RStudio File Edit Code View Plots Session Build Debug Tools Window Help Thu 9:08 AM

~/Documents/DynDocs/rainfall/rain - Shiny

http://127.0.0.1:6173 Open in Browser

HW5_spamDerivedVars.R | Untitled1* | server.R | index.html

Preview

```
<h2>Outline Justification for the Generalized Pareto Model</h2>
<p>
 In this section, we will look at a proof of how GEV and GPD models are
related one or another. Let X follows GEV distribution over a high thresho
u and y is the excess (x-&mu;).
</p>
<p>
  $$Pr\{X > u + y | X > u\} = {\frac{1-Pr(X < x + u)}{1-Pr(X < u)}}$$
  $$ = {\frac{[1 + \frac{&xi;(u + y - &mu;)}{&sigma;}]^{\frac{-1}{&xi
;}}}{[1 + \frac{&xi;(u - &mu;)}{&sigma;}]^{\frac{-1}{&xi;}}}}  $$

  $$ = [1 + {\frac{\frac{&xi;(u + y - &mu;)}{&sigma;}}{1+\frac{&xi;(u -
&mu;)}{&sigma;}}}]^{\frac{-1}{&xi;}} $$
  $$ = [1 + {\frac{&xi; y }{&sigma;}}]^{\frac{-1}{&xi;}} $$
</p>
<p>
  where &sigma;<SUB><FONT size=-1>u</font></sub> = &sigma; + &xi; (u-&mu
;).&sigma;<SUB><FONT size=-1>u</font></sub> is the scale parameter of the
Generalizd Pareto model and &sigma; is the scale of the point process mode
This derivation makes the dependence of the GPD scale &sigma;<SUB><FONT si
=-1>u</font></sub> and threshold u very clear that the GPD shape parameter
is the same as the Generalized Extreme Value(GEV) shape parameter.
</p>

    <h3>Threshold Selection</h3>
<p>
 Selection of a threshold that is too low will result biased parameter
estimates, but a threshold that is too high will result in large variance
the parameter estimates. The idea of selecting a reasonable approximation
to pick a low threshold for the limit subject model.
</p>
```

```
304 146.0435
647 114.8824
333  19.4000
```

Rainfall in Colorado | Analysis | Simulation

Outline Justification for the Generalized Pareto Model

In this section, we will look at a proof of how GEV and GPD models are related one or another. Let X follows GEV distribution over a high threshold u and y is the excess (x-μ).

$$Pr\{X > u + y | X > u\} = \frac{1 - Pr(X < x + u)}{1 - Pr(X < u)}$$

$$= \frac{[1 + \frac{\xi(u+y-\mu)}{\sigma}]^{\frac{-1}{\xi}}}{[1 + \frac{\xi(u-\mu)}{\sigma}]^{\frac{-1}{\xi}}}$$

$$= [1 + \frac{\frac{\xi(u+y-\mu)}{\sigma}}{1 + \frac{\xi(u-\mu)}{\sigma}}]^{\frac{-1}{\xi}}$$

$$= [1 + \frac{\xi y}{\sigma}]^{\frac{-1}{\xi}}$$

where $\sigma_u = \sigma + \xi\,(u-\mu)$.$\sigma_u$ is the scale parameter of the Generalizd Pareto model and σ is the scale of the point process model. This derivation makes the dependence of the GPD scale σ_u and threshold u very clear that the GPD shape parameter is the same as the Generalized Extreme Value(GEV) shape parameter.

Threshold Selection

Selection of a threshold that is too low will result biased parameter estimates, but a threshold that is too high will result in large variance of the parameter estimates. The idea of selecting a reasonable approximation is to pick a low threshold for the limit subject model.

Data input

Choose a weather station:

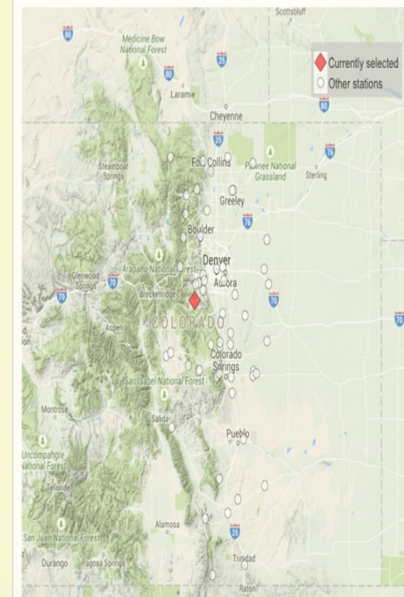
Weather station 17

Threshold:

80 96 100

80 82 84 86 88 90 92 94 96 98 100

Note: this is the quantile at which the threshold is set.



RStudio File Edit Code View Plots Session Build Debug Tools Window Help Thu 9:07 AM

HW5_spamDerivedVars.R | Untitled1* | server.R | index.html

```
<body >
  <h1>Extreme Value Distributions</h1>
  <h2>Introduction</h2>

  <p> In this project, we will look at the probability plot, quantile plot
return
  level plot, and density plot which involves the fitting of data to a
Generalized Pareto Distribution using a range of different thresholds and
estimate its return values of a weather station in Colorado.
  </p>

  <h2>Asymptootic Model Characterization</h2>
  <p> The classical asymptotically model for excesses above a high
  threshold is the generalised Pareto distribution. It is often useful to
look
  at the exceedances over a given threshold instead of the maximum or
minimum of the data.
  </p>

  <h2>The Generalized Pareto Distribution </h2>
  <p>
  Suppose X<SUB><FONT size=-1>1</font></sub>,X<SUB><FONT size=-1>2</font
></sub>,... be a sequence of i.i.d. random variables with distribution
function, <I>F</i>, and let M<SUB><FONT size=-1>n</font></sub> =
  <TT>max</tt>{X<SUB><FONT size=-1>1</font></sub>,...,X<SUB><FONT size=-1>
```

```
304 146.0435
647 114.8824
333  19.4000
 ggplot(data=ruspini, aes(x=x, y=y, color=cluster)) +
eom_point() + theme(legend.position="right") +
eom_point(data = centers,
          aes(x=x, y=y, color=as.factor(c(1,2, 3, 4))),
          size=10, alpha=.3)
```

~/Documents/DynDocs/rainfall/rain - Shiny

http://127.0.0.1:6173 Open in Browser

Rainfall in Colorado | Analysis | Simulation

Extreme Value Distributions

Introduction

In this project, we will look at the probability plot, quantile plot, return level plot, and density plot which involves the fitting of data to a Generalized Pareto Distribution using a range of different thresholds and estimate its return values of a weather station in Colorado.

Asymptootic Model Characterization

The classical asymptotically model for excesses above a high threshold is the generalised Pareto distribution. It is often useful to look at the exceedances over a given threshold instead of the maximum or minimum of the data.

The Generalized Pareto Distribution

Suppose $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with distribution function, *F*, and let $M_n = \max\{X_1,\ldots,X_n\}$. Now, assuming that *F* satisfies certain conditions, then we have that $\Pr\{M_n \le z\} \approx G(z)$, where

?

for some μ, σ > 0 and ξ. Then for a large enough threshold, u, the distribution function of GPD (X-μ), conditional on X > u, is then given by

?

where $z = \frac{(x-\mu)}{\sigma}$, and k, σ, μ are the shape, scale, and location parameters respectively. The scale of σ must be positive, the shape and location can take any real numbers.

Data input

Choose a weather station:

Weather station 6

Threshold:

80 92 100

80 82 84 86 88 90 92 94 96 98 100

Note: this is the quantile at which the threshold is set.

RStudio File Edit Code View Plots Session Build Debug Tools Window Help Thu 9:04 AM

Go to file/function

HW8-4.Rmd HW5_spamDerivedVars.R Untitled1*

```
  })

  #The 4 diagnostic plots
  output$plots2 <- renderPlot({
    fit = fevd(precip, rainfall[[datasetInput(
    par(mfrow = c(2,2), bg = '#FAFAD2',mai = c
    plot(fit, "probprob", main ="Probability P
    plot(fit, "qq", main ="Quantile Plot")
    plot(fit, type = "rl", rperiods= c(2,5,10,
    plot(fit, type = "density", main = "Densit
  })

  #The return levels
  output$rl <- renderText({
    RL2 = ci(fevd(precip, rainfall[[datasetInp
                threshold = quantile(rainfal
                type="GP", units = "mm"),
            return.period=c(2,5,10,20,50,80,1
    paste(capture.output(myprint(RL2)), collap
  })

  # Simulation QQ plot
  output$plots4 <-  renderPlot({
    par(mfrow = c(1,1),bg = '#FAFAD2', mai = c
    plot(fevd(precip, rainfall[[datasetInput2(
```

19:34 <function>(input, output,session)

Console ~/

```
2 43.91304 146.0435
3 98.17647 114.8824
4 68.93333  19.4000
> g1 <- ggplot(data=ruspini, aes(x=x, y=y, color=clu
+     geom_point() + theme(legend.position="right")
+     geom_point(data = centers,
+                aes(x=x, y=y, color=as.factor(c(1,2
+                size=10, alpha=.3)
> g1
```

~/Documents/DynDocs/rainfall/rain - Shiny

http://127.0.0.1:6173 Open in Browser Publish

Rainfall in Colorado Analysis Simulation

Parameter Estimation

Having determined a threshold, the parameters of the generalized Pareto distribution can be estimated by maximum likelihood. Maximum likelihood estimation of the parameters (k, α) was considered by DuMouchel (1983), Davison (1984), R. L. Smith (1984, 1987), J. A. Smith (1986), and Joe (1987). To compute the GPD maximum likelihood estimates, there are two values of (k, α) that must be investigated, such that $\mathfrak{R} = \{k < 0, \alpha > 0\} \cup \{0 < k \leq 1, \frac{\alpha}{k} > X_n\}$. The first is the local maximum of the loglikelihood in space $\mathfrak{R}$ and the second is at the boundary space $\mathfrak{R}$, where k = 1. Suppose that $X = \{X_1, X_2 \ldots, X_n\}$ is a random sample from the GPD with largest value of X_n. Then, the loglikelihood is given by

$$l(\sigma, \xi) = klog\sigma - (1 + \frac{1}{\sigma})\sum_{i=1}^{k} log(1 + \frac{\xi y_i}{\sigma}), k \neq 0$$

$$l(\sigma, \xi) = -klog\sigma - \frac{1}{\sigma}\sum_{i=1}^{k} y_i, k = 0$$

where σ > 0 for k > 0 and $\sigma > kX_{(n)}$ for k > 0. If k > 1, then there is no maximum likelihood estimate because $\lim_{\frac{\alpha}{k} \to X^+} L(k, \alpha; X) = \infty$

Return Levels

When measuring the extreme value models, it is more convenient to interpret them in terms of return levels or quantiles rather than individual parameter values. Return levels are best to measure on the annual scale, so that the N-year return level is the level that is expected to be exceeded once in every N years. Suppose that a generalized Pareto distribution with parameters σ and ξ is a suitable model for exceedances of a threshold u by a variable X, such that

$$Pr(X > x | X > u) = [1 + \xi(\frac{x - u}{\sigma})]^{\frac{-1}{\xi}}$$

Data input

Choose a weather station:

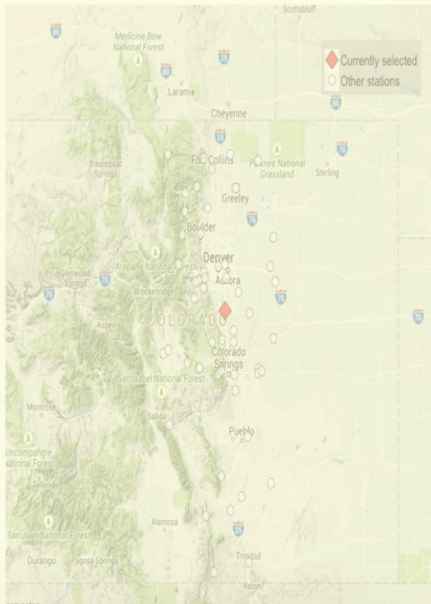
Weather station 7

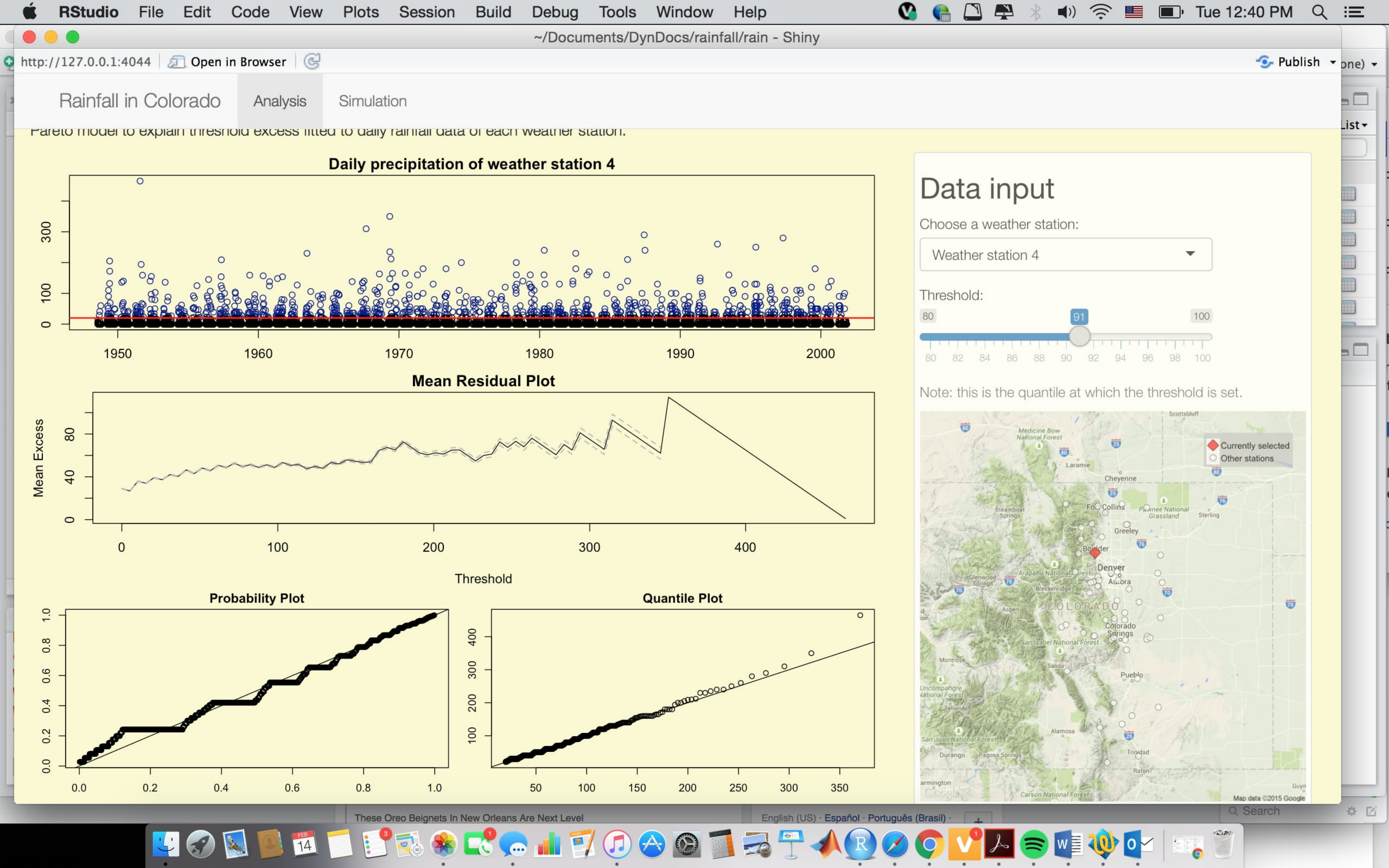
Threshold:

80 88 100

80 82 84 86 88 90 92 94 96 98 100

Note: this is the quantile at which the threshold is set.




RStudio File Edit Code View Plots Session Build Debug Tools Window Help
Tue 12:40 PM
~/Documents/DynDocs/rainfall/rain - Shiny
http://127.0.0.1:4044
Open in Browser
Publish
Rainfall in Colorado
Analysis
Simulation
Pareto model to explain threshold excess fitted to daily rainfall data of each weather station.
Daily precipitation of weather station 4
Mean Residual Plot
Mean Excess
Threshold
Probability Plot
Quantile Plot
Data input
Choose a weather station:
Weather station 4
Threshold:
80
91
100
Note: this is the quantile at which the threshold is set.
Currently selected
Other stations
Cheyenne
Fort Collins
Greeley
Boulder
Denver
Aurora
Colorado Springs
Pueblo
COLORADO
Map data ©2015 Google
These Oreo Beignets In New Orleans Are Next Level
English (US) · Español · Português (Brasil) ·
Search

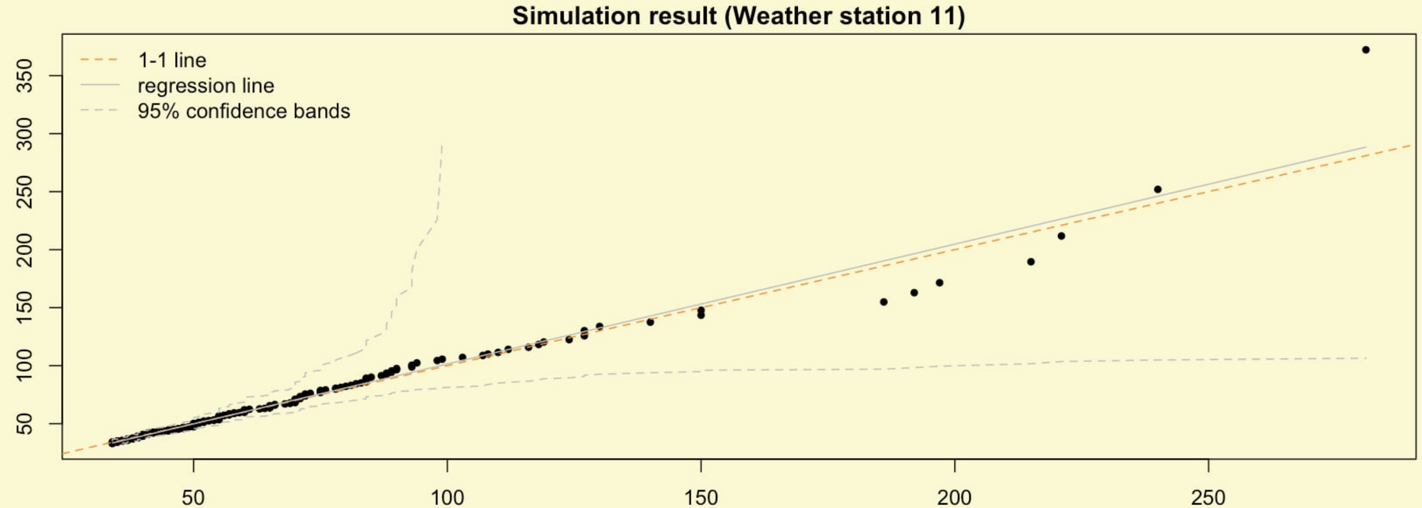
RStudio File Edit Code View Plots Session Build Debug Tools Window Help Sat 1:14 AM

~/Documents/DynDocs/rainfall/rain - Shiny

http://127.0.0.1:7438 Open in Browser Publish

Rainfall in Colorado | Analysis | Simulation

Simulation



Re-Run simulation

Generate a QQ plot of quantiles from model-simulated data against the data.

Summer project, *by Wuji and Ryan.*

Data input

Choose a weather station:

Weather station 11

Threshold:

80 95 100

80 82 84 86 88 90 92 94 96 98 100

Note: this is the quantile at which the threshold is set.