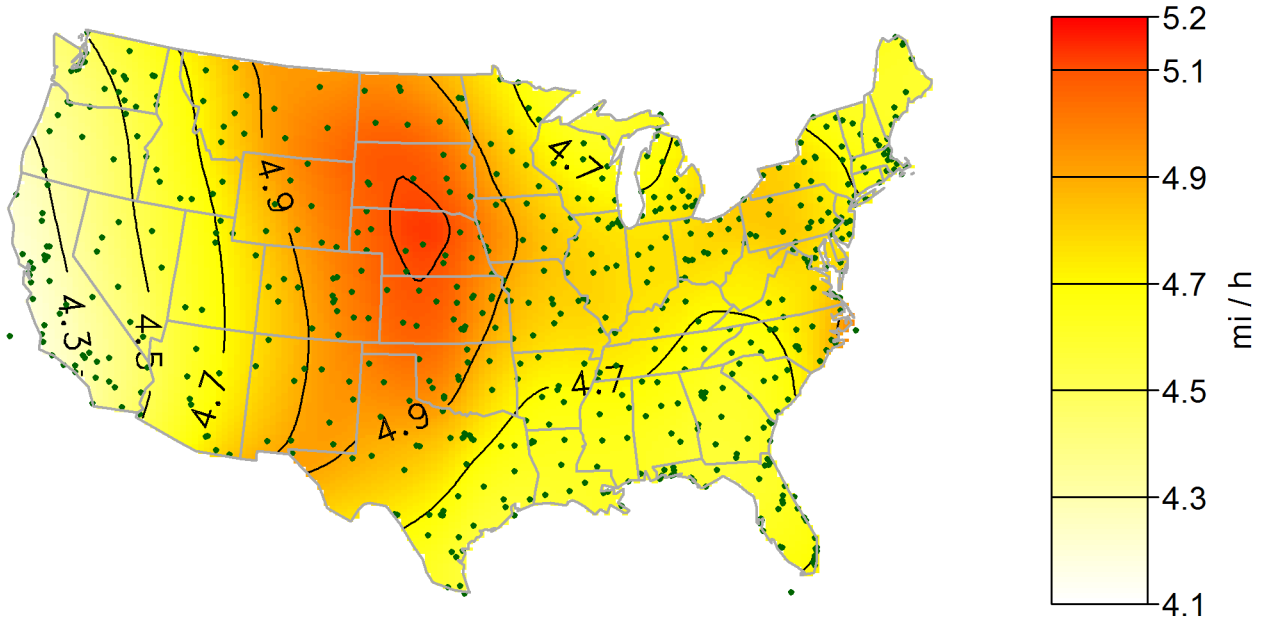


(b) Map of standard errors of estimated expected 10-year return values.

Figure 46: -0.1 tail length maps for a 10 year mean recurrence interval.

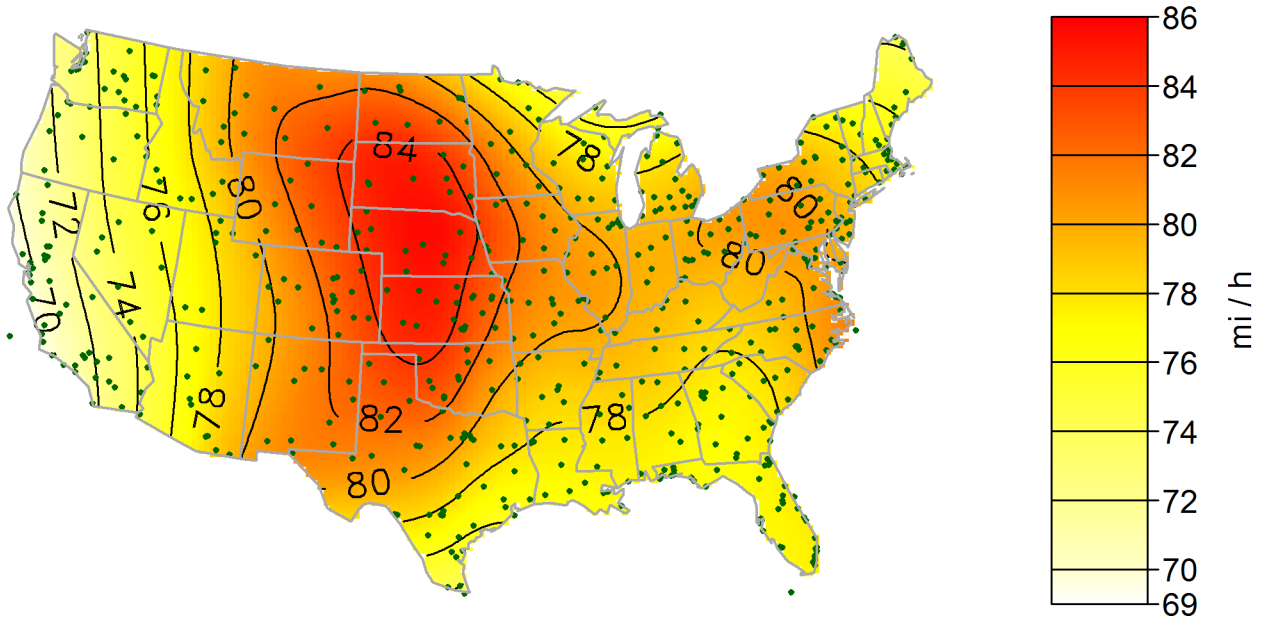


(a) Map of estimated expected 25-year return values.

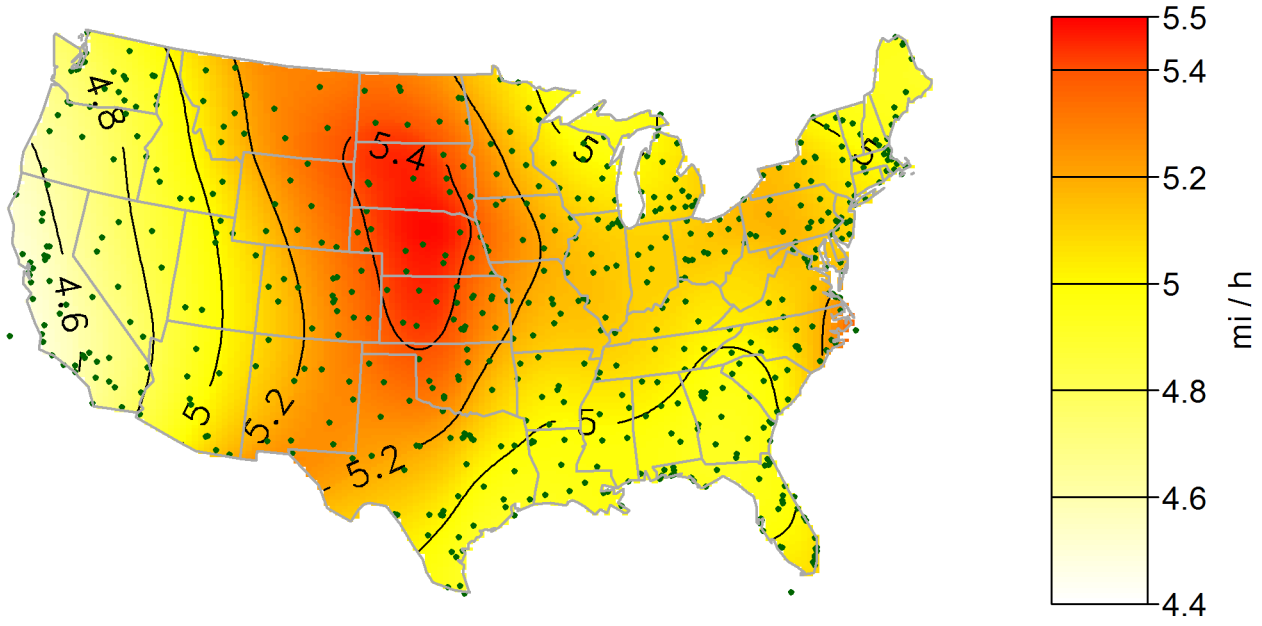


(b) Map of standard errors of estimated expected 25-year return values.

Figure 47: -0.1 tail length maps for a 25 year mean recurrence interval.

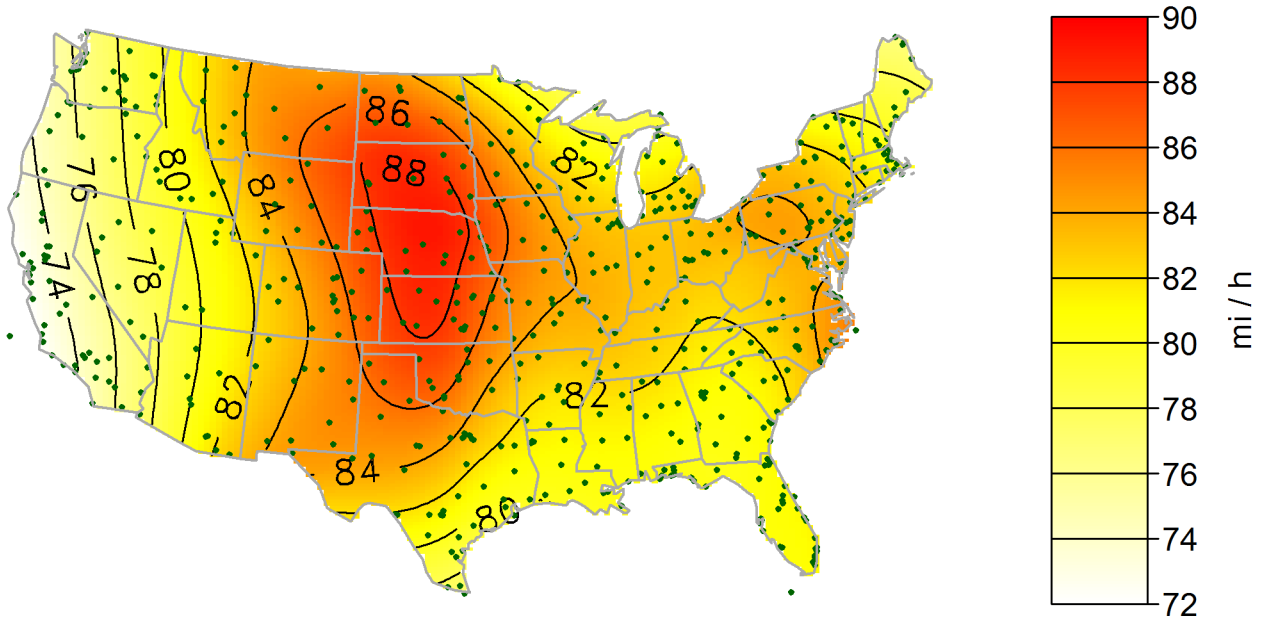


(a) Map of estimated expected 50-year return values.

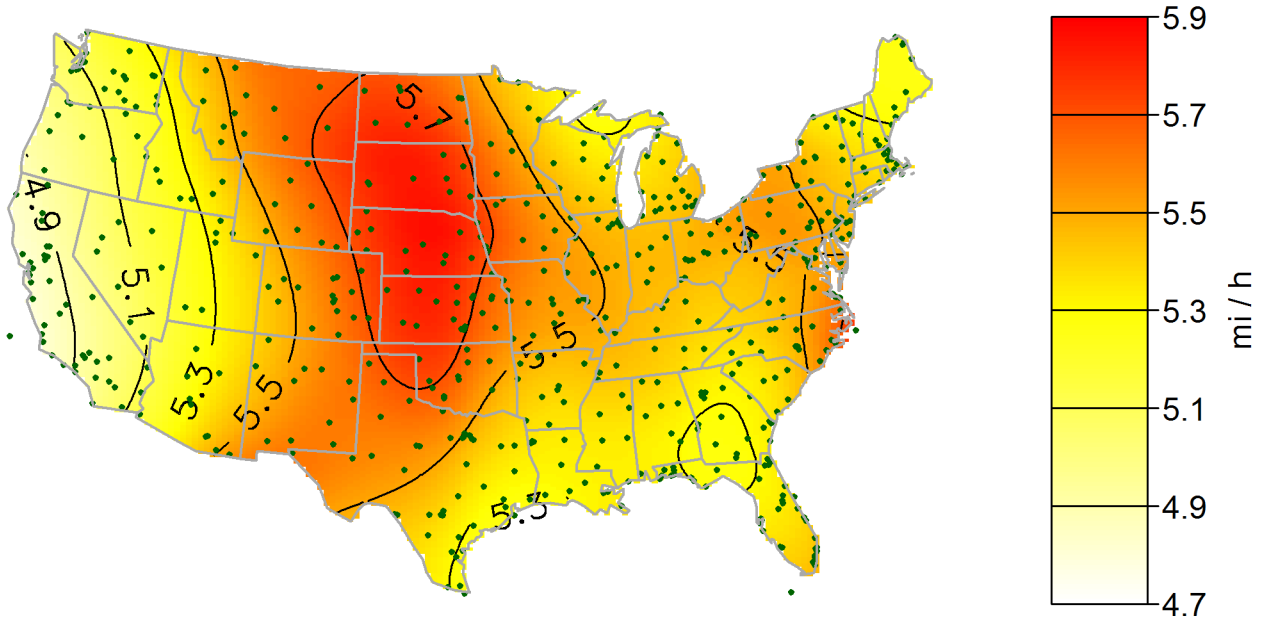


(b) Map of standard errors of estimated expected 50-year return values.

Figure 48: -0.1 tail length maps for a 50 year mean recurrence interval.

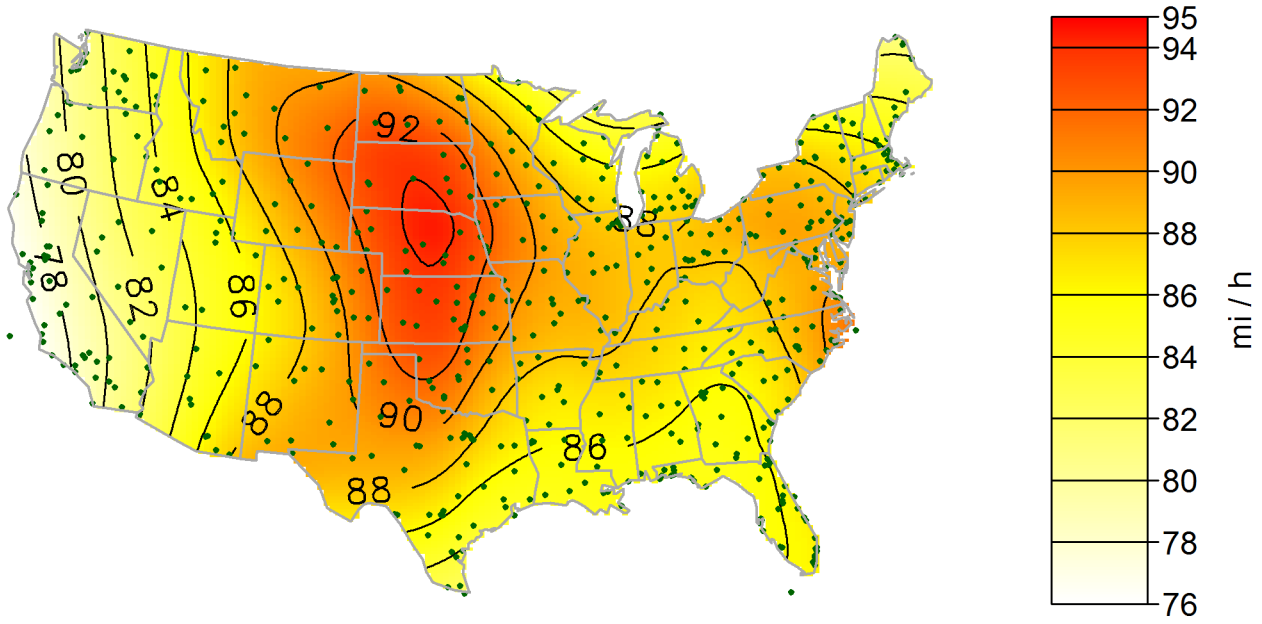


(a) Map of estimated expected 100-year return values.

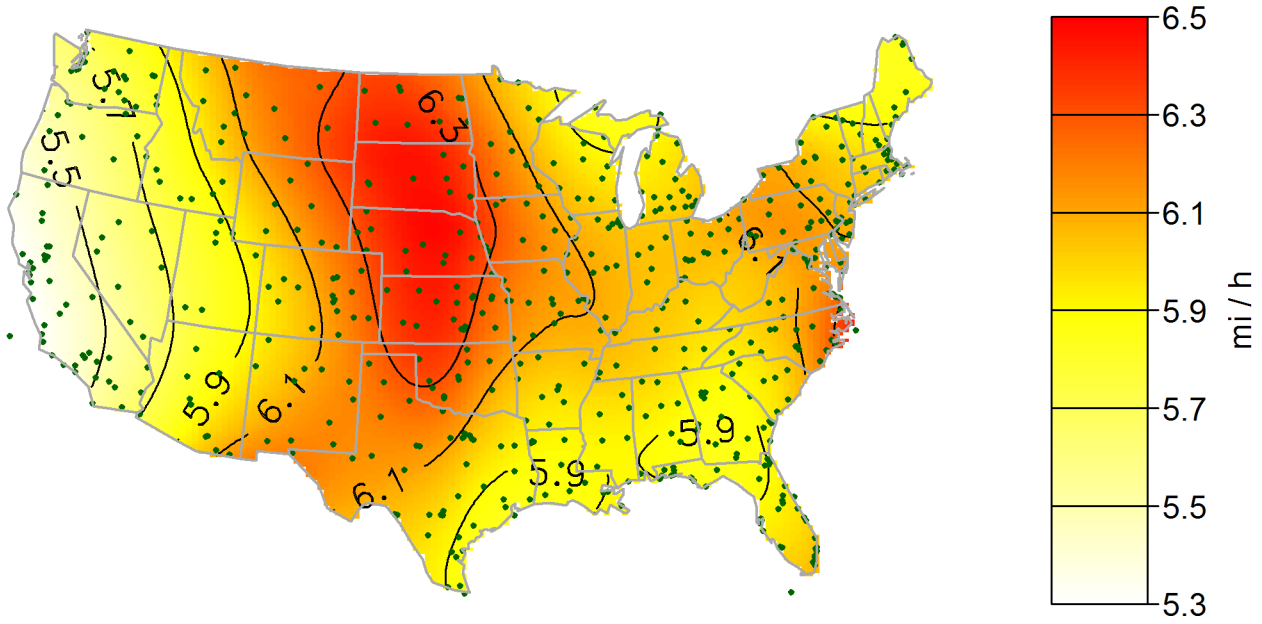


(b) Map of standard errors of estimated expected 100-year return values.

Figure 49: -0.1 tail length maps for a 100 year mean recurrence interval.

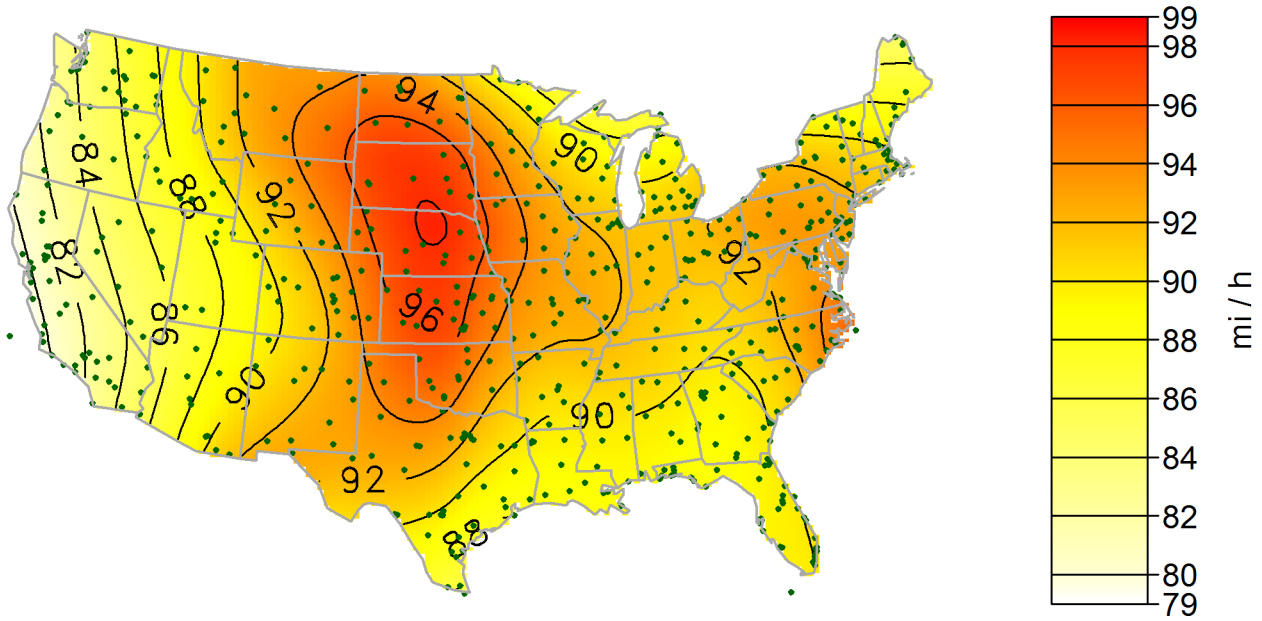


(a) Map of estimated expected 300-year return values.

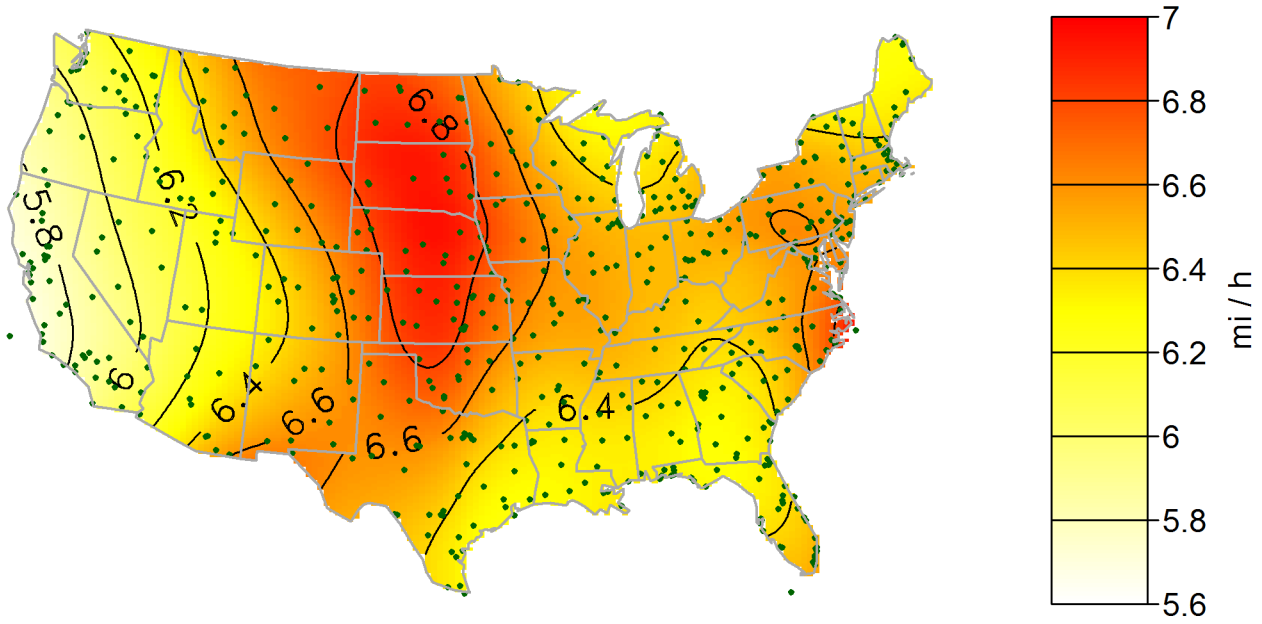


(b) Map of standard errors of estimated expected 300-year return values.

Figure 50: -0.1 tail length maps for a 300 year mean recurrence interval.

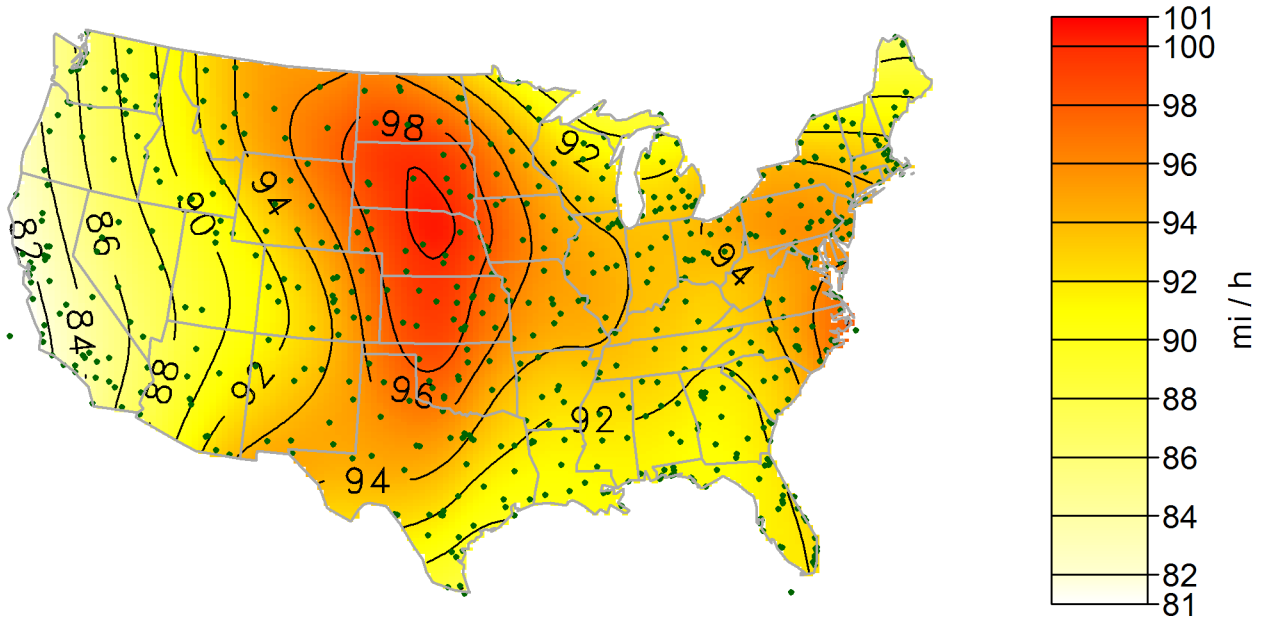


(a) Map of estimated expected 700-year return values.

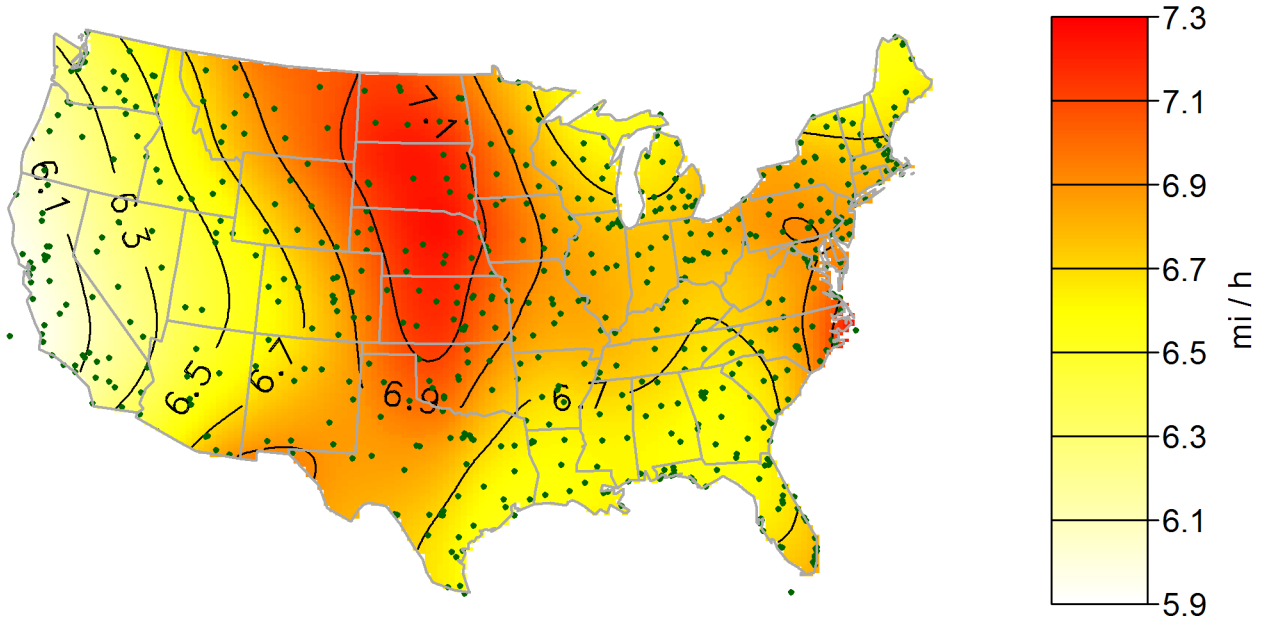


(b) Map of standard errors of estimated expected 700-year return values.

Figure 51: -0.1 tail length maps for a 700 year mean recurrence interval.

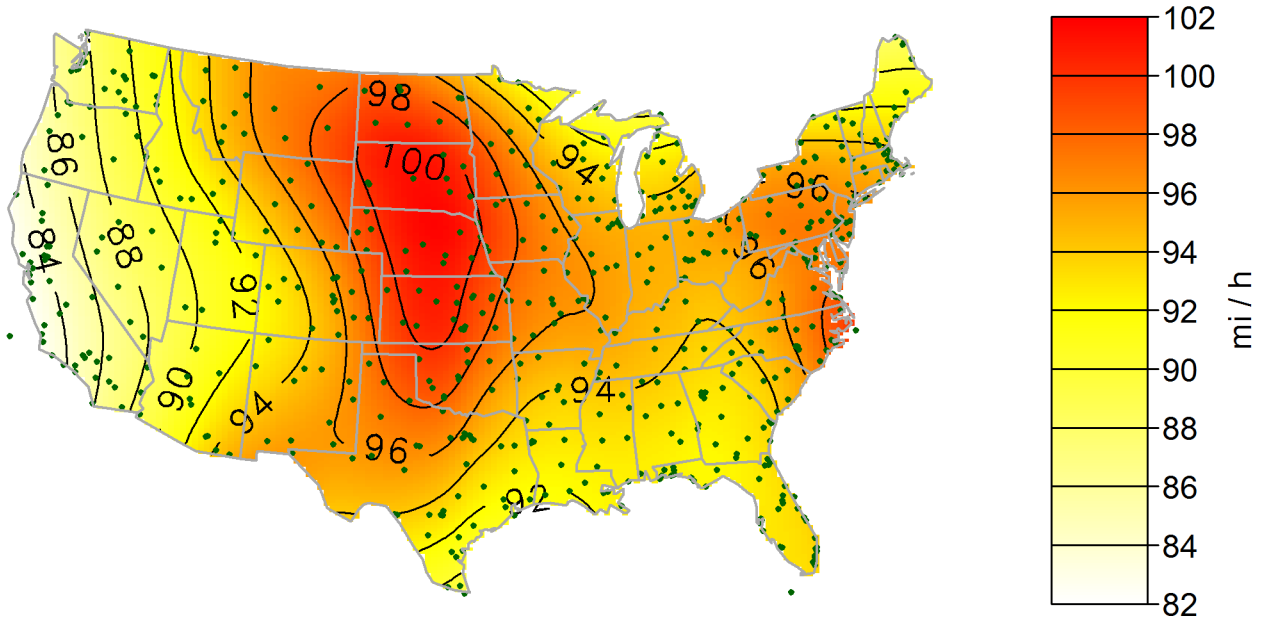


(a) Map of estimated expected 1200-year return values.

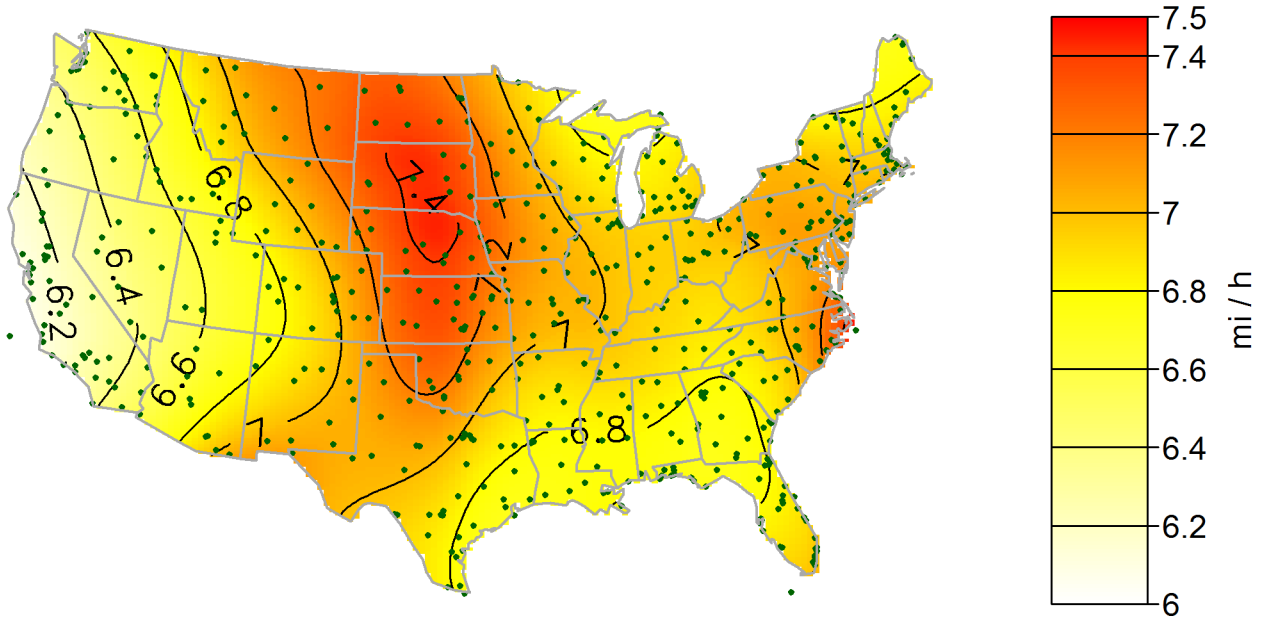


(b) Map of standard errors of estimated expected 1200-year return values.

Figure 52: -0.1 tail length maps for a 1200 year mean recurrence interval.

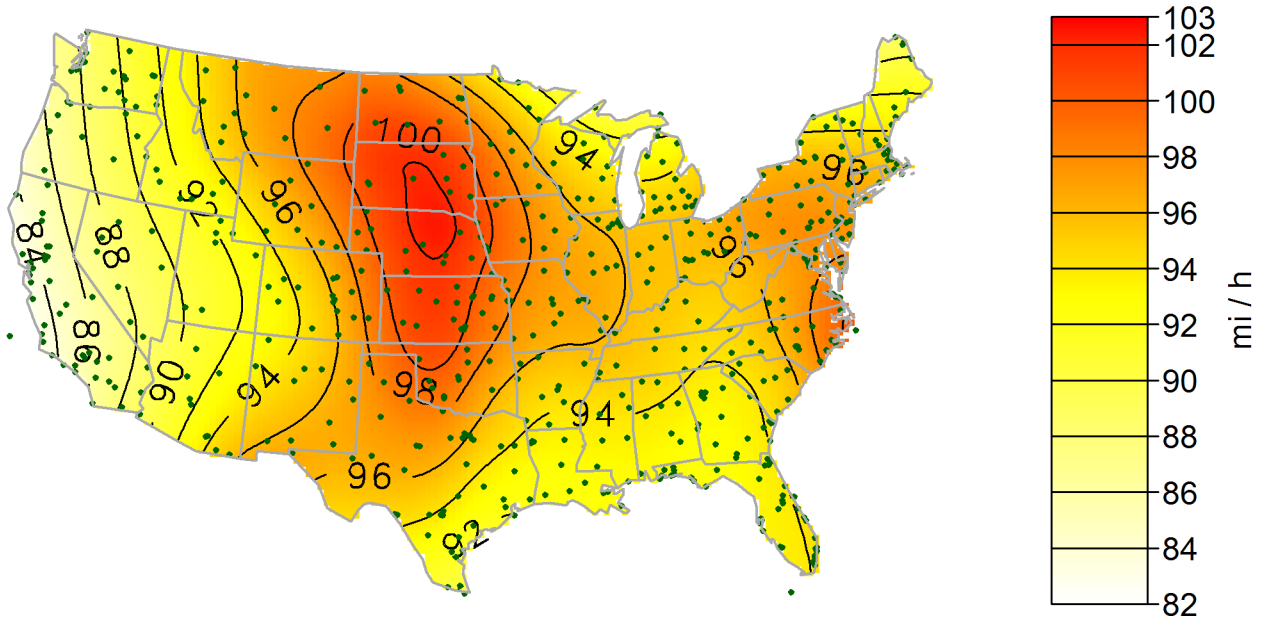


(a) Map of estimated expected 1700-year return values.

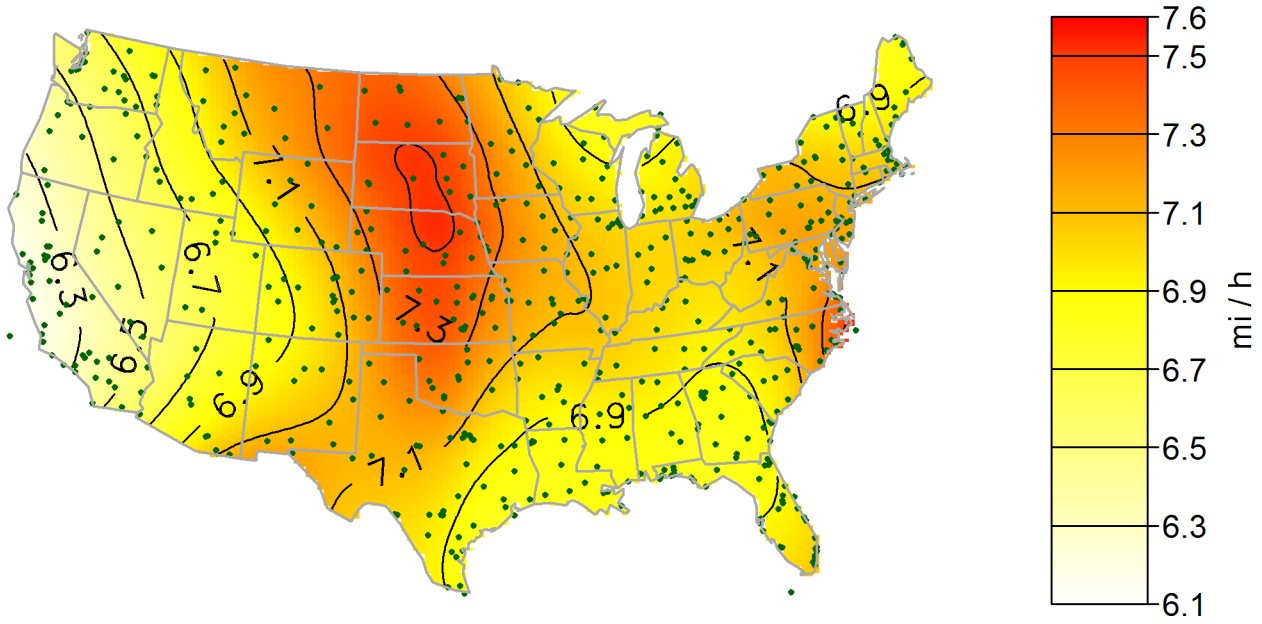


(b) Map of standard errors of estimated expected 1700-year return values.

Figure 53: -0.1 tail length maps for a 1700 year mean recurrence interval.

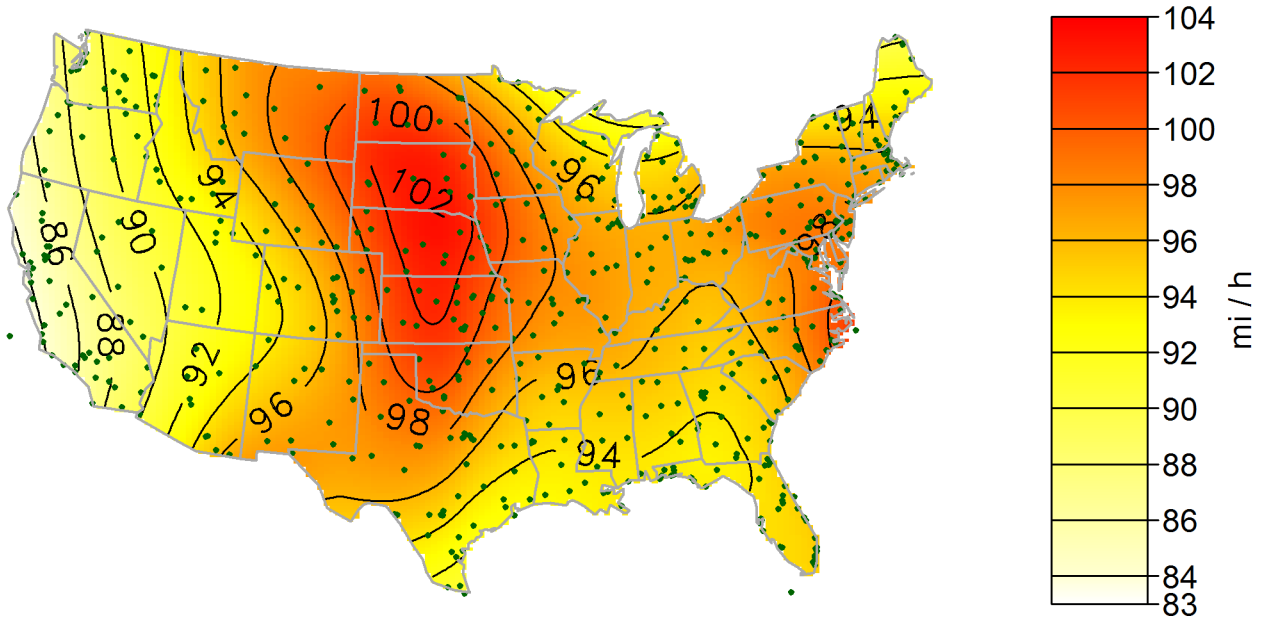


(a) Map of estimated expected 2000-year return values.

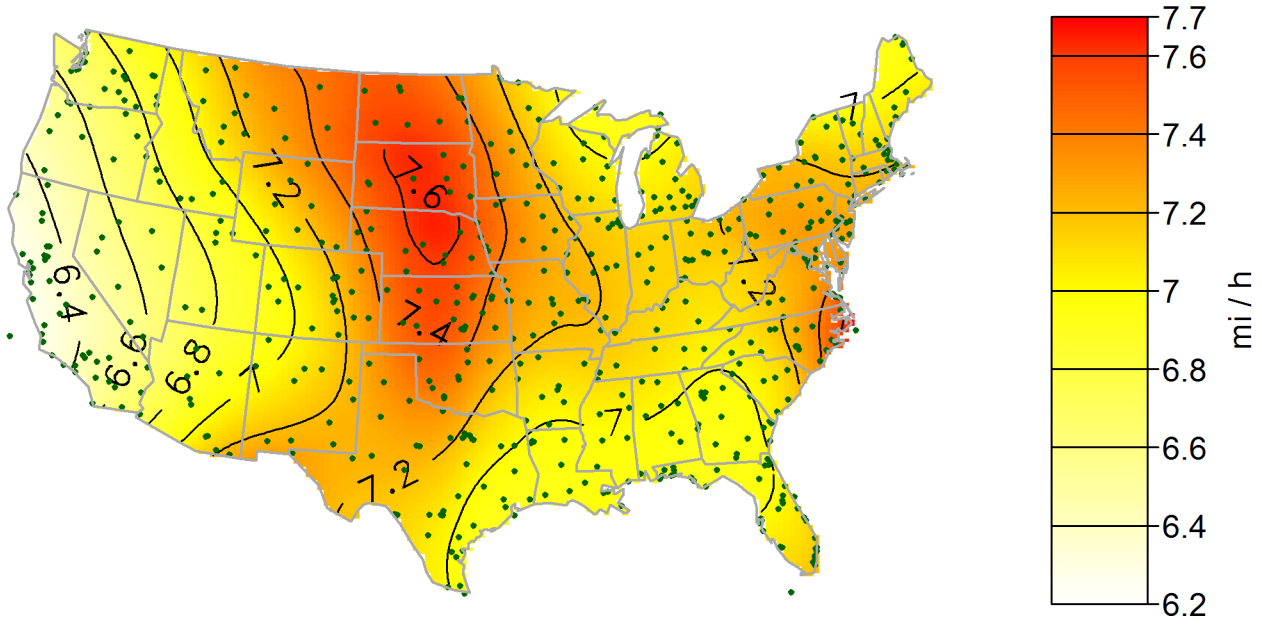


(b) Map of standard errors of estimated expected 2000-year return values.

Figure 54: -0.1 tail length maps for a 2000 year mean recurrence interval.

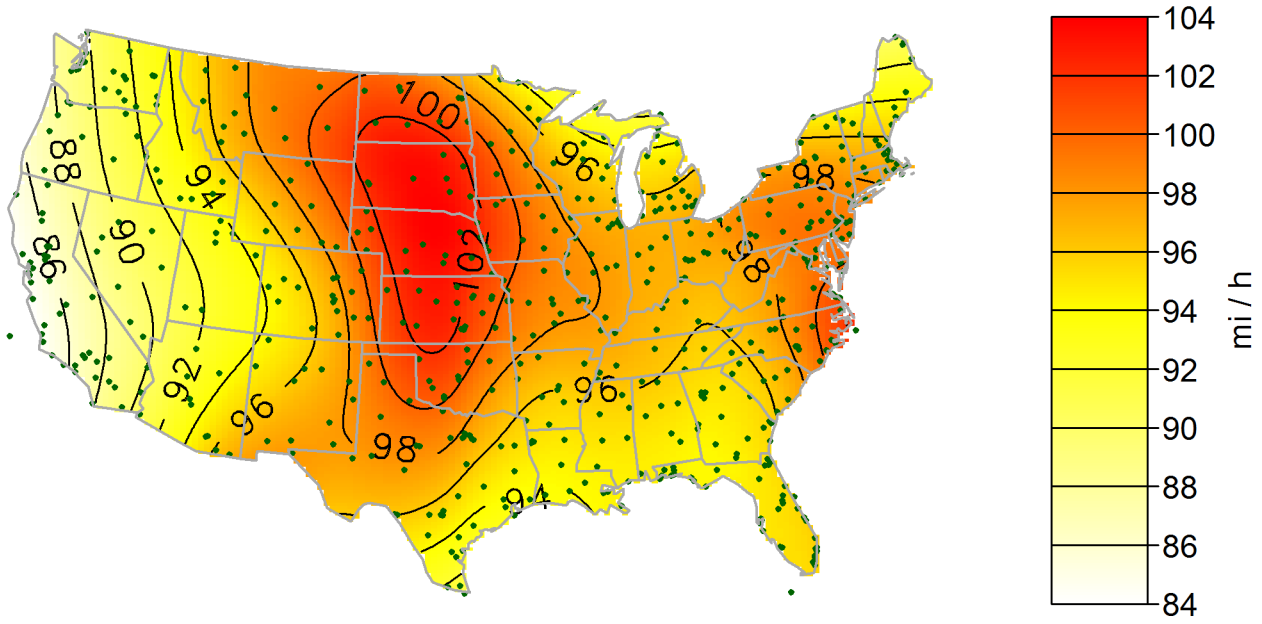


(a) Map of estimated expected 2500-year return values.

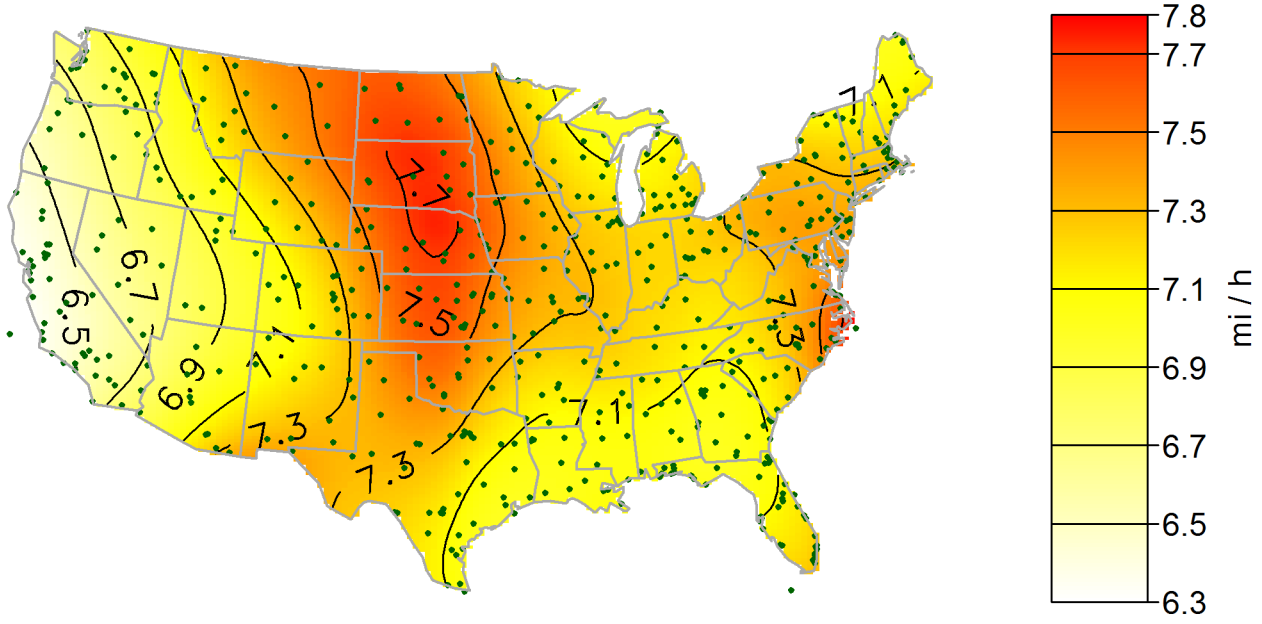


(b) Map of standard errors of estimated expected 2500-year return values.

Figure 55: -0.1 tail length maps for a 2500 year mean recurrence interval.

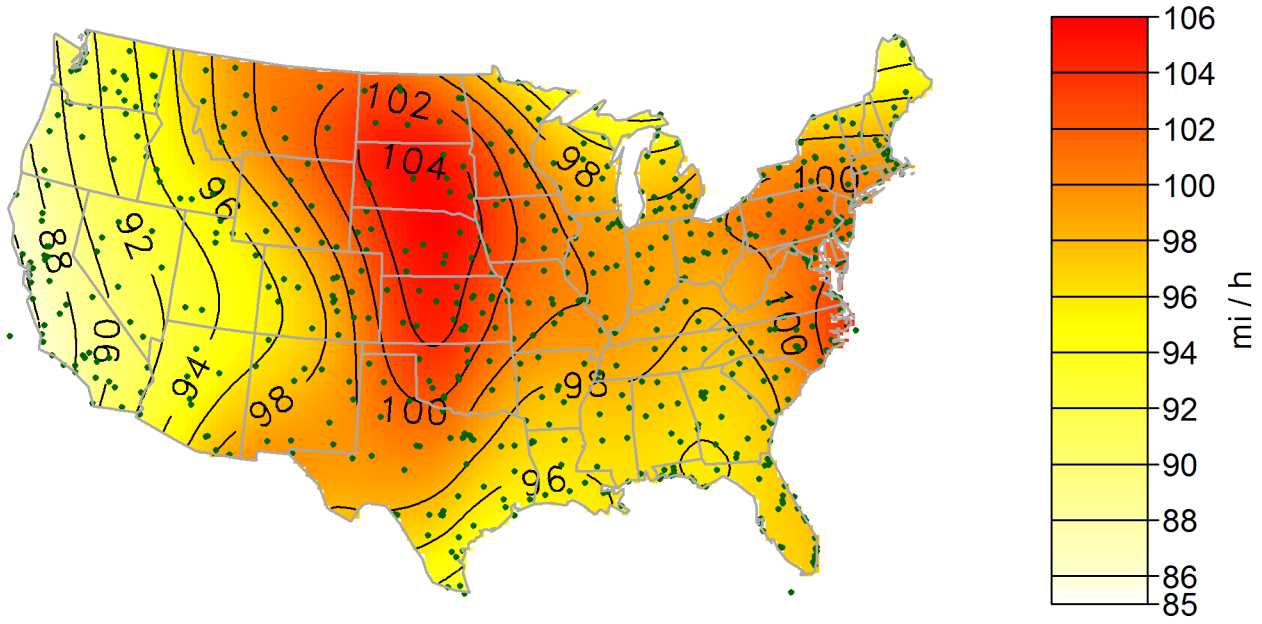


(a) Map of estimated expected 3000-year return values.

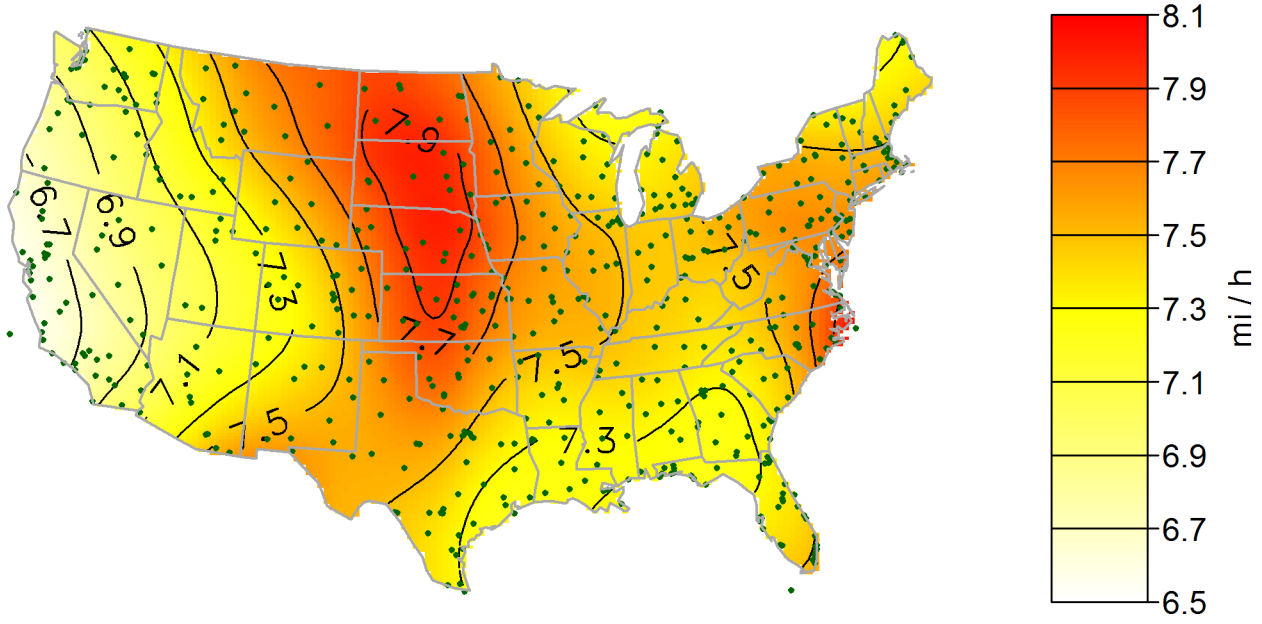


(b) Map of standard errors of estimated expected 3000-year return values.

Figure 56: -0.1 tail length maps for a 3000 year mean recurrence interval.

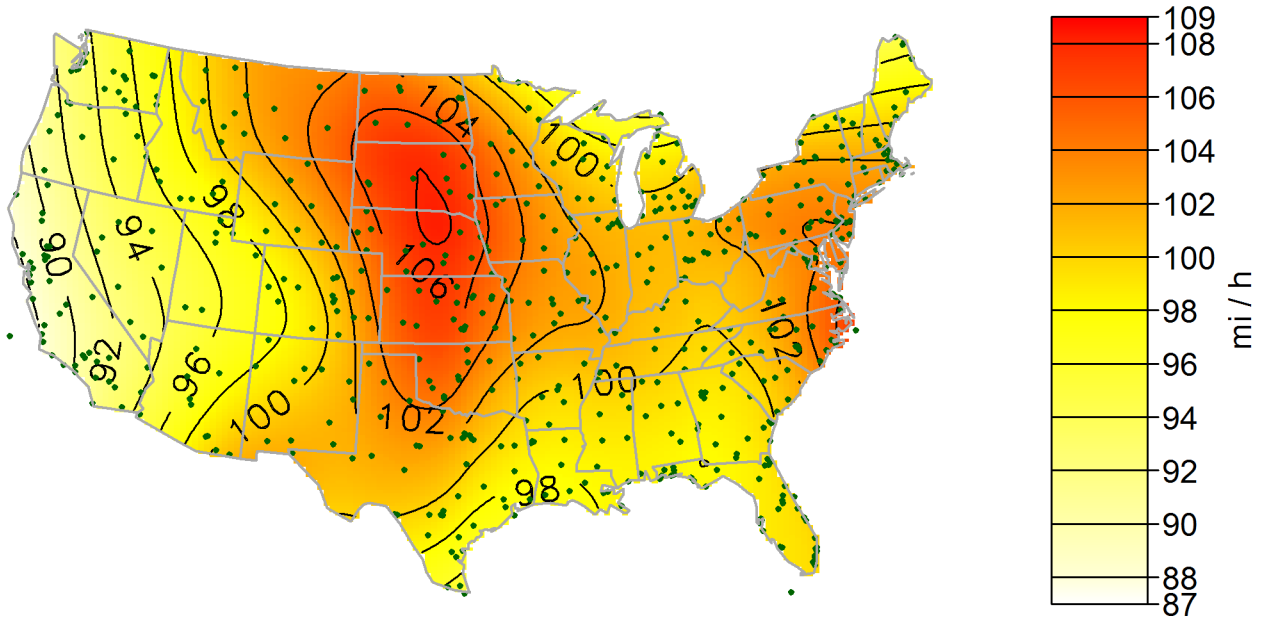


(a) Map of estimated expected 5000-year return values.

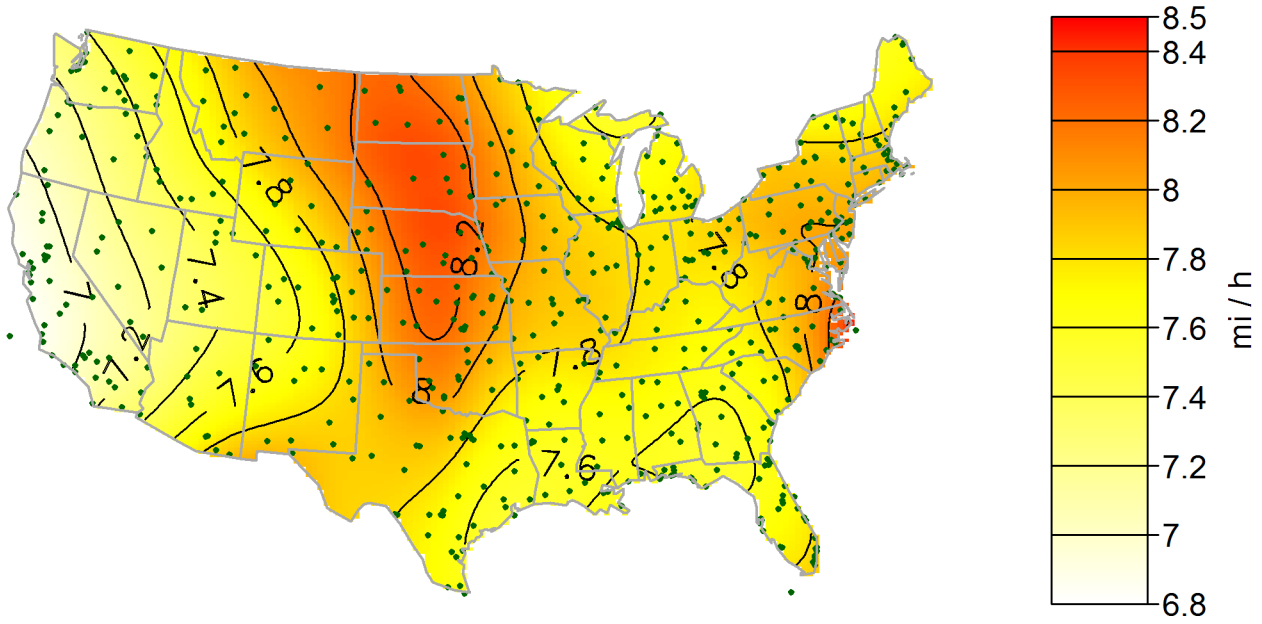


(b) Map of standard errors of estimated expected 5000-year return values.

Figure 57: -0.1 tail length maps for a 5000 year mean recurrence interval.

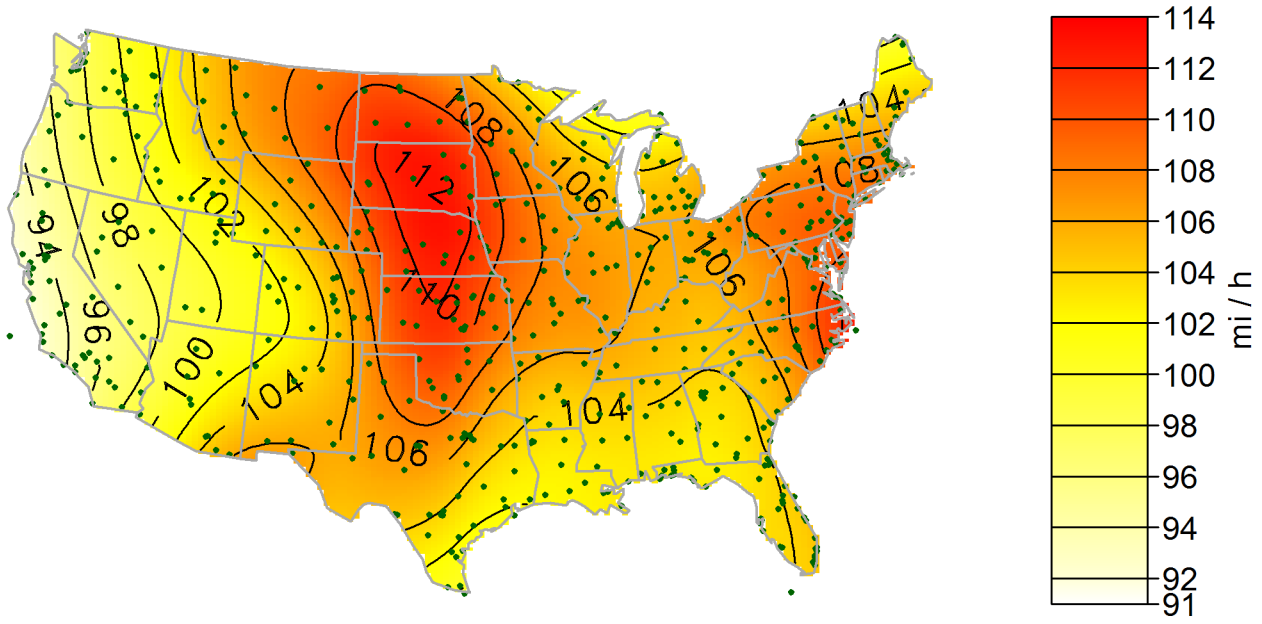


(a) Map of estimated expected 10 000-year return values.

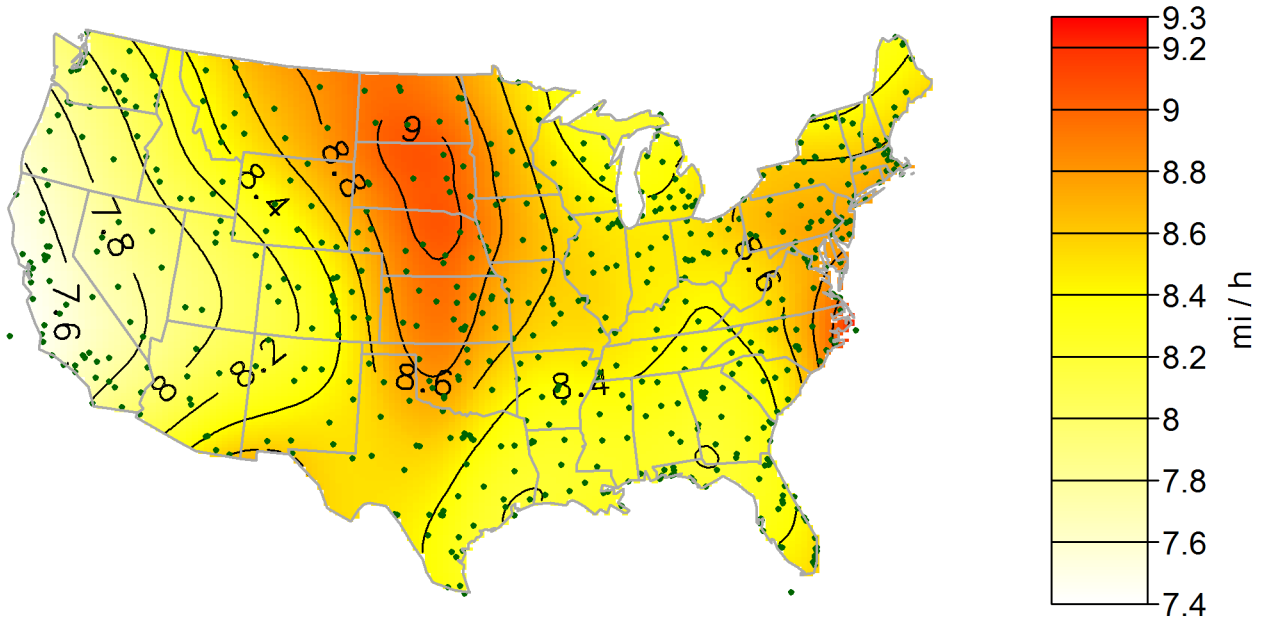


(b) Map of standard errors of estimated expected 10 000-year return values.

Figure 58: -0.1 tail length maps for a 10 000 year mean recurrence interval.

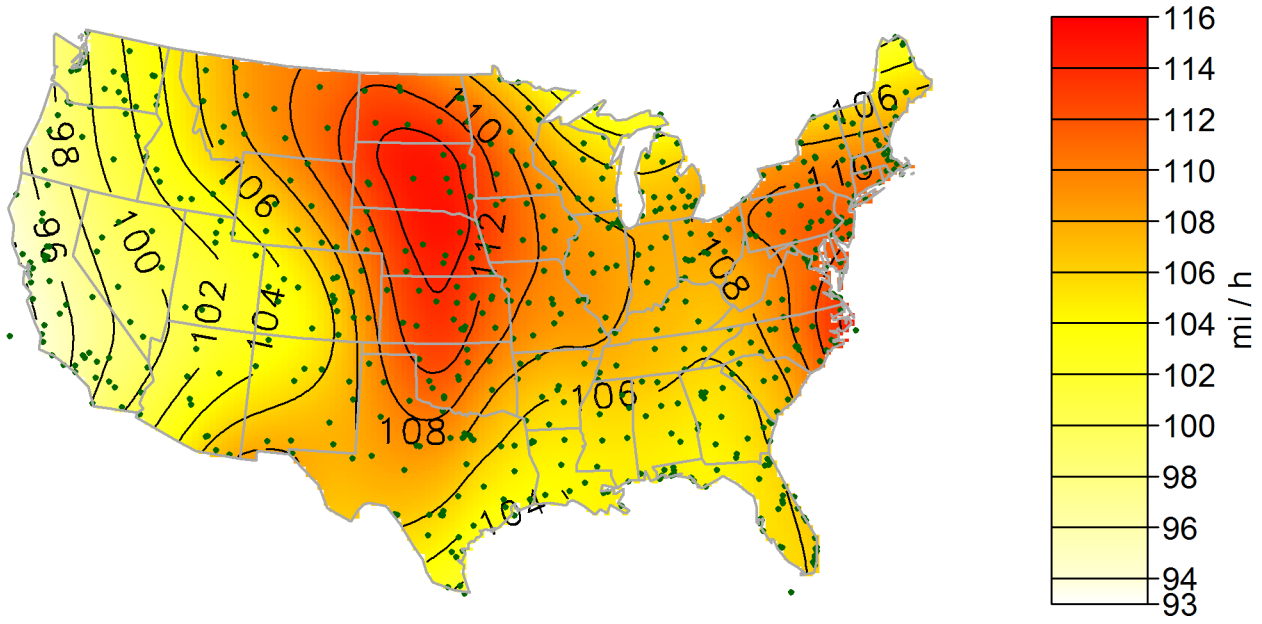


(a) Map of estimated expected 50 000-year return values.

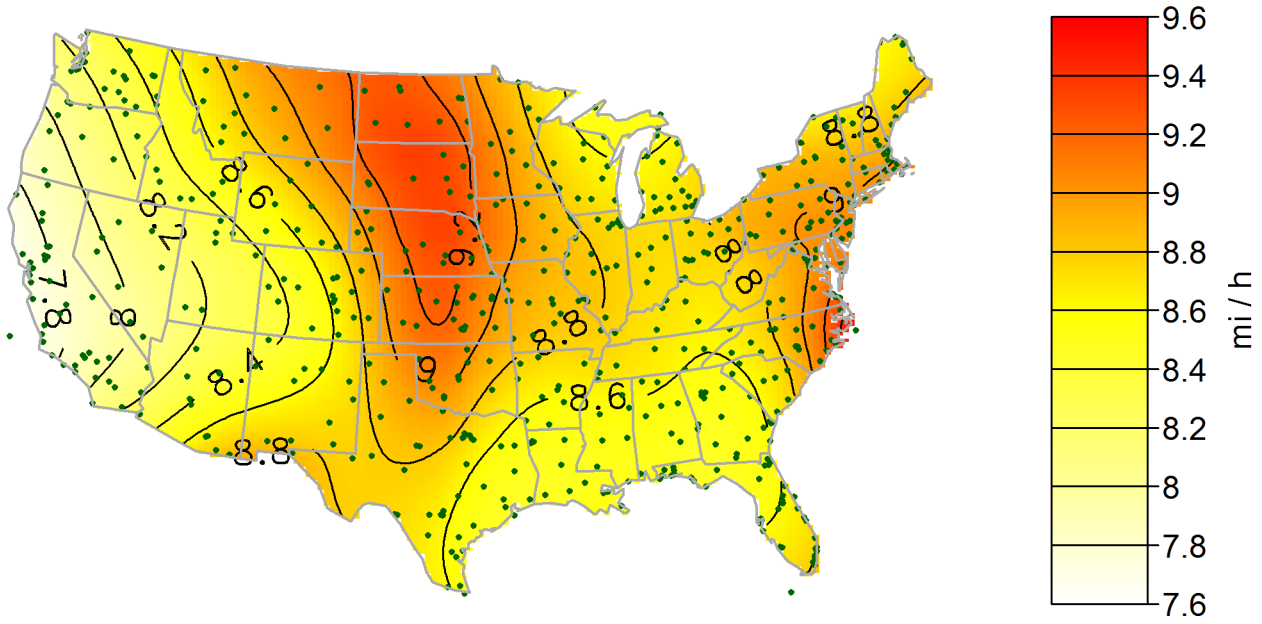


(b) Map of standard errors of estimated expected 50 000-year return values.

Figure 59: -0.1 tail length maps for a 50 000 year mean recurrence interval.



(a) Map of estimated expected 100 000-year return values.



(b) Map of standard errors of estimated expected 100 000-year return values.

Figure 60: -0.1 tail length maps for a 100 000 year mean recurrence interval.

D Station Used in the Creation of the Final Maps

The list of station numbers and geographic coordinates for the 575 meteorological stations meeting all of the requirements of Sec. 3.1.6 and used in development of the final maps presented in App. C is provided in Table 1. The locations of these stations are shown in Figure 61. Links to web sites with more information about the data, and to download the data for these stations, were provided in Sec. 2.

Station	Longitude	Latitude
690190	-99.86	32.42
722010	-81.75	24.55
722020	-80.30	25.82
722024	-80.28	25.91
722025	-80.15	26.07
722026	-80.38	25.49
722029	-80.43	25.65
722030	-80.10	26.68
722039	-80.17	26.20
722040	-80.65	28.10
722045	-80.42	27.66
722049	-80.11	26.25
722050	-81.33	28.43
722053	-81.33	28.55
722056	-81.06	29.18
722057	-81.24	28.78
722060	-81.69	30.49
722065	-81.68	30.23
722067	-81.87	30.22
722068	-81.52	30.34
722070	-81.20	32.12
722080	-80.04	32.90
722085	-80.72	32.48
722090	-81.56	31.89
722103	-80.38	27.50
722104	-82.63	27.77
722106	-81.86	26.59
722108	-81.75	26.54
722110	-82.54	27.96
722115	-82.56	27.40
722116	-82.69	27.91
722135	-82.51	31.54
722137	-81.39	31.25
722140	-84.35	30.39
722146	-82.27	29.69
722160	-84.19	31.54

722166	-83.28	30.78
722170	-83.65	32.69
722180	-81.97	33.37
722190	-84.43	33.64
722195	-84.52	33.78
722196	-84.30	33.88
722200	-85.03	29.73
722210	-86.53	30.48
722215	-86.52	30.78
722223	-87.19	30.47
722225	-87.32	30.35
722226	-87.02	30.72
722230	-88.25	30.69
722245	-85.68	30.21
722246	-86.52	30.65
722250	-84.99	32.34
722255	-84.94	32.52
722260	-86.39	32.30
722267	-86.01	31.86
722270	-84.52	33.91
722280	-86.75	33.56
722286	-87.62	33.21
722287	-85.86	33.59
722310	-90.25	29.99
722315	-90.03	30.04
722317	-91.15	30.54
722320	-89.41	29.35
722340	-88.75	32.33
722350	-90.08	32.32
722358	-90.47	31.18
722359	-90.08	33.50
722400	-93.23	30.12
722405	-91.99	30.20
722410	-94.02	29.95
722420	-94.86	29.27
722430	-95.36	29.99
722435	-95.28	29.64
722445	-96.36	30.59
722446	-94.75	31.23
722448	-95.40	32.35
722470	-94.71	32.38
722480	-93.82	32.45
722484	-93.74	32.54
722485	-93.66	32.50
722486	-92.04	32.51

722487	-92.30	31.39
722500	-97.43	25.91
722505	-97.65	26.23
722506	-98.24	26.18
722510	-97.51	27.77
722516	-97.82	27.50
722517	-98.03	27.74
722523	-98.47	29.34
722526	-99.22	28.46
722530	-98.46	29.53
722533	-99.17	29.36
722535	-98.58	29.38
722540	-97.68	30.18
722545	-97.68	30.18
722550	-96.93	28.86
722556	-97.67	28.37
722560	-97.23	31.61
722585	-96.97	32.73
722590	-97.04	32.90
722594	-97.32	32.97
722595	-97.44	32.77
722596	-97.36	32.82
722597	-98.06	32.78
722599	-96.87	32.68
722610	-100.92	29.37
722615	-100.78	29.36
722630	-100.49	31.35
722636	-102.55	36.02
722650	-102.21	31.93
722656	-103.20	31.78
722660	-99.68	32.41
722670	-101.82	33.67
722680	-104.54	33.31
722686	-103.32	34.38
722687	-104.26	32.34
722690	-106.48	32.38
722700	-106.38	31.81
722725	-107.72	32.26
722730	-110.34	31.59
722735	-109.60	31.47
722740	-110.95	32.13
722745	-110.88	32.17
722780	-111.99	33.44
722784	-112.08	33.69
722800	-114.61	32.66

722860	-117.26	33.88
722868	-116.50	33.83
722886	-118.49	34.21
722897	-120.64	35.24
722900	-117.17	32.73
722950	-118.41	33.94
723013	-77.91	34.27
723030	-79.01	35.17
723035	-78.88	34.99
723040	-75.62	35.23
723060	-78.79	35.87
723068	-77.89	35.85
723075	-76.03	36.82
723080	-76.19	36.90
723085	-76.28	36.93
723086	-76.49	37.13
723087	-76.61	37.13
723090	-76.88	34.90
723095	-77.05	35.07
723096	-77.43	34.70
723100	-81.12	33.94
723106	-79.73	34.19
723110	-83.33	33.95
723120	-82.22	34.90
723125	-82.71	34.49
723140	-80.94	35.21
723145	-81.39	35.74
723150	-82.54	35.43
723170	-79.94	36.10
723183	-82.40	36.48
723193	-80.22	36.13
723230	-86.79	34.64
723235	-87.61	34.74
723240	-85.20	35.03
723260	-83.99	35.82
723265	-85.08	35.95
723270	-86.69	36.12
723306	-88.45	33.65
723320	-88.77	34.26
723340	-89.98	35.06
723345	-89.87	35.36
723346	-88.92	35.59
723403	-92.23	34.75
723405	-92.15	34.92
723407	-90.65	35.83

723415	-93.10	34.48
723417	-91.94	34.17
723418	-94.01	33.45
723419	-92.81	33.22
723440	-94.36	35.33
723445	-94.17	36.01
723446	-93.16	36.26
723489	-89.57	37.23
723495	-94.50	37.15
723510	-98.49	33.98
723525	-99.05	35.01
723526	-99.20	35.34
723527	-99.78	36.30
723530	-97.60	35.39
723544	-97.65	35.53
723545	-97.09	36.16
723546	-97.10	36.73
723550	-98.40	34.65
723560	-95.89	36.20
723566	-95.78	34.90
723575	-98.42	34.57
723604	-100.29	34.43
723627	-108.79	35.51
723630	-101.71	35.22
723650	-106.62	35.04
723656	-106.09	35.62
723658	-108.23	36.74
723676	-103.60	35.18
723677	-105.14	35.65
723723	-112.42	34.65
723740	-110.72	35.02
723783	-112.16	35.95
723805	-114.62	34.77
723810	-117.88	34.91
723815	-116.79	34.85
723816	-118.22	34.74
723820	-118.08	34.63
723825	-117.38	34.60
723860	-115.16	36.08
723870	-116.03	36.62
723890	-119.72	36.78
723910	-119.12	34.12
723925	-119.84	34.43
723940	-120.47	34.92
723965	-120.63	35.67

724010	-77.32	37.51
724020	-75.50	37.94
724030	-77.45	38.94
724035	-77.30	38.50
724037	-77.31	38.72
724040	-76.40	38.28
724045	-75.51	38.34
724050	-77.03	38.87
724060	-76.68	39.17
724066	-77.73	39.71
724070	-74.58	39.46
724075	-75.08	39.37
724080	-75.23	39.87
724085	-75.01	40.08
724086	-75.15	40.20
724089	-75.60	39.67
724090	-74.35	40.03
724095	-74.81	40.28
724096	-74.59	40.02
724100	-79.21	37.34
724106	-79.34	36.57
724110	-79.97	37.32
724177	-77.98	39.40
724210	-84.67	39.04
724220	-84.61	38.04
724235	-85.66	38.23
724240	-85.97	37.91
724243	-84.08	37.09
724275	-80.65	40.18
724280	-82.88	39.99
724286	-81.89	39.94
724288	-83.08	40.08
724290	-84.22	39.91
724295	-83.83	39.83
724297	-84.42	39.10
724336	-89.25	37.78
724340	-90.37	38.75
724345	-90.66	38.66
724350	-88.77	37.06
724375	-86.62	39.14
724380	-86.27	39.71
724386	-86.94	40.41
724390	-89.68	39.84
724396	-91.19	39.94
724400	-93.39	37.24

724450	-92.22	38.82
724455	-92.54	40.10
724456	-91.77	38.13
724457	-92.14	37.74
724458	-92.16	38.59
724460	-94.72	39.30
724466	-94.56	38.84
724467	-93.55	38.73
724468	-94.74	38.85
724490	-94.91	39.77
724500	-97.43	37.65
724504	-97.22	37.75
724505	-97.27	37.62
724506	-97.86	38.07
724507	-95.48	37.67
724510	-99.97	37.77
724515	-100.72	37.93
724550	-96.76	39.05
724555	-96.68	39.13
724556	-96.19	38.33
724560	-95.63	39.07
724565	-95.66	38.95
724580	-97.65	39.55
724585	-98.83	38.87
724586	-97.66	38.81
724620	-105.87	37.44
724635	-103.53	38.05
724636	-102.69	38.07
724640	-104.50	38.29
724645	-104.34	37.26
724650	-101.69	39.37
724655	-99.83	39.38
724660	-104.71	38.81
724665	-103.72	39.19
724676	-106.87	39.22
724680	-104.77	38.70
724690	-104.87	39.75
724698	-103.23	40.17
724755	-113.10	37.70
724756	-112.15	37.71
724760	-108.54	39.13
724765	-107.90	38.51
724796	-111.85	41.79
724800	-118.36	37.37
724830	-121.49	38.51

724836	-121.40	38.67
724837	-121.44	39.14
724838	-121.57	39.10
724839	-121.59	38.70
724855	-117.09	38.06
724860	-114.84	39.30
724885	-118.70	39.42
724917	-121.61	36.66
724926	-120.95	37.63
724930	-122.22	37.76
724935	-122.12	37.66
724936	-122.05	37.99
724945	-121.93	37.36
724955	-122.28	38.21
724957	-122.81	38.51
725025	-74.06	40.85
725035	-73.10	40.79
725036	-73.88	41.63
725037	-73.71	41.07
725046	-72.05	41.33
725064	-70.73	41.91
725065	-70.96	41.68
725066	-70.61	41.39
725067	-70.28	41.67
725070	-71.43	41.72
725075	-73.17	42.70
725080	-72.68	41.94
725088	-70.92	42.58
725090	-71.01	42.36
725097	-70.93	42.15
725098	-71.17	42.19
725103	-75.96	40.37
725115	-76.76	40.19
725118	-76.85	40.22
725125	-78.90	41.18
725126	-78.32	40.30
725130	-75.73	41.34
725140	-76.92	41.24
725144	-76.57	40.44
725150	-75.98	42.21
725156	-76.89	42.16
725170	-75.45	40.65
725180	-73.80	42.75
725185	-73.61	43.34
725190	-76.10	43.11

725196	-75.40	43.23
725197	-75.38	43.15
725200	-80.23	40.50
725205	-79.92	40.35
725210	-81.44	40.92
725240	-81.85	41.41
725246	-82.52	40.82
725250	-80.67	41.25
725260	-80.18	42.08
725266	-78.64	41.80
725280	-78.74	42.94
725290	-77.68	43.12
725300	-87.91	41.99
725305	-88.25	41.91
725306	-87.82	42.08
725314	-90.16	38.57
725315	-88.28	40.04
725316	-88.87	39.98
725320	-89.68	40.67
725330	-85.21	41.01
725335	-86.15	40.65
725340	-87.75	41.79
725360	-83.80	41.59
725366	-83.67	41.01
725370	-83.35	42.22
725374	-83.74	42.22
725375	-83.01	42.41
725376	-83.23	42.24
725390	-84.58	42.78
725395	-84.46	42.26
725396	-85.25	42.31
725430	-89.09	42.20
725440	-90.52	41.47
725450	-91.71	41.88
725455	-91.12	40.78
725460	-93.67	41.54
725465	-92.45	41.11
725470	-90.70	42.40
725480	-92.40	42.55
725485	-93.33	43.16
725500	-95.90	41.31
725510	-96.76	40.83
725520	-98.31	40.96
725525	-98.43	40.60
725527	-96.18	41.76

725533	-95.59	40.08
725540	-95.76	41.12
725555	-99.64	41.44
725560	-97.44	41.98
725570	-96.38	42.39
725610	-102.98	41.10
725620	-100.67	41.12
725625	-100.59	40.21
725635	-102.80	42.06
725636	-103.09	42.84
725640	-104.81	41.16
725645	-105.67	41.31
725650	-104.66	39.83
725660	-103.59	41.87
725690	-106.47	42.90
725705	-109.51	40.44
725717	-107.73	39.53
725720	-111.97	40.79
725744	-109.06	41.59
725745	-107.20	41.81
725750	-112.01	41.20
725760	-108.73	42.82
725775	-111.03	41.27
725780	-112.57	42.92
725785	-112.07	43.52
725805	-118.56	40.07
725825	-115.79	40.83
725830	-117.81	40.90
725866	-114.49	42.48
725867	-113.77	42.54
725910	-122.25	40.15
725920	-122.31	40.52
725945	-124.11	40.98
725955	-122.47	41.78
725970	-122.87	42.39
726050	-71.50	43.20
726055	-70.82	43.08
726060	-70.30	43.64
726083	-68.31	47.29
726088	-68.82	44.81
726116	-72.31	43.63
726145	-72.58	44.20
726170	-73.15	44.47
726185	-69.80	44.32
726223	-74.85	44.94

726225	-73.47	44.65
726227	-76.02	43.99
726228	-74.21	44.38
726350	-85.52	42.88
726355	-86.42	42.13
726357	-85.55	42.23
726360	-86.24	43.17
726370	-83.75	42.97
726375	-83.42	42.66
726379	-84.08	43.53
726380	-84.69	44.37
726387	-85.58	44.74
726390	-83.58	45.07
726400	-87.90	42.95
726410	-89.34	43.14
726416	-90.18	43.21
726430	-91.26	43.75
726435	-91.48	44.87
726440	-92.49	43.90
726450	-88.12	44.51
726456	-88.56	43.98
726463	-89.63	44.93
726499	-94.75	43.41
726500	-95.20	43.16
726510	-96.75	43.58
726540	-98.23	44.38
726545	-98.04	43.78
726546	-97.15	44.93
726550	-94.05	45.55
726555	-94.13	46.41
726556	-95.08	44.55
726557	-95.39	45.88
726580	-93.23	44.88
726590	-98.42	45.45
726620	-103.05	44.05
726625	-103.10	44.15
726660	-106.98	44.77
726665	-107.95	43.97
726686	-100.29	44.38
726770	-108.54	45.81
726776	-109.47	47.05
726785	-112.51	45.95
726796	-112.55	45.26
726797	-111.15	45.79
726810	-116.22	43.56

726815	-115.87	43.04
726830	-118.95	43.59
726835	-121.15	44.25
726837	-117.01	44.02
726880	-118.83	45.70
726886	-117.81	44.84
726930	-123.21	44.13
726940	-123.00	44.91
726980	-122.60	45.59
726985	-122.40	45.55
726988	-121.17	45.62
727033	-67.79	46.12
727120	-68.03	46.87
727125	-67.88	46.95
727340	-84.37	46.47
727347	-84.80	45.57
727415	-89.47	45.63
727435	-84.64	45.87
727437	-88.11	45.82
727440	-88.51	47.17
727450	-92.19	46.84
727455	-92.84	47.39
727470	-93.40	48.57
727476	-94.61	48.73
727530	-96.81	46.92
727535	-98.68	46.93
727575	-97.40	47.96
727576	-97.18	47.95
727640	-100.75	46.77
727645	-102.80	46.80
727670	-103.64	48.20
727675	-101.36	48.43
727676	-101.28	48.26
727680	-106.62	48.21
727720	-111.96	46.61
727730	-114.09	46.92
727750	-111.38	47.47
727790	-114.26	48.30
727796	-112.38	48.61
727810	-120.53	46.56
727825	-120.21	47.40
727826	-119.51	47.30
727827	-119.32	47.21
727830	-117.01	46.38
727840	-119.60	46.57

727845	-119.12	46.27
727846	-118.29	46.09
727850	-117.53	47.62
727855	-117.66	47.62
727857	-117.11	46.74
727890	-119.52	48.46
727920	-122.90	46.97
727930	-122.31	47.46
727934	-122.21	47.49
727935	-122.30	47.53
727937	-122.28	47.91
727938	-122.58	47.27
727976	-122.54	48.79
742060	-122.48	47.14
742070	-122.58	47.08
742300	-105.89	46.43
743920	-69.93	43.90
743945	-71.44	42.93
744655	-88.48	41.77
744860	-73.80	40.66
744900	-71.29	42.47
744904	-71.12	42.72
744915	-72.72	42.16
745060	-122.33	34.79
745310	-104.81	38.97
745700	-84.05	39.83
745940	-76.87	38.81
745980	-76.36	37.08
746716	-86.44	36.98
746930	-78.94	35.13
746943	-76.17	36.26
747020	-119.95	36.33
747185	-115.58	32.83
747187	-116.16	33.63
747188	-114.72	33.62
747686	-88.92	30.41
747750	-85.58	30.07
747770	-86.69	30.43
747804	-81.15	32.01
747810	-83.19	30.97
747880	-82.52	27.85
747900	-80.47	33.97
747946	-80.69	28.61
747950	-81.61	28.24

Table 1: Number and coordinates of the 575 stations contributing to the final maps.

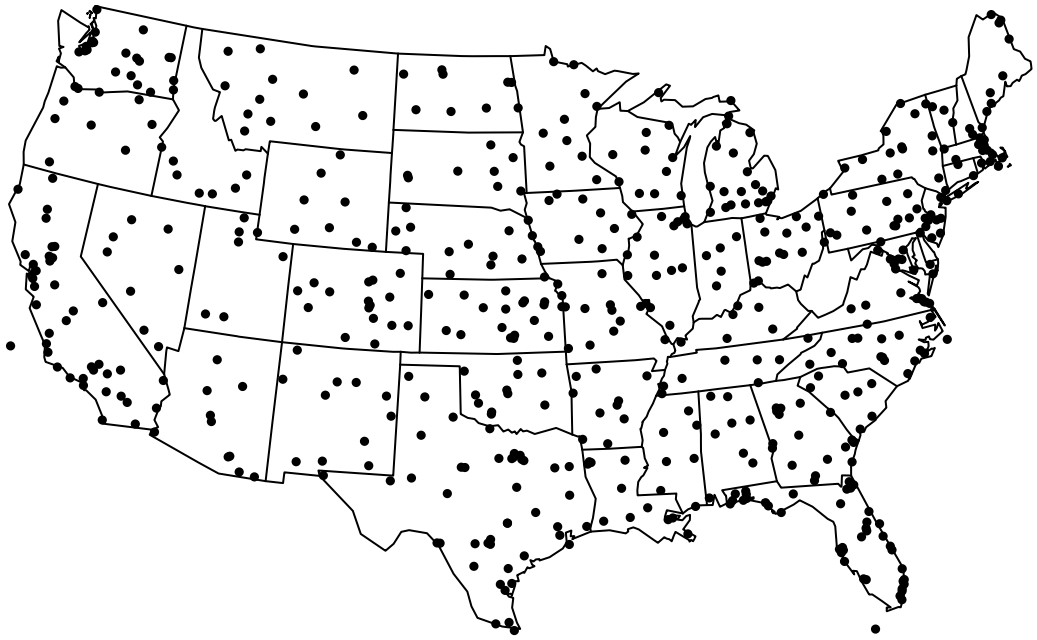


Figure 61: Locations of the 575 stations contributing to the final maps.