$$[Y|X,\theta][X|\theta][\theta]$$

# Observed monthly precipitation, min and max temperatures for the coterminous US 1895-1997

These are complete "data products" where missing station values have been filled in using spatial statistics. When using the complete data for further statistical analysis care should be taken with the infilled values. Although they are good estimates of the mean they may not reproduce the variability that one would expect from actual point observations of the meteorology. In statistical language, the infilled values are the mean of the conditional distribution for the measurement given the observed data. They are not samples from this conditional distribution.

A convenient subset of min/max and precip for Colorado stations has been formatted as an R data object. See Colorado Monthly Meteorological Data

Much of our analysis uses the R package and we also recommend the Fields library for plotting and spatial analysis.

## Precipitation:

**Data file format**>Acquire the (100Mb) tar file: NCAR_pinfill_others.tar

In UNIX

```
tar -xvf NCAR_pinfill_others.tar
```

Will extract to the subdirectory **NCAR_pinfill**. Precipitation units are in total millimeters per month and the time span is 1895-1997. There are a total of 11918 station locations, thus each yearly file has 11918 lines.

Metadata on these stations is found in **METAinfo**. The first row gives the columns headings and subsequent rows have the information:
**station code, longitude, latitude, elevation**
where some of the station codes contain characters. The (244kb) text file USmonthly.names.txt is a table ( station code, place name) that can be used to find the geographic name of a station. Not all the precipitation stations are in this list however.

The complete precipitation files based on regular station data have the names **ppt.complete.Y**_nnn_ where _nnn_ = **001, 002, ..., 103** and 001=1895 and 103=1997.

Each separate data file consists of the precipitation for a single year. Each line of the file is data for one station according to the format: station id, 12 temps ( jan-dec), 12 missing value/infill codes (1=missing, 0=present) and is written with the FORTRAN statement `format(a8,12I5,2x, 12I1)`. The stations appear in *exactly* the same order as in the metadata file.

**Statistical methodology for infilling monthly precipitation:** When a data value is missing a statistically infilled value appears and the statistical details of this process are given in the technical report: Johns, C., D., Nychka, T. Kittel, and C., Daly, **2001**: Infilling Sparse Records of Spatial Fields (to appear in JASA). Some details of the models and estimates are collected in Supplement to JASA article Finally, the entire analysis and infill process can be reproduced using Matlab, R and F77 programs and the interested researcher should contact Doug Nychka (nychka "at" ucar "dot' edu) for details. Archived volume for this project is 678MB.

**NOAA related dataset and data product**:The infilled precipitation and temperature records were subsequently used to create a fine (4km) gridded, publicly available data product: "103-Year High-Resolution Climate Data Set for the Conterminous United States " maintained and distributed by NOAA/NCDC. The FTP distribution for this final product along with supporting meta-data can be found at www1.ncdc.noaa.gov/pub/data/prism100.

## Getting a particular station

In UNIX in the directory **NCAR_pinfill**:

```
grep 010008 ppt.complete* > first.station.data
wc first.station.data
    103    1442   10403
```

This will have all the years for a station in the right order.

## Reading data into R

To read in metadata:

```
temp<- read.table( "METAinfo")
# check out locations
plot( temp$lon, temp$lat, pch=".")
```

To read in a particular station source the R code in the file get.station.R

```
id<- '010008'
look<- get.station(id, with.infill=T, type="ppt")
```

To read in a particular year, **yr** and deal with missing obs. see the source code in single.year.R (This is ugly only because it hard to read fixed format numbers without spaces into R.) Here is an example that is used to create the fields example data set **RMprecip**. It assumes that you are in the directory **NCAR_pinfill**.

```
single.year( 1963, type="ppt")-> dat
scan("METAinfo", skip=1, what=list( "a", 1,1,1))-> look
names( look)<-c("station.id", "lon", "lat","elev")
ind<- look$lon<  -102 &  look$lon> -112 & look$lat< 55 &  look$lat>35
```

```
x<- cbind(look$lon[ind],look$lat[ind] )
dimnames( x) <- list( look$station.id[ind], c("lon", "lat"))

elev<- look$elev[ind]
y<- dat[ind,8] # column 8 is Aug.

ind2<- !is.na( y)
y<- y[ ind2]
x<- x[ind2,]
elev<- elev[ind2]
RMprecip<- list( x=x, elev=elev, y=y)
```

To create a complete time series use a "for" loop with the year file names and accumulate what you need ... a convenient time to get some coffee while this is running.

### Hey, I just want a small subset of stations, there must be an easier way

First read in the metadata file to help in your choice of station.

```
scan("METAinfo", skip=1, what=list( "a", 1,1,1))-> look
names( look)<-c("station.id", "lon", "lat","elev")
```

To find a subset that covers Colorado (with a bit extra):

```
ind<- look$lon<  -101 &  look$lon> -109.5
ind<- ind&look$lat<41.5 &  look$lat>36.5
# check the results
library( fields)
US()
points( look$lon[ind], look$lat[ind])
# Colorado station id's
colo.id<- look$station.id[ind]
```

To grab first station in CO subset, only real precip data included

```
look<- get.station( colo.id[1],with.infill=FALSE, type="ppt")
```

This output dataset will be a 103X12 matrix with a missing value code (NA) where observations were not taken. See the R example for the "soup to nuts" process of creating R datasets from these files.

---

# Temperatures:

Acquire the (100Mb) tar file: NCAR_tinfill_others.tar

In UNIX

```
tar -xvf NCAR_tinfill_others.tar
```

Will extract to the subdirectory **NCAR_tinfill**

Metadata on these stations is found in **METAinfo** . Columns of this file are:
station code, elevation, longitude, latitude
(however elevation is not used in any of the infill procedures.) The stations for temperature **may not be** the same as those reporting precip. *Do not be fooled, station ids contain some characters!* The (244Kb) text file [USmonthly.names.txt](USmonthly.names.txt) is a table ( station code, place name) that can be used to find the geographic name of a station. Not all the temperature stations may be listed.

There are a total of 8125 station locations. The data file names are of the form: **tmax.complete.Ynnn** and **tmin.complete.Ynnn** with **nnn** $= 001, 002,$ ..., 103; and consist of the values for a particular year with 001=1895 and 103=1997. Temperature appears as a integer in *tenths* of degree C. So 73 should be interpreted as 7.3 degrees C or (9/5)* 7.3 + 32= 45.14 degrees F.

The format for each line of the data is the same as the description of the precipitation data set above including flags for infilled verses real data. The R code to read in a single year is the same as the sample file single.year.R for precip given above. To read the temperature files just change the "ppt" part of the file name to either "tmin" or "tmax".

## Getting a particular station

In UNIX:

```
grep 010148 tmax.complete* > first.station.tmax.data
wc first.station.data
     103    1442    10506


grep 010148 tmin.complete* > first.station.tmin.data
wc first.station.tmin.data
 103    1442    10506
```

The alphanumeric order of the files will insure that all the temps are in the right time order.

## Reading data into R

First read in the metadata file to help in your choice of station.

```
scan("METAinfo", skip=1, what=list( "a", 1,1,1))-> look
names( look)<-c("station.id", "elev", "lon", "lat")
```

As an example find a subset that covers Colorado with a bit extra :

```
ind<- look$lon<  -101 &  look$lon> -109.5
ind<- ind&look$lat<41.5 &  look$lat>36.5
# check the results
library( fields)
US()
points( look$lon[ind], look$lat[ind])
# colorado station id's
colo.id<- look$station.id[ind]
```

Source a useful R function: get.station.R

To extract the first member of CO subset, just the real data:

```
look.tmax<- get.station( colo.id[1], with.infill=FALSE,type="tmax")
look.tmin<- get.station( colo.id[1], with.infill=FALSE,type="tmin")
```

Both of these output data sets will be 103X12 matrices with a missing value code (NA) where observations were not taken. See the R processing script for a "soup to nuts" example of creating a R data sets from these files.

<div align="center">

UCAR : NCAR : CISL : IMAGe : [ CMG   DAReS   GSP   RISC   TNT   ]
For problems with the page, contact Tim Hoar - thoar 'at' ucar 'dot' edu

</div>