United States Department of Agriculture
Natural Resources Conservation Service

How To Choose And Assimilate Data For Hydrological Prediction

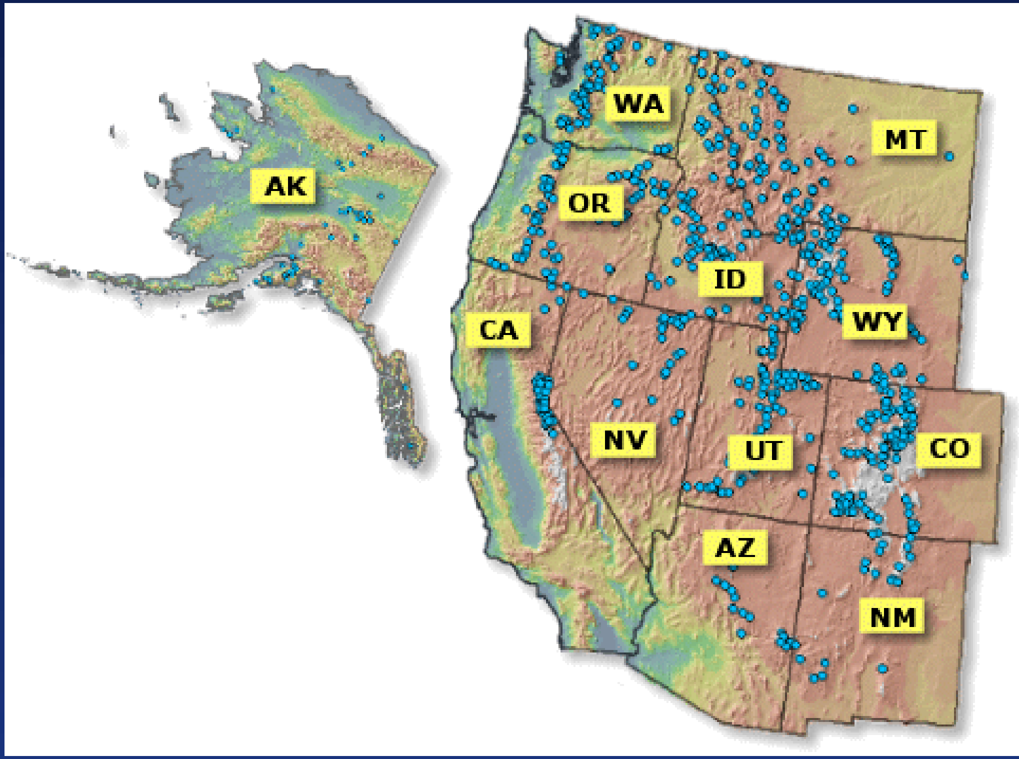
David Garen

United States Department of Agriculture
Natural Resources Conservation Service
National Water and Climate Center
Portland, Oregon, USA

Presented at
IAHS - Predictions in Ungauged Basins Workshop
Canmore, Alberta, Canada
May 2011

Snow Survey and Water Supply Forecasting Program

- Data collection
- Water supply forecasts
- Climate services



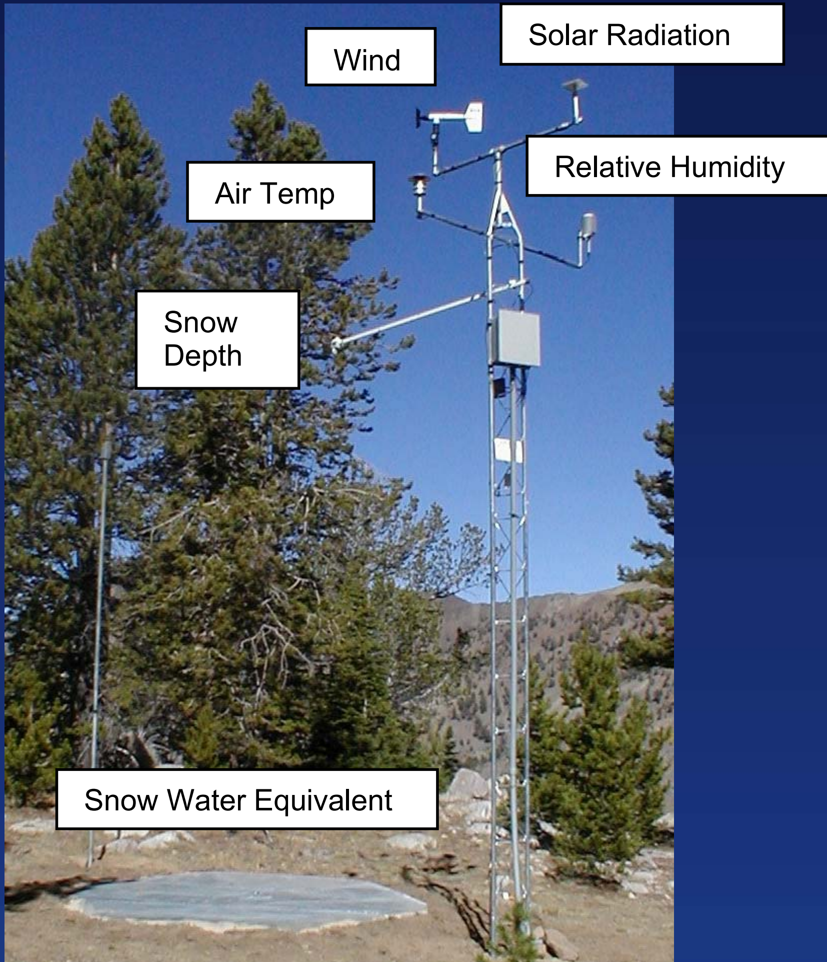
United States Department of Agriculture
Natural Resources Conservation Service

USDA NRCS

SNOTEL Network

Currently over 800 sites
in 13 western states

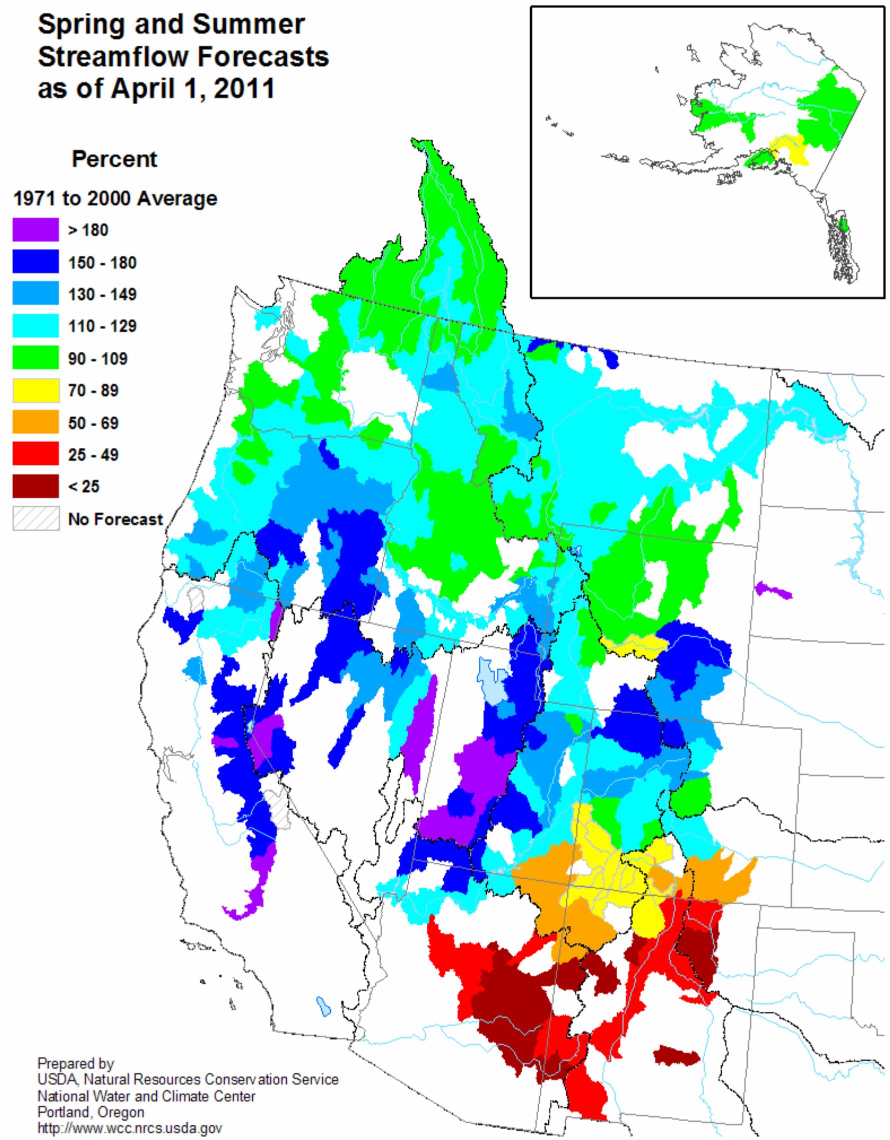
http://www.wcc.nrcs.usda.gov/snow



United States Department of Agriculture
Natural Resources Conservation Service

Water Supply Forecasting

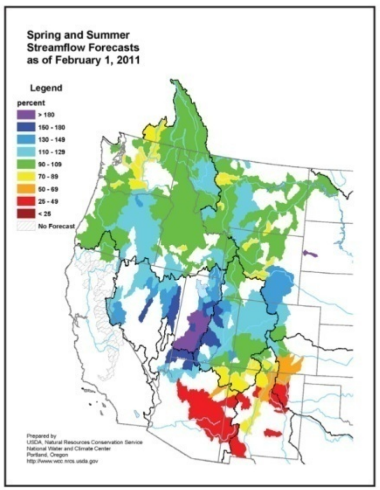
- Seasonal streamflow volume
- Published January through June
- Cooperative effort with National Weather Service
- Over 700 forecast points in western US



United States Department of Agriculture
Natural Resources Conservation Service

Part 622 Snow Survey and Water Supply Forecasting
National Engineering Handbook

Chapter 7 Water Supply Forecasting

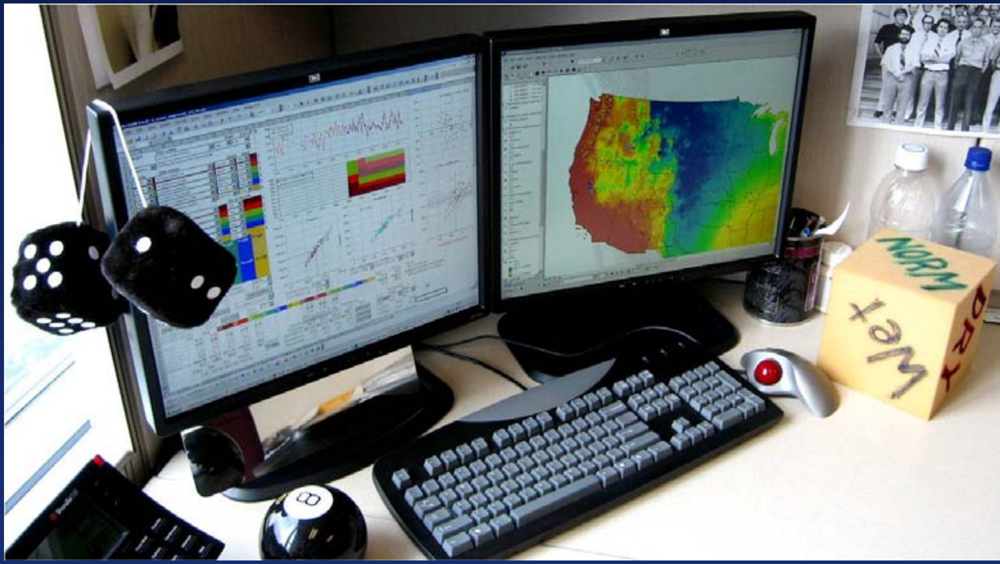


(210–VI–NEH, March 2011)

Where is this?

Different Kinds of Prediction

- Simulating the hydrograph (e.g., comparing accuracies of different models)
- Estimating changes in the hydrograph due to watershed or climate changes
- Real-time streamflow forecasting



NRCS "enhanced" forecasting workstation

Station Data Needs for Different Model Types

Statistical models:

Meteorological station data only need to be good indices of the target. Absolute magnitudes of measured quantities do not have to be correct -- they just need to have a consistent relationship with the target.

Simulation models:

Meteorological station data need to measure quantities accurately in absolute amounts.

Station Data Usage for Different Model Types

<u>Statistical models:</u>

Meteorological station data used can be optimized by predictor variable search algorithms. All stations are not necessarily used.

<u>Simulation models:</u>

All meteorological station data are usually used to define forcing fields. Only "anomalous" stations that have large local influences and are not spatially representative would be omitted.

Station Data Availability for Different Model Applications

<u>Simulation and impact assessment (i.e., research mode):</u>

Meteorological data can be from any stations available in the historical record. They may or may not be currently operated, and the data may or may not be available in real-time.

<u>Real-time forecasting (i.e., operational mode):</u>

All meteorological data must be from currently operated stations and be available in real-time.

Data Processing Considerations

In a research mode, much time and effort can be afforded to process input data -- retrieval, quality control, formatting, pre-processing, etc. -- and it can be done manually without a lot of automation.

In an operational mode, most of this processing needs to be automated so that it can be done in a timely manner. This requires significant database and software infrastructure as well as "intelligent" algorithms to perform these tasks accurately with minimal human intervention.

Data Availability Considerations

Poor coverage of meteorological stations

→ Uncertain estimates of inputs

→ Only simple models can be used

→ Uncertain estimates of outputs

→ Only simple questions can be answered

Good coverage of meteorological stations

→ Opposite of above

United States Department of Agriculture
Natural Resources Conservation Service

Quality of Output Depends on Quality of Input

The first prerequisite for modelling anything is to make the best estimate possible of the system forcings. If the forcings cannot be estimated reasonably well, even the best model cannot produce good predictions.

The starting point, then, is to devise techniques to attempt to make the best estimates possible of meteorological forcings in a watershed.

Spatial Interpolation

To create model forcings, we are inevitably faced with a spatial interpolation task: How do we generalize meteorological station data collected at a point scale to the spatial domain of a watershed?

Various algorithms exist to do this -- I will show some examples of my attempts.

Detrended Kriging Concepts

- Goal is to distribute meteorological station data values over a gridded domain.
- Variability can be divided into vertical and horizontal components.
- Elevation is the primary deterministic factor and defines the vertical variability.
- Horizontal variability is described by ordinary kriging.

Detrended Kriging Assumptions

- Region has relatively homogeneous precip / temp regime.
- Station density adequately represents precip / temp regime (horizontally and vertically).
- Distances are short enough so as not to reach a sill in the variogram, thus allowing the use of a linear variogram, which produces time-invariant station weights.
- Most appropriate for mesoscale regions, ~ 10^2-10^4 km^2.

Detrended Kriging Algorithm

Compute kriging station weights for each grid cell --
Linear semivariogram, time-invariant weights

→

Compute linear precip / temp vs. elevation trend using station data

↓

Compute station residuals

↓

Interpolate residual field with ordinary kriging

↓

Compute elevation trend field based on DEM and linear trend

↓

Add residual field to trend field to obtain final precip / temp field

↓

Repeat for each time step (→ back to first step)

Selecting Stations

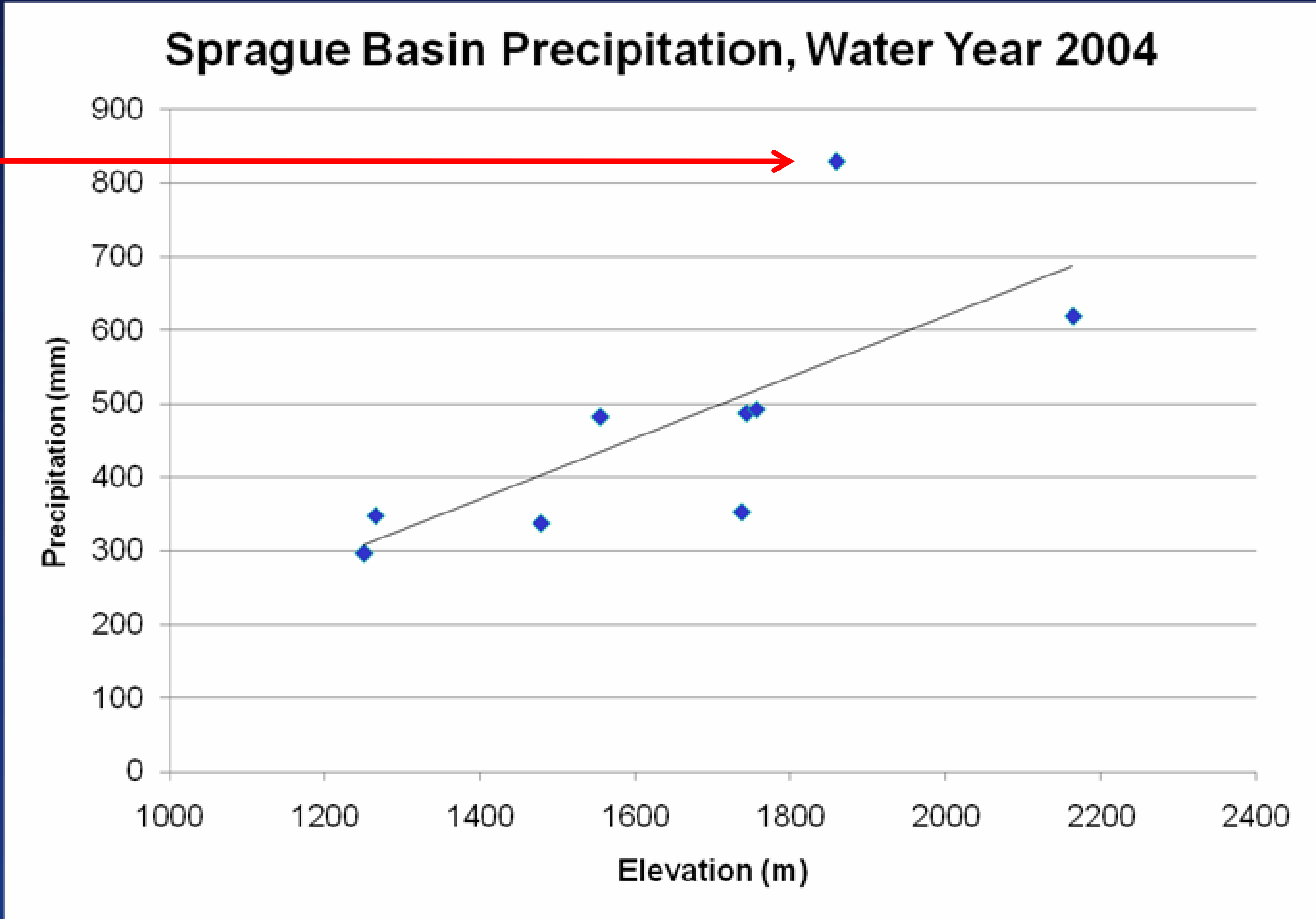
Care must be taken to select stations that represent the general precipitation or temperature fields and are not strongly affected by local influences that are not spatially representative.

This can be difficult to assess, especially with a sparse network. It can be difficult to know if a station with consistently high or low values represents a region around it or whether it is subject to very localized orographic or air drainage influences.

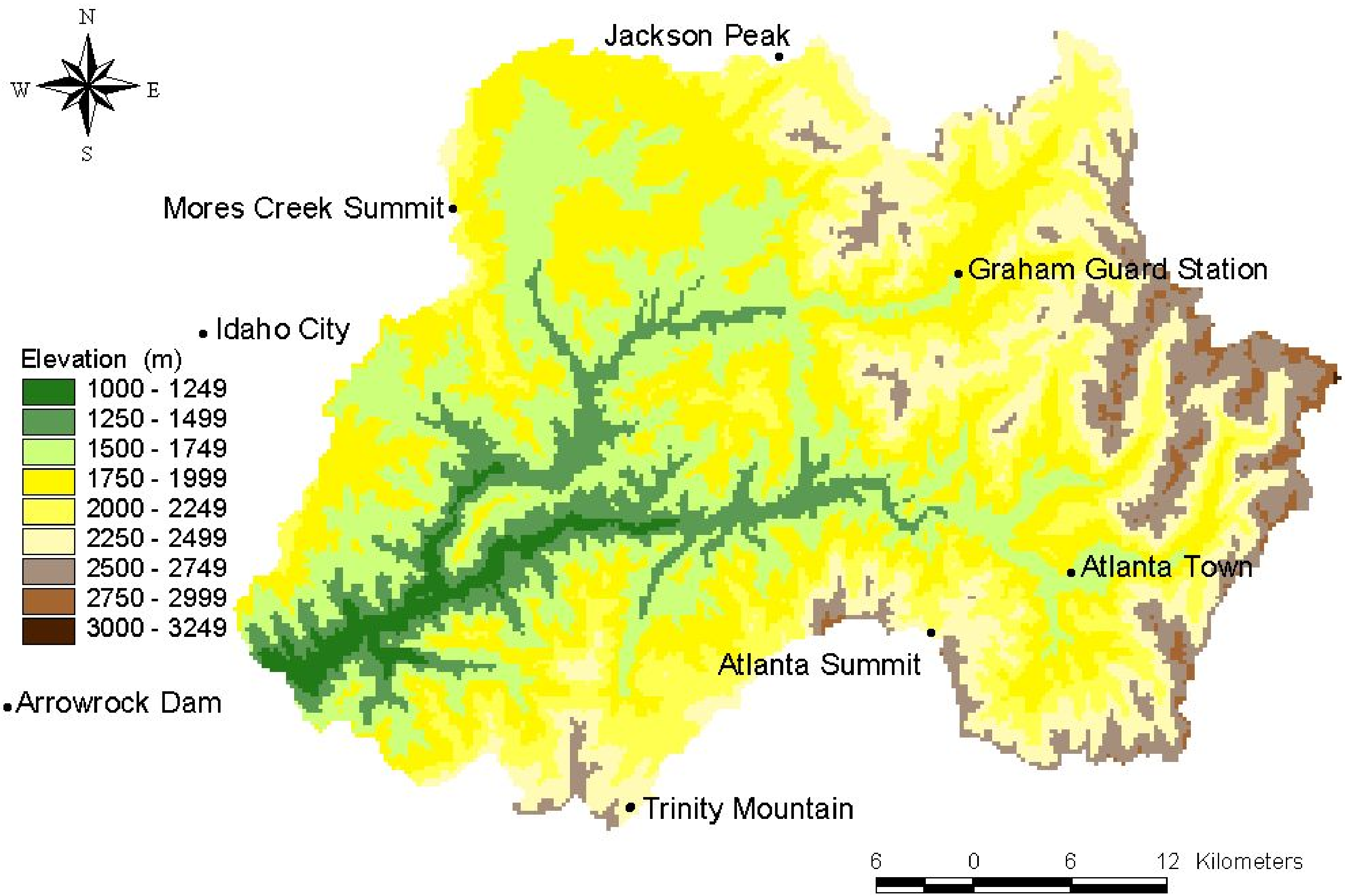
United States Department of Agriculture
Natural Resources Conservation Service

Selecting Stations

Would you want to use this station?



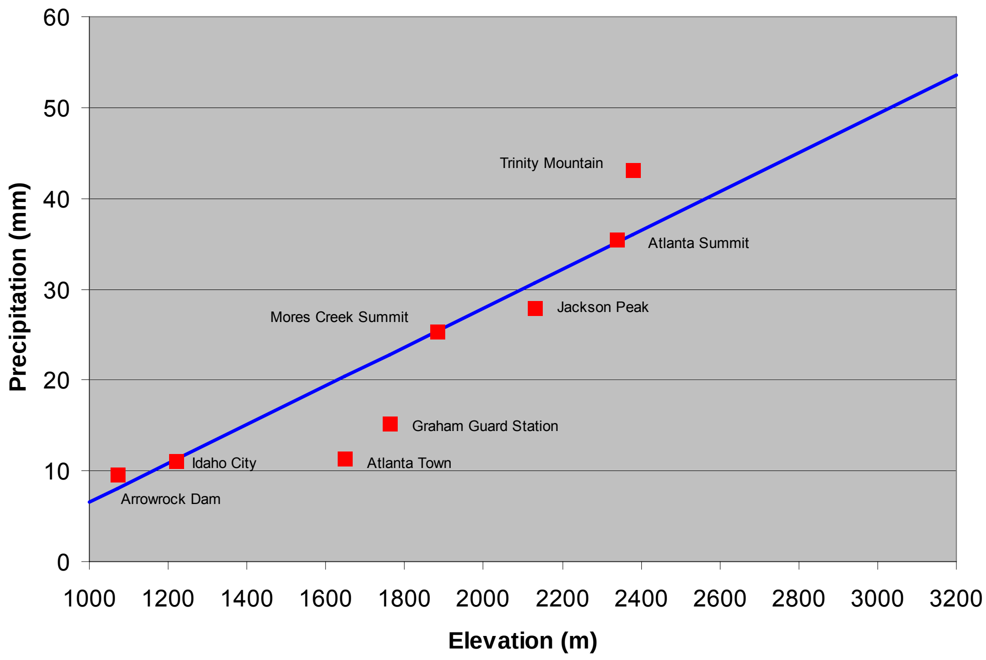
Boise River Elevation and Data Site Locations



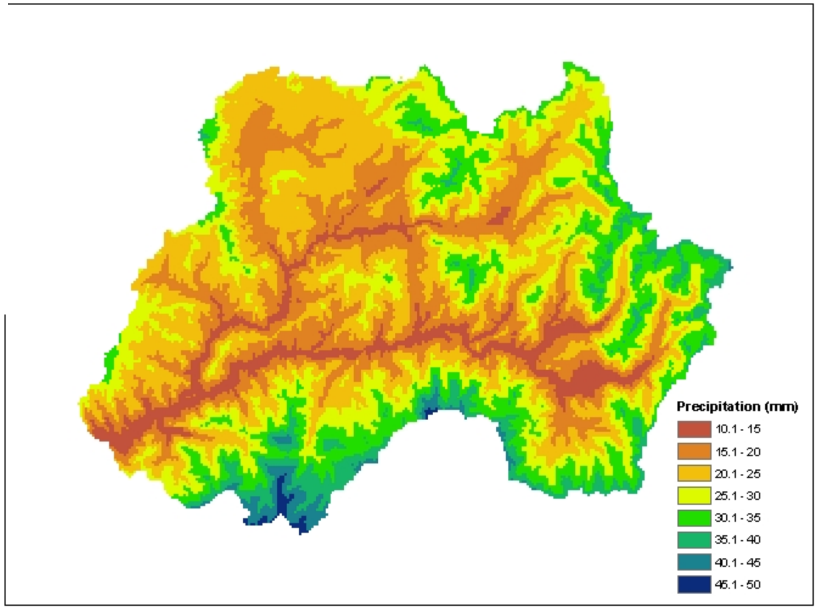
United States Department of Agriculture
Natural Resources Conservation Service

USDA NRCS

Boise River Precipitation, 21 Feb 1998



Precipitation-elevation trend



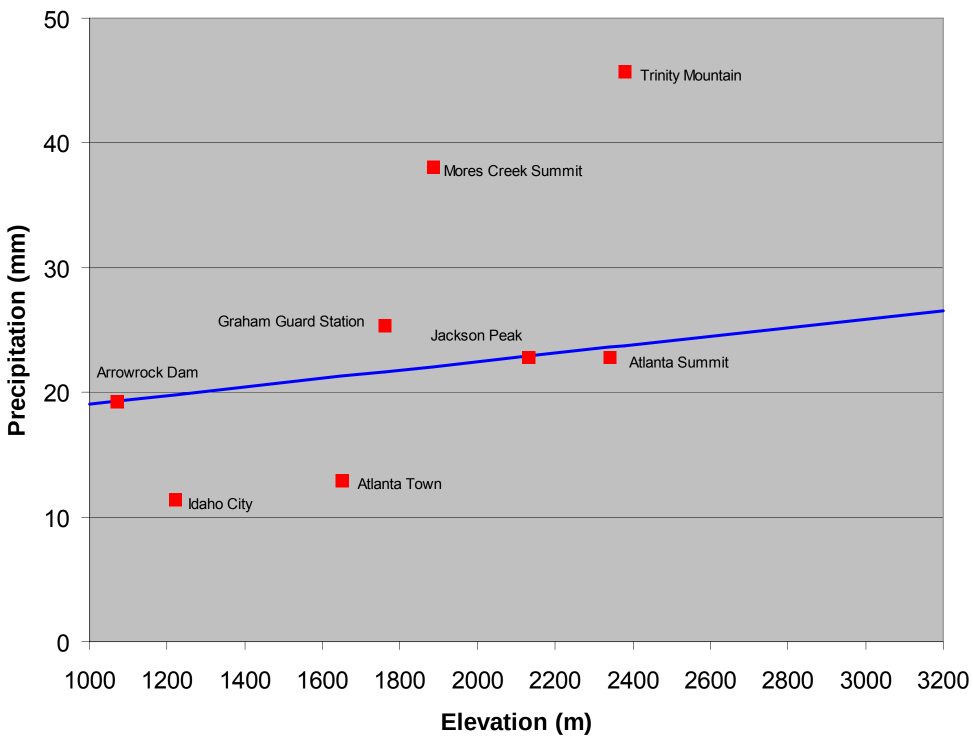
Spatial field

United States Department of Agriculture
Natural Resources Conservation Service

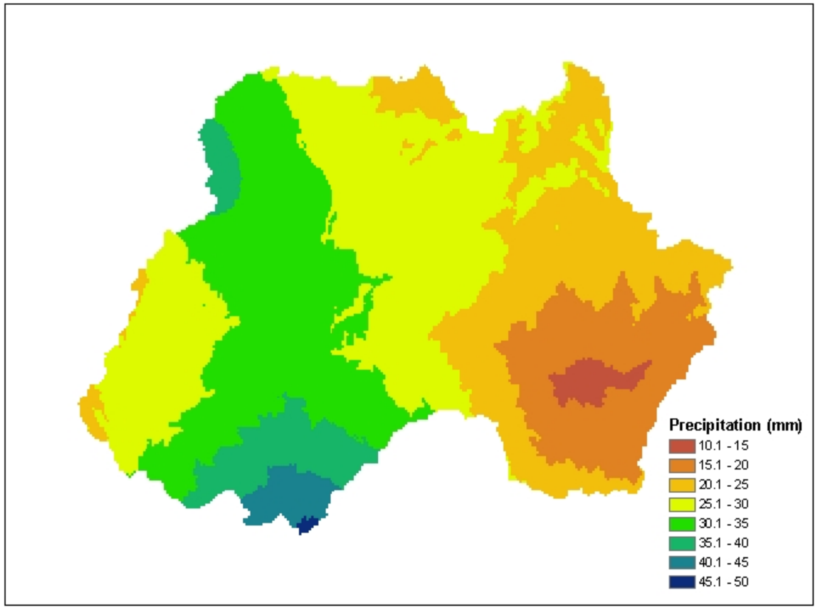
USDA NRCS

Boise River Precipitation, 22 Mar 1998

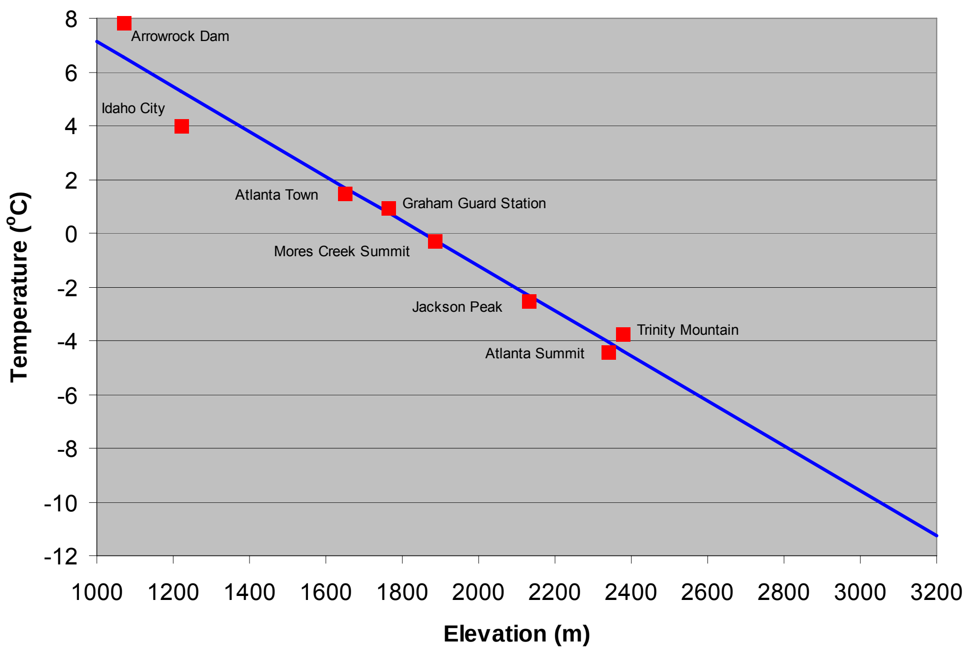
Precipitation-elevation trend



Spatial field

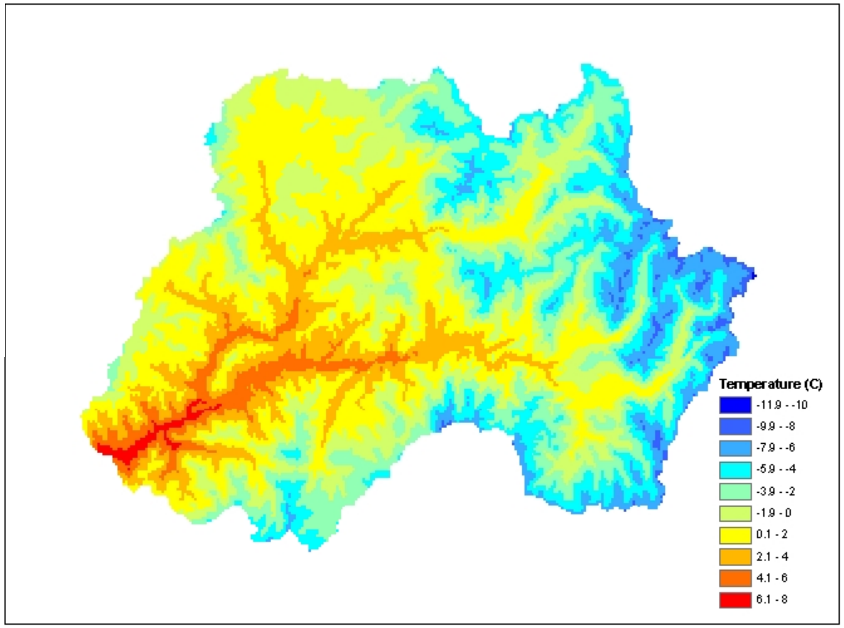


Boise River Temperature
21 Feb 1998, 1200-1500



Temperature-elevation trend

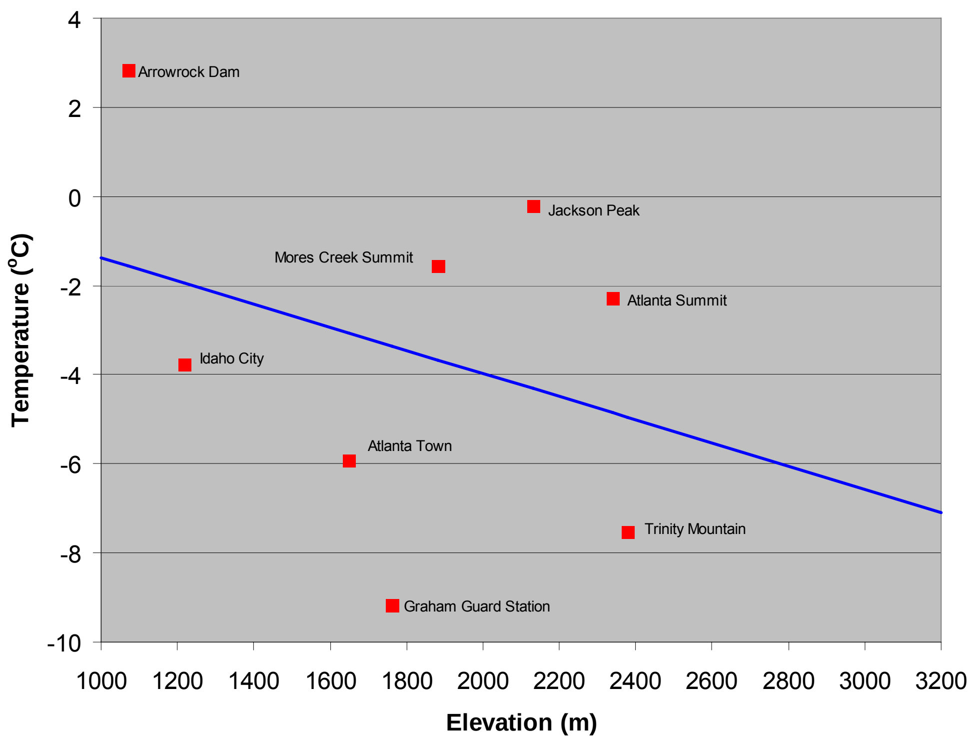
Spatial field



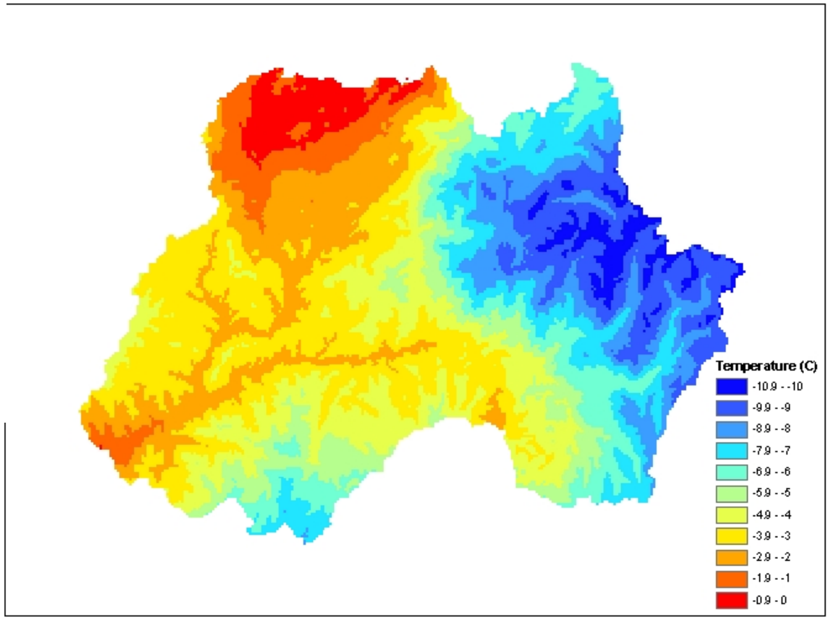
United States Department of Agriculture
Natural Resources Conservation Service

USDA NRCS

Boise River Temperature 14 Mar 1998, 0300-0600



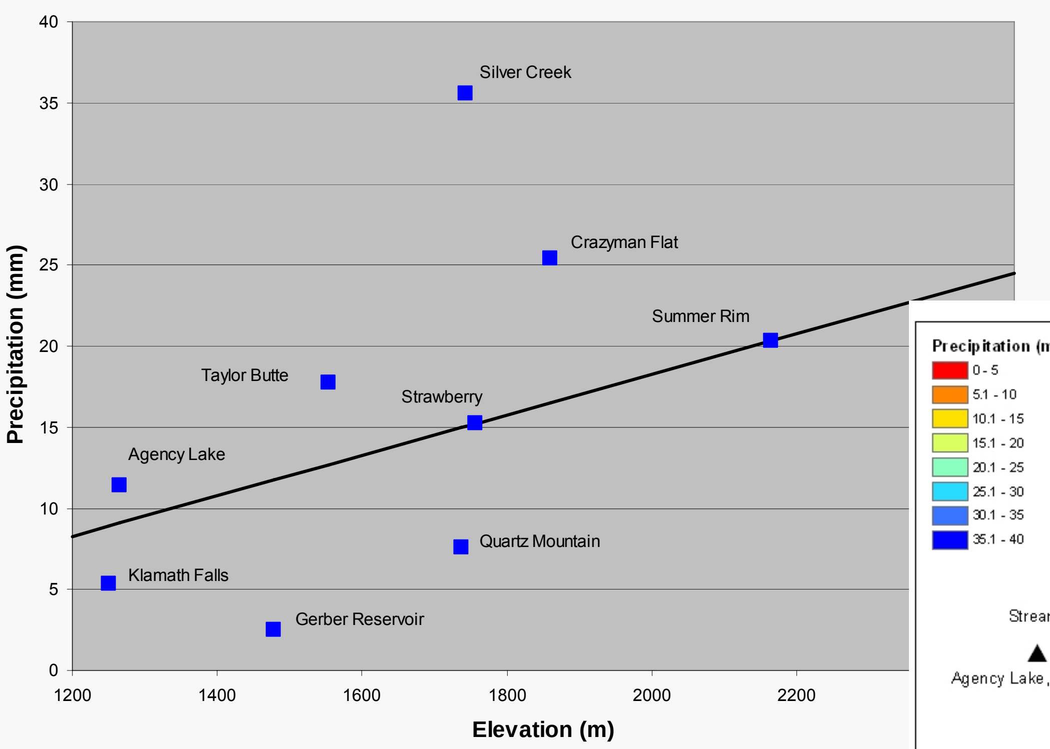
Temperature-elevation trend



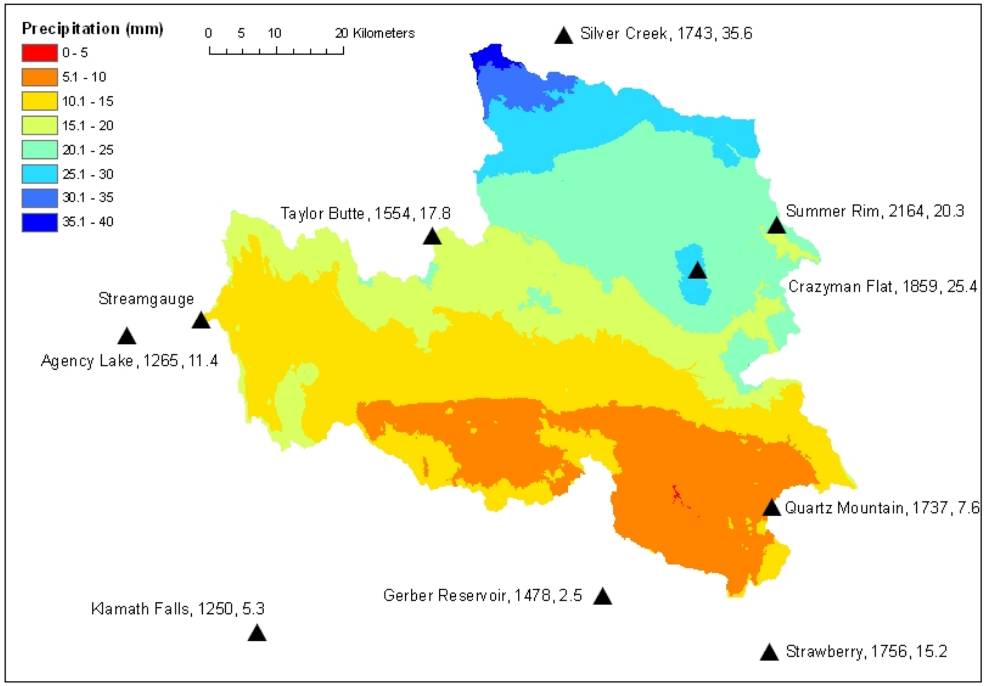
Spatial field

United States Department of Agriculture
Natural Resources Conservation Service

Sprague River Precipitation, 1 Jan 2004



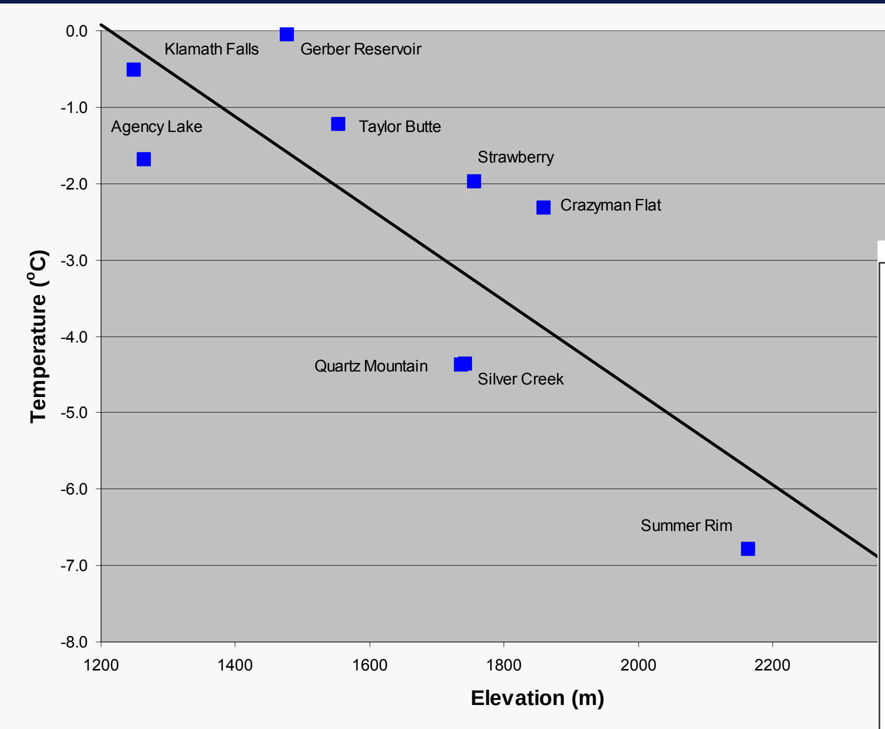
Precipitation-elevation trend



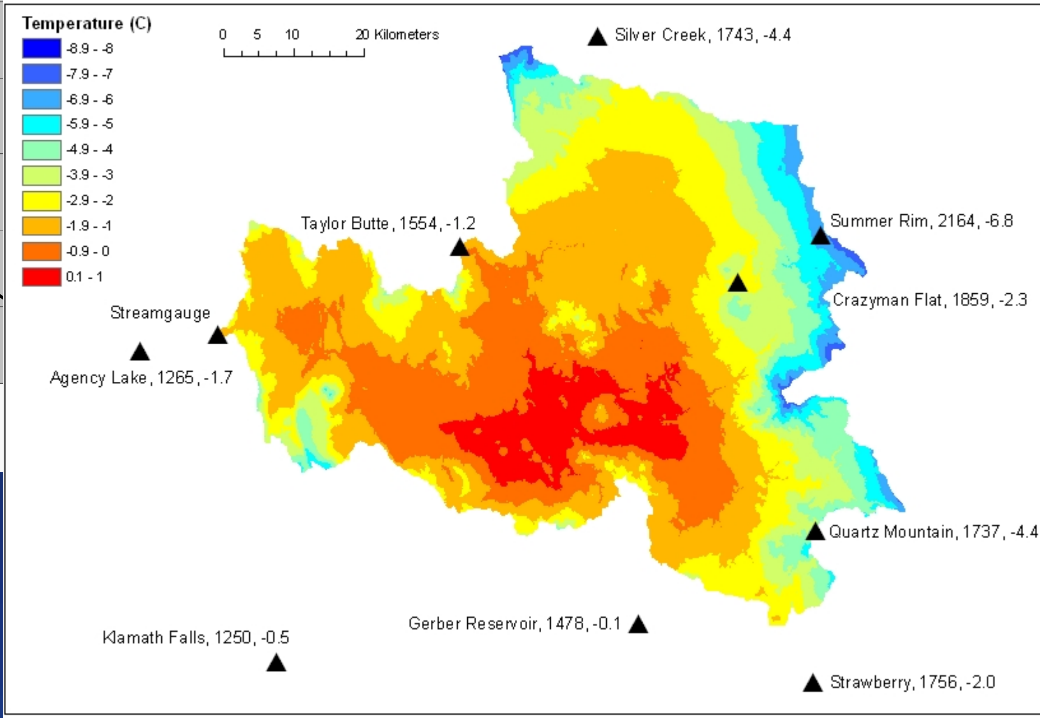
Spatial field

United States Department of Agriculture
Natural Resources Conservation Service

Sprague River Temperature 1 Jan 2004, 1200-1500



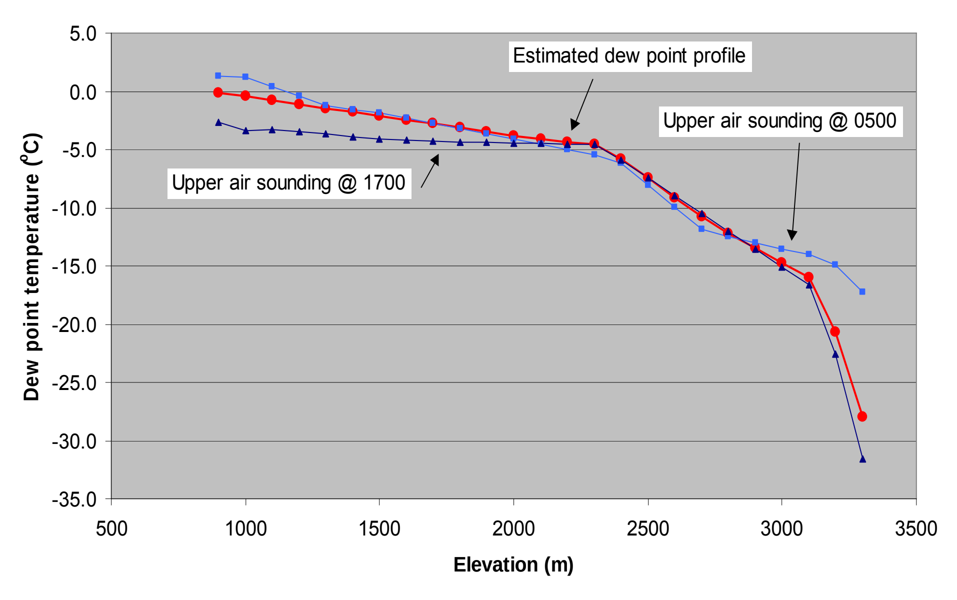
Temperature-elevation trend



Spatial field

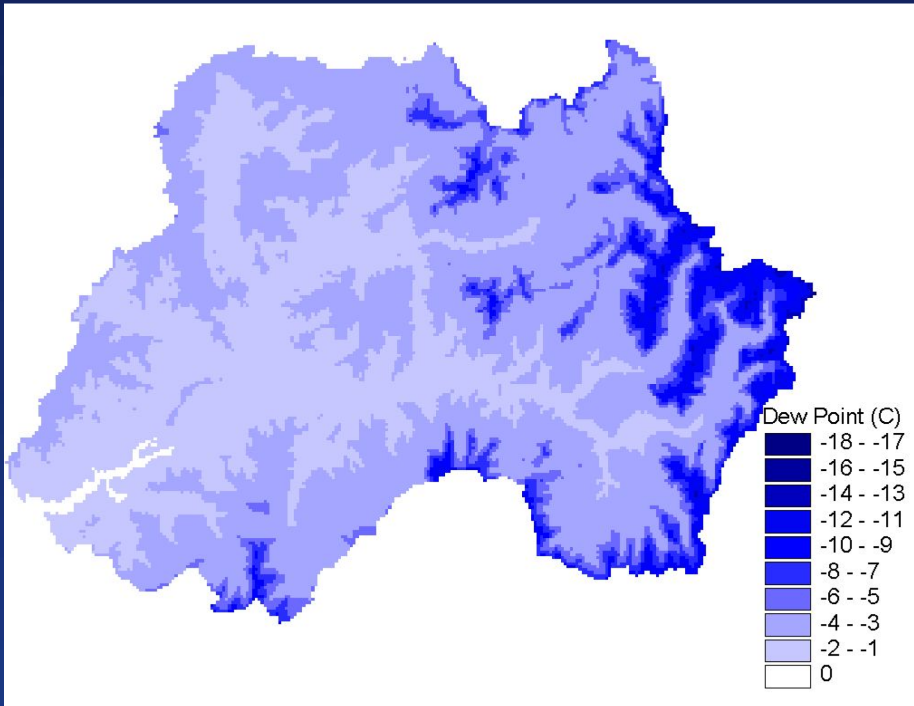
United States Department of Agriculture
Natural Resources Conservation Service

Boise River Dew Point Temperature 20 Apr 1998, 1200-1500

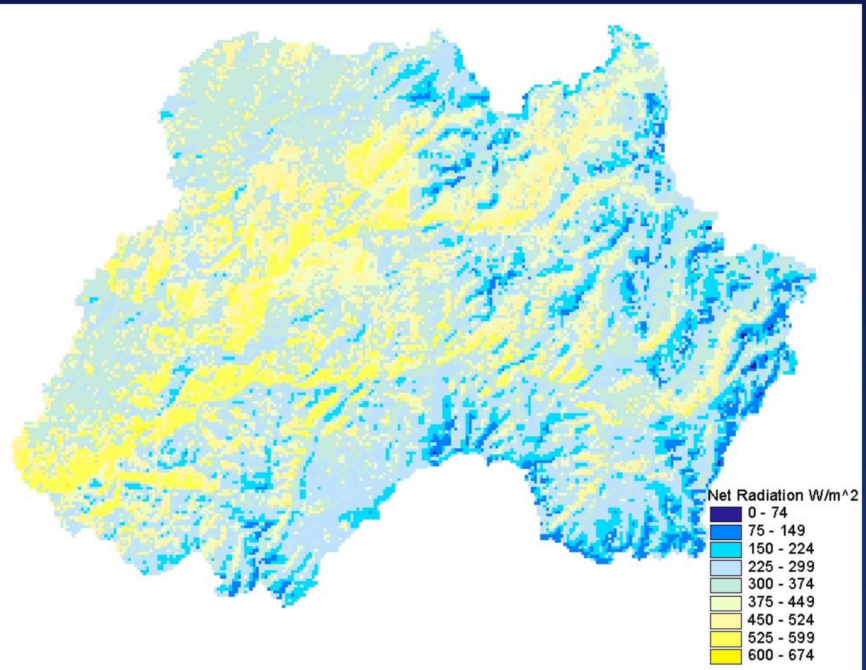


Dew point-elevation profile

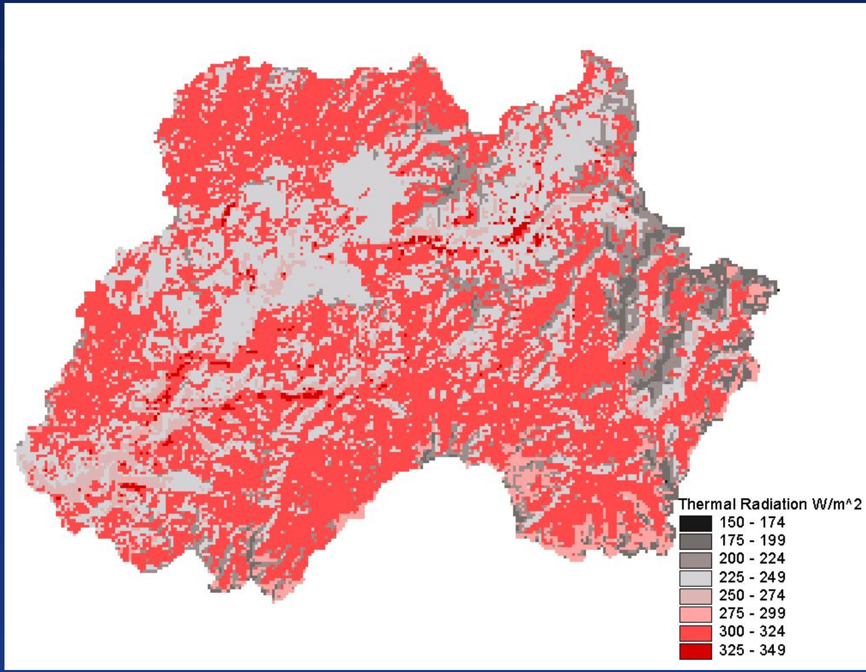
Spatial field



Boise River Solar and Thermal Radiation 20 Apr 1998, 0900-1200



Net solar radiation



Thermal radiation

Model Spatial Units

Once spatial field time series of forcing data have been produced, the forcings can be used as-is (for a fully distributed model) or they can be spatially aggregated to the scale of whatever spatial units the hydrologic model uses (for a semi-distributed or lumped model).

Snowpack Modelling

The Boise River data shown were part of a simulation of snowpack using a spatial energy budget model by Danny Marks (isnobal):

D. C. Garen and D. Marks (2005). Spatially distributed energy balance modelling in a mountainous river basin: estimation of meteorological inputs and verification of model results. Journal of Hydrology, 315:126-153.

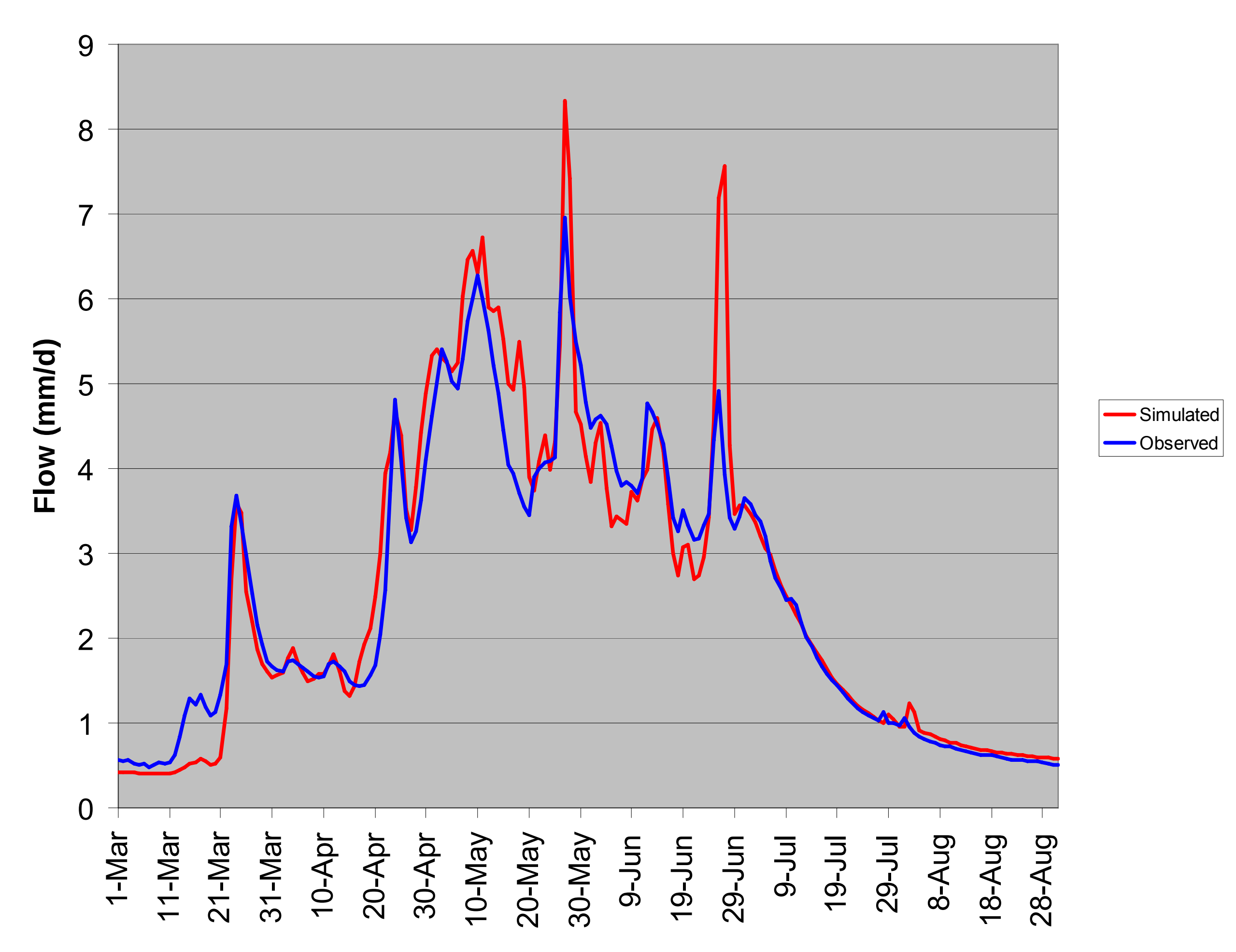
Snowpack Modelling -- Forcing and Snowpack Parameters

Snowpack is a relatively well-defined system with well-defined physics. But, even so, there were still parameters to estimate or assume having to do with some of the forcings (e.g., solar radiation, albedo, forest canopy effects) or with snowpack characteristics (e.g., grain size, liquid water holding capacity).

Snowmelt Output Used as Input to a Watershed Model

Spatial snowmelt fields from the snow model were used as moisture input to a spatially distributed water balance model for the Boise River. Many of the model parameters were calculated directly from spatial layers of elevation, vegetation, and soil texture. Only a few calibration parameters remained, those representing the behavior of conceptual moisture storages.

Example Results -- Boise River, 1998



Data Issues

- Meteorological station data can be spatially interpolated, but the adequacy of the result is highly dependent on station density and spatial representativeness.
- Precipitation and temperature are the easiest to interpolate from ground-based station data.
- Other quantities (humidity, wind, radiation) require special techniques involving models, observations from other sources, or assumptions.

More Issues

- Ensuring data quality is a major effort, both for historical and real-time data.
- Input data for physically-based models can be voluminous, and its preparation can be difficult to automate.
- Much overhead and infrastructure is required to establish and maintain an operational modelling and prediction system.

Conclusion

Questions?

Comments?