Howdy folks!

Show him your card!

NRCS Water Supply Forecasting Technology

Tom Perkins, Tom Pagano and David Garen

Montana Hydrology Conference

Great Falls, Montana

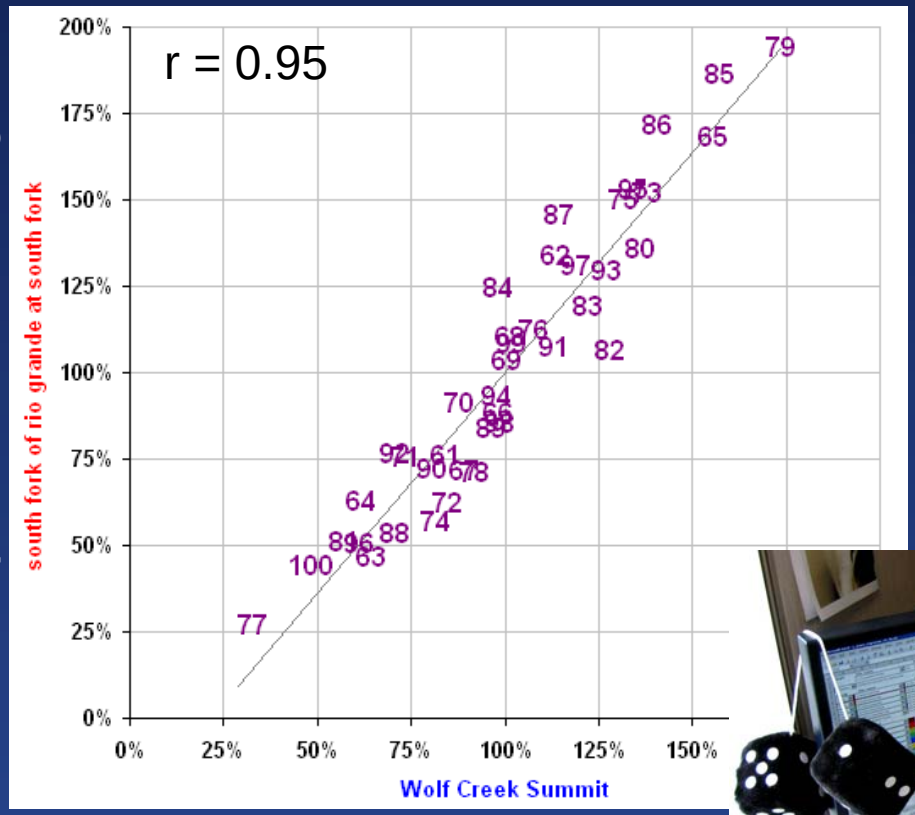
May 27-29, 2008

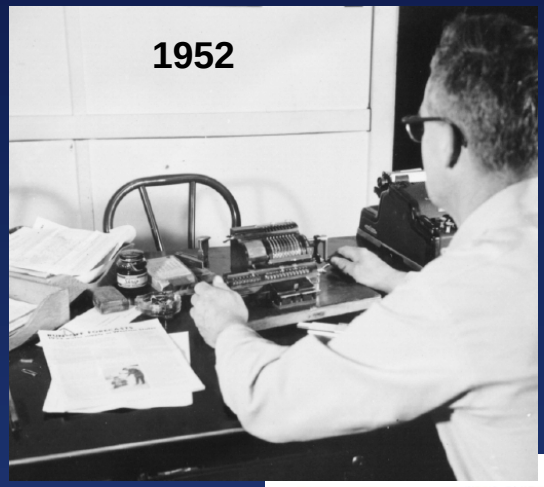
USDA Natural Resources Conservation Service
National Water & Climate Center

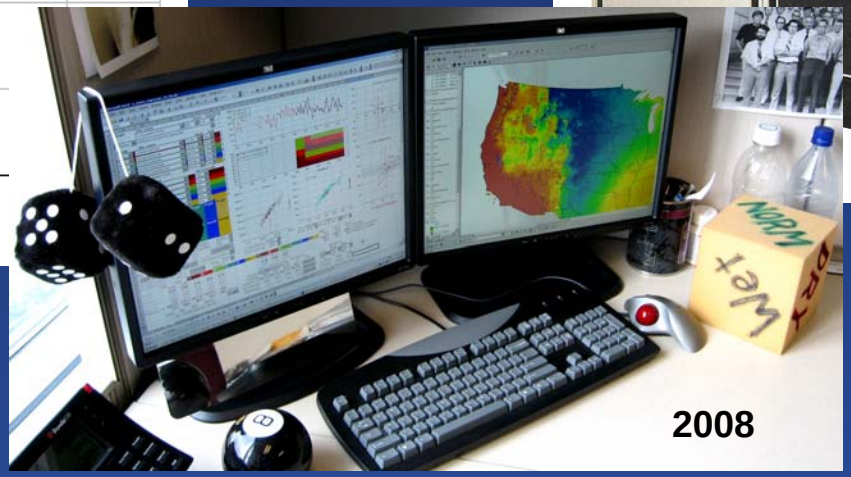
100 YEARS

Current Snowpack

South Fork Rio Grande, Colorado







Philosophy of technology development

VIPER

Daily Update Guidance

Technology Development

Platforms: Microsoft Office - ArcMap - Object Modeling System

Forecasters capable of developing rapid prototypes in VBA

Joint development with contractors to develop operational applications from prototypes (e.g. VIPER)

Dedicated server available for automated products

Visually Interactive Prediction and Estimation Routines (VIPER)

Replacement of legacy forecasting software

Functional prototype 2005 forecast year

Major changes in forecasting capabilities

Operational on January 1, 2008

2-day training workshop in Portland July 2007
Follow-up workshop in Salt Lake City, May 2008

User Guide available

VIPER features

Predictor search routines
Predictor time optimization
Report generation
Exporting to GIS
Saving and managing forecast equations/models
(Configuration management)
Automated data acquisition from web
Data management and editing
Forecast coordination and management
Downstream relational forecasts

Production of final water supply outlook report is only remaining tie to legacy software

The General Linear Regression Model

$$Y = b_0 + \sum_{i=1}^{n} b_i X_i$$

where:

Y = dependent variable

X_i = independent variables

b_i = regression coefficients

n = number of independent variables

The Problem

$$Y = b_0 + \sum_{i=1}^{n} b_i X_i$$

If X's are intercorrelated, they contain redundant information (collinearity), and the b's cannot be meaningfully estimated.

However, we don't want to have to throw out most of the X's but prefer to retain them for robustness.

The Solution

$$Y = b_0 + \sum_{i=1}^{n} b_i X_i$$

Possibilities:

1) Pre-combine X's into composite index(es), e.g., Z-score method

2) Principal components regression

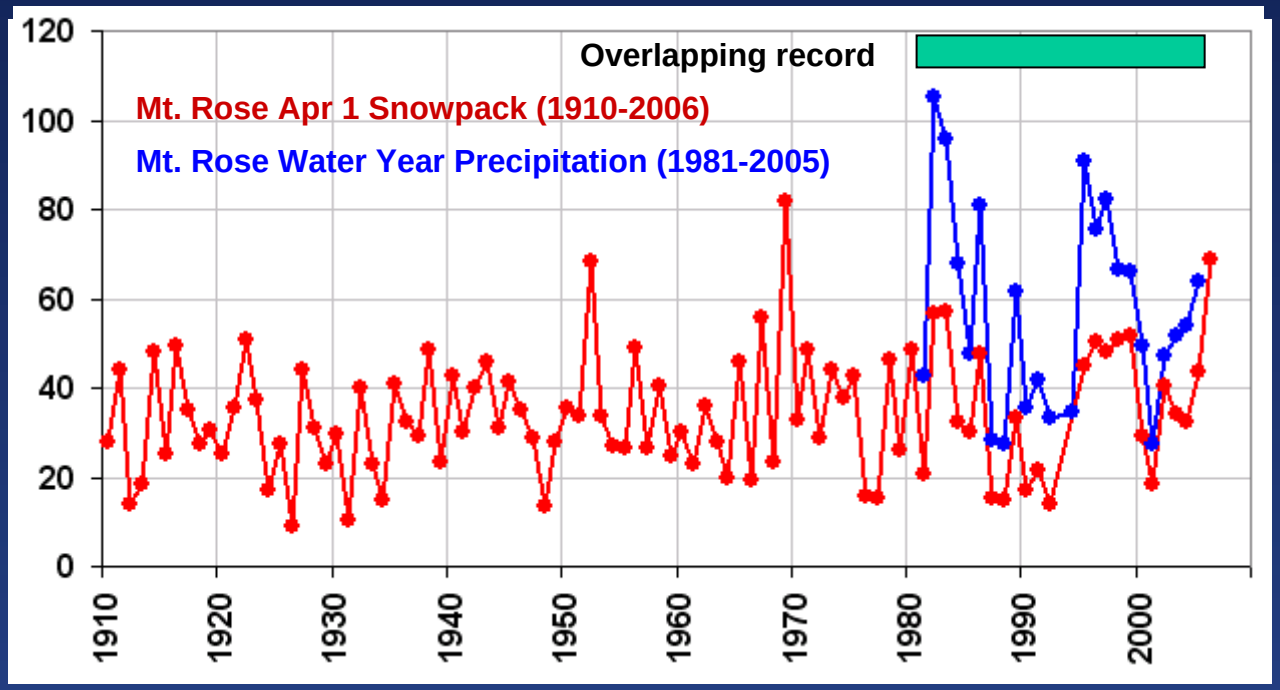
These are similar in concept but differ in the mathematics.

Uneven record lengths

Some stations
have many years

Others have fewer

Typical regression
requires completeness



The choice in this situation has been:
Use fewer stations or use fewer years

Why this is a problem

To use new, younger stations,
older information has to be “forgotten”.

Otherwise, a station must exist
for a long time before becoming useable.

If one piece of data is missing in realtime
then no forecast at all is available,
even if 95% of the “information” is there.

What does z-score regression do?

1. Combines predictors into weighted indices, emphasizing good stations, minimizing bad ones.

2. Compensates for missing data with remaining data.

3. Regresses index against target predictand

What is a z-score?

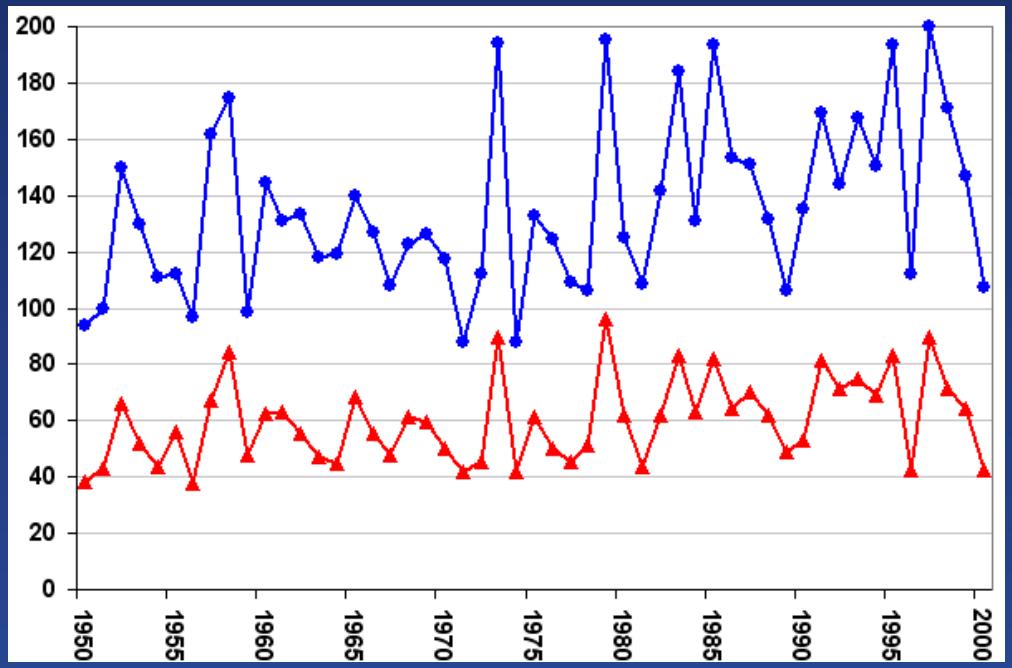
A z-score is a "normalized anomaly":

$$Z = \frac{\text{value - average}}{\text{standard deviation}}$$

What is a z-score?

A z-score is a "normalized anomaly":

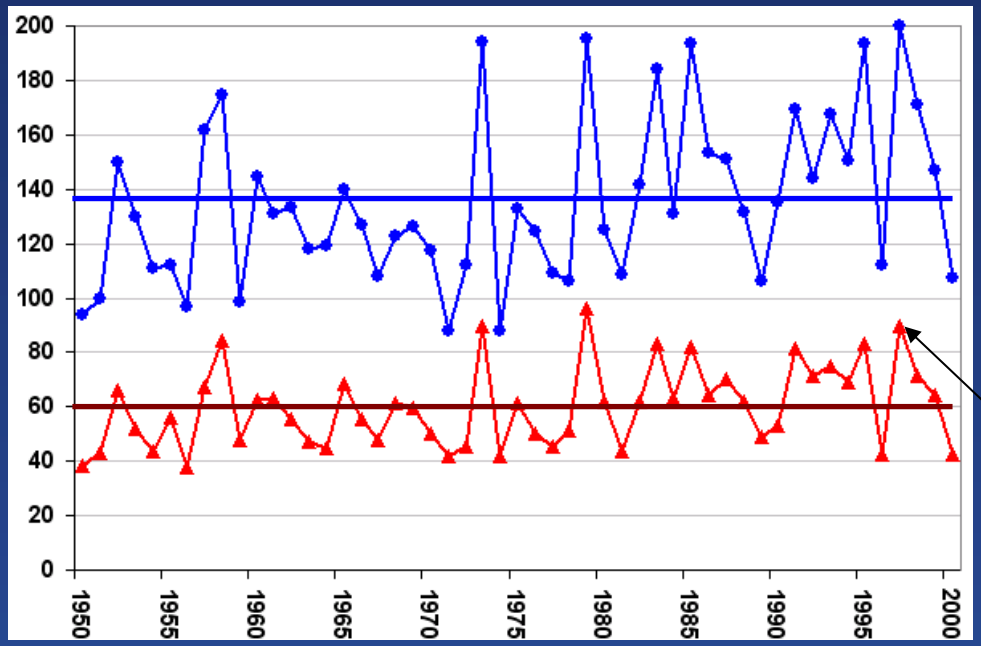
$$Z = \frac{\text{value - average}}{\text{standard deviation}}$$



What is a z-score?

A z-score is a "normalized anomaly":

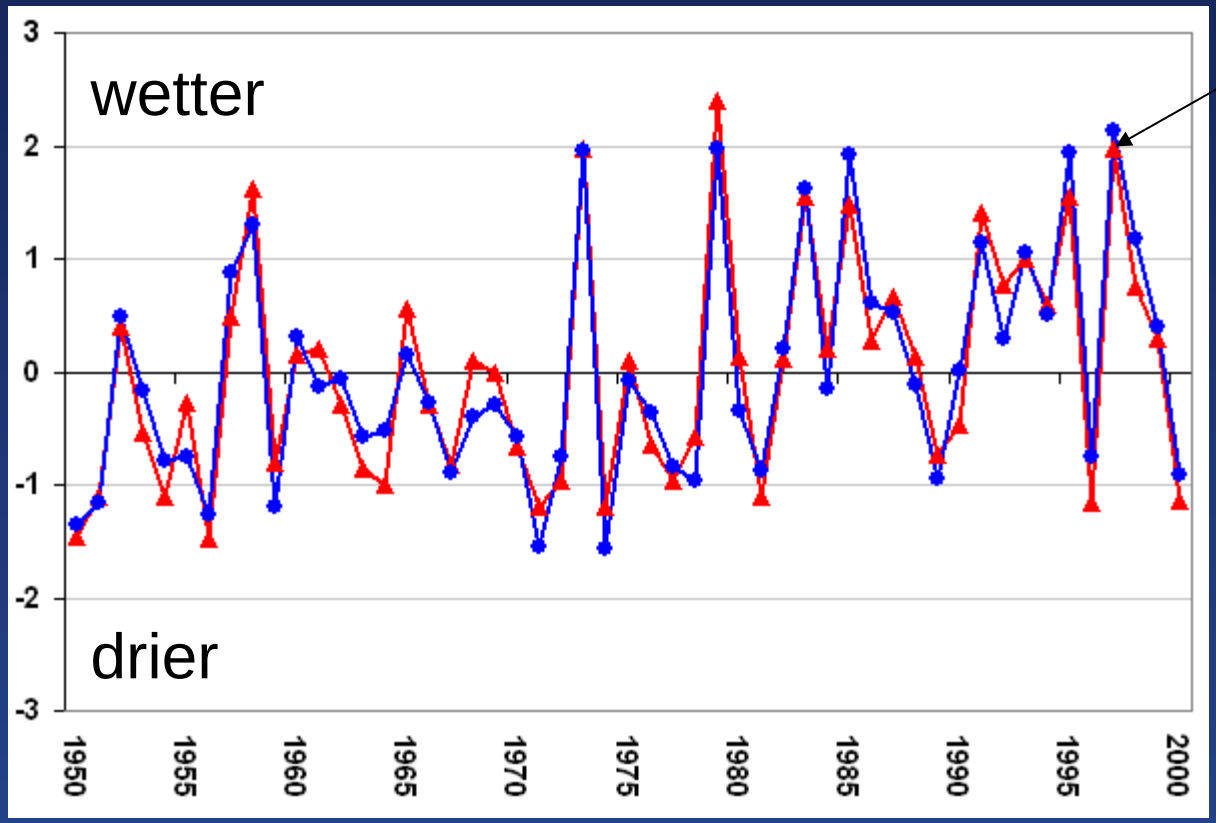
$$Z = \frac{\text{value - average}}{\text{standard deviation}}$$



avg	stdev
135	30
60	15

Z = (90 – 60)/15 = +2

What is a z-score?



avg 0

stdev 1

Stations are now on an “even footing”

The use of “groups” (aka components)

In the case of multiple signals,
stations with a like signal (e.g. fall precipitation)
are combined by the user
into their own respective “group index”,
weighted by their combination with flow.

All the group indices are then
combined into a “master index”,
weighted, again, by their correlation with flow.

The master index is regressed against flow.

How good are the results

Under conditions of serially compete data,
and relatively “normal” conditions
PCA and Z-Score are effectively indistinguishable

Skill and behavior is similar to the official published outlooks

However…
Any tool is a weapon
if you hold it right.

Summary

Z-score regression –

A regression methodology that, within reason, can handle uneven record lengths and missing data.

It groups stations into indices, emphasizing good stations, minimizing the effect of poor stations. Multiple signals can be managed (e.g. snow, fall precip, baseflow).

Can be abused especially if the input data set is highly uneven.

Principal Components Analysis

Principal components regression is just like standard regression except the independent variables are principal components rather than the original X variables.

Principal components are linear combinations of the X's.

Principal Components Analysis

Principal components are calculated from eigenvectors that are derived from a matrix equation whose input is the correlation matrix of all the X's with each other.

The eigenvectors (weights) are based solely on the intercorrelations among the X's and have no knowledge of Y (in contrast to Z-score, for which the opposite is true).

Principal components are new variables that are not correlated with each other.

Principal Components Analysis -- Example

Independent Variables:

$X_1 – X_5$	Snow water equivalent at 5 stations
$X_6 – X_{10}$	Water year to date precipitation at 5 stations
X_{11}	Antecedent stream flow
X_{12}	Climate teleconnection index

Pacific Decadal Oscillation (PDO)
Southern Oscillation Index (SOI)
Nino 3.4 Sea Surface Temperature Index (NINO34)
Tni Climate Index (TNI)

Correlation Matrix

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	Y
X_1	1.0	.72	.67	.76	.81	.54	.31	.54	.38	.50	.18	.64	.65
X_2		1.0	.67	.45	.80	.62	.45	.47	.31	.49	.14	.39	.60
X_3			1.0	.49	.72	.84	.76	.86	.68	.85	.48	.56	.80
X_4				1.0	.62	.42	.26	.36	.56	.38	.28	.59	.68
X_5					1.0	.62	.49	.51	.44	.62	.32	.59	.73
X_6						1.0	.93	.87	.83	.90	.63	.43	.85
X_7							1.0	.82	.85	.90	.67	.32	.76
X_8								1.0	.74	.84	.64	.39	.70
X_9									1.0	.80	.70	.49	.84
X_{10}										1.0	.64	.46	.79
X_{11}											1.0	.36	.51
X_{12}												1.0	.64

	PC_1	PC_2	PC_3	PC_4	PC_5
X_1	0.265	0.444	0.004	0.074	-0.104
X_2	0.249	0.325	-0.483	-0.030	0.315
X_3	0.335	0.016	-0.178	0.149	-0.314
X_4	0.229	0.353	0.456	-0.595	-0.009
X_5	0.287	0.332	-0.148	0.120	0.412
X_6	0.339	-0.168	-0.162	-0.106	-0.040
X_7	0.308	-0.329	-0.150	-0.058	-0.015
X_8	0.317	-0.197	-0.114	0.027	-0.261
X_9	0.304	-0.240	0.299	-0.313	-0.103
X_{10}	0.330	-0.197	-0.197	0.072	-0.129
X_{11}	0.235	-0.349	0.351	0.168	0.692
X_{12}	0.232	0.262	0.473	0.675	-0.212
% var.	62.7	15.8	7.8	3.8	3.2

First Five Eigenvectors

Eigenvalues

Principal Components Regression Procedure

- Try the PC's in order
- Test for regression coefficient significance (t-test)
- Stop at first insignificant component
- Transform regression coefficients to be in terms of original variables
- Sign test – coefficient signs must be same as correlation with Y

	Type	Target	Start	End
>	Forecast Point	06191500, MT, Yellowstone R At Corwin Springs	Apr	Jul

		Type	Predictors	Start	End	
1	✓	SnotelSwe	878, WY, Younts Peak	Apr F	Apr F	1
2	✓	SnotelSwe	837, WY, Two Ocean Plateau	Apr F	Apr F	2
3	✓	SnotelSwe	816, WY, Thumb Divide	Apr F	Apr F	3
4	✓	SnotelSwe	384, WY, Canyon	Apr F	Apr F	4
5	✓	SnotelSwe	806, WY, Sylvan Lake	Apr F	Apr F	5
6	☐	USGSStrm	06191500, MT, Yellowstone River At Corwin Sp	Oct-1	Mar	6
7	✓	SnotelPrcp	878, WY, Younts Peak	Oct-1 F	Mar L	7
8	✓	SnotelPrcp	837, WY, Two Ocean Plateau	Oct-1 F	Mar L	8
9	☐	SnotelPrcp	816, WY, Thumb Divide	Oct-1 F	Mar L	9
10	✓	SnotelPrcp	384, WY, Canyon	Oct-1 F	Mar L	10
11	✓	SnotelPrcp	806, WY, Sylvan Lake	Oct-1 F	Mar L	11
12	✓	SnotelSwe	683, WY, Parker Peak	Apr F	Apr F	12
13	✓	SnotelSwe	670, MT, Northeast Entrance	Apr F	Apr F	13
14	✓	SnotelSwe	480, MT, Fisher Creek	Apr F	Apr F	14
15	✓	SnotelPrcp	683, WY, Parker Peak	Oct-1 F	Mar L	15
16	✓	SnotelPrcp	670, MT, Northeast Entrance	Oct-1 F	Mar L	16
17	✓	SnotelPrcp	480, MT, Fisher Creek	Oct-1 F	Mar L	17
18	✓	SnotelSwe	862, MT, White Mill	Apr F	Apr F	18

Global month changes: **Instantaneous** May F **Accumulated** Oct-1 F Apr L

Station	1	2	3	4	5	6
Correl	0.798	0.849	0.820	0.867	0.814	
Years	37/37	37/37	37/37	37/37	27/27	
CurrZ	0.289	1.358	0.195	1.192	0.381	
PctNorm	112%	136%	108%	136%	110%	
Pred	1687.49	2084.87	1657.09	2033.26	1687.18	

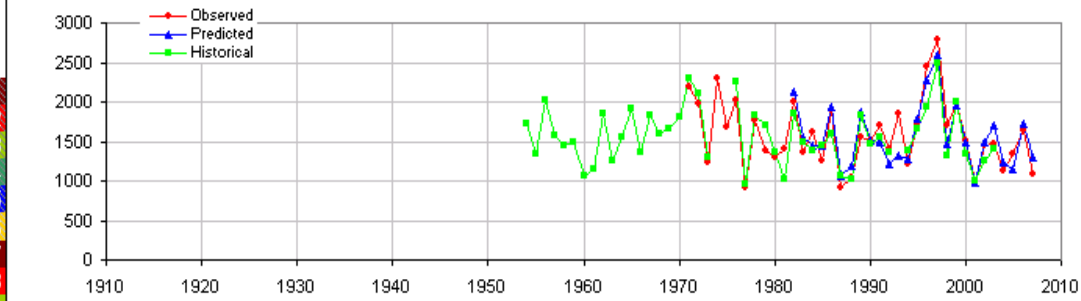
Station	7	8	9	10	11	12
Correl	0.817	0.824		0.837	0.865	0.820
Years	27/27	27/27		37/37	27/27	26/26
CurrZ	0.636	0.743		1.483	1.086	1.627
PctNorm	119%	121%		141%	127%	136%
Pred	1776.19	1815.52		2122.38	1954.51	2137.41

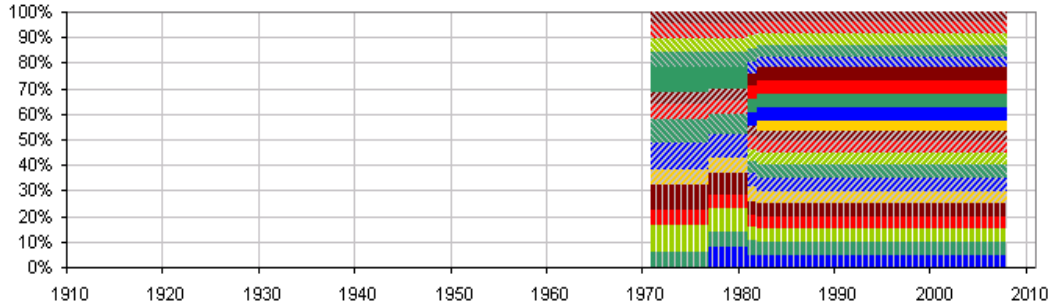
Station	13	14	15	16	17	18
Correl	0.761	0.857	0.786	0.808	0.845	0.876
Years	37/37	37/37	27/27	37/37	37/37	37/37
CurrZ	0.166	0.876	2.299	0.506	0.310	0.726
PctNorm	106%	124%	150%	113%	108%	120%
Pred	1642.60	1911.42	2322.85	1764.05	1701.11	1862.07

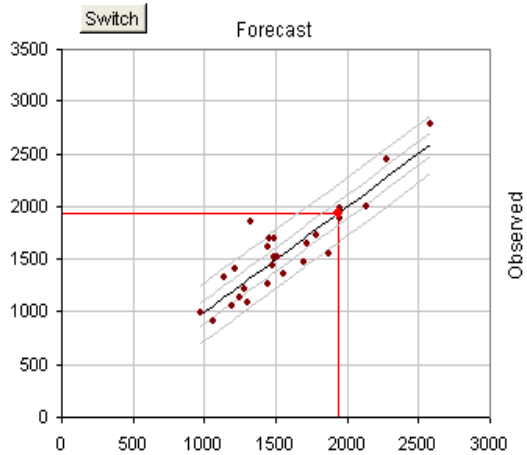
Group	SnotelSwe	SnowSwe	SnotelPrcp	CoopPrcp	USGSStrm	NRCSStrm
Correl	0.895		0.907			
Years	37		37			
CurrZ	0.785		0.780			
Pred	1898.67		1890.40			

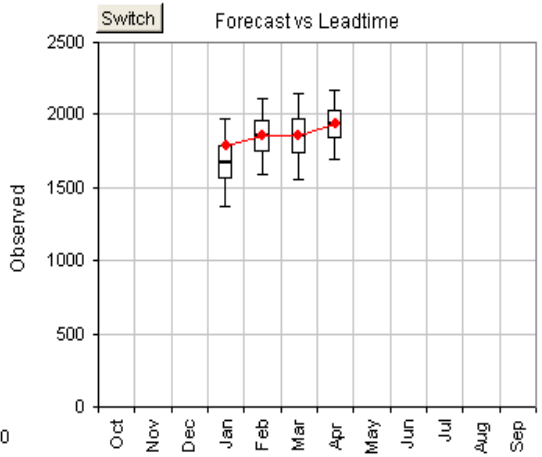
Group	SOI	Reservoir	Routed	BEARStrm	BCMTMprc	All
Correl						0.898
Years						26
CurrZ						0.783
Pred						1937.45194

Statistics	Average	Median	Min	MinYear	Max	MaxYear
POR	1588.1	1520.85	909.43	1977	2777.79	1997
71-00	1654.7	1647.65	909.43	1977	2777.79	1997









Analysis Type	Principal Components

Forecast	Volume	Pct
10	2213.45	134%
30	2048.75	124%
50	1937.45	117%
70	1826.15	111%
90	1661.45	101%
	standard	jackknife
r2	0.807	0.775
StdErr	193.74	209.44
StdErrSS	0.552	0.515

Offl Norm 1650

Transformation	None
First Year Used	1971
Last Year Used	9999
Target Data Src	USGS
Publication Date	April
Published?	✓

More Predictors... (5 active)

Advanced Settings... (none active)

Helper Predictand... (not active)

Clear

Recalculate

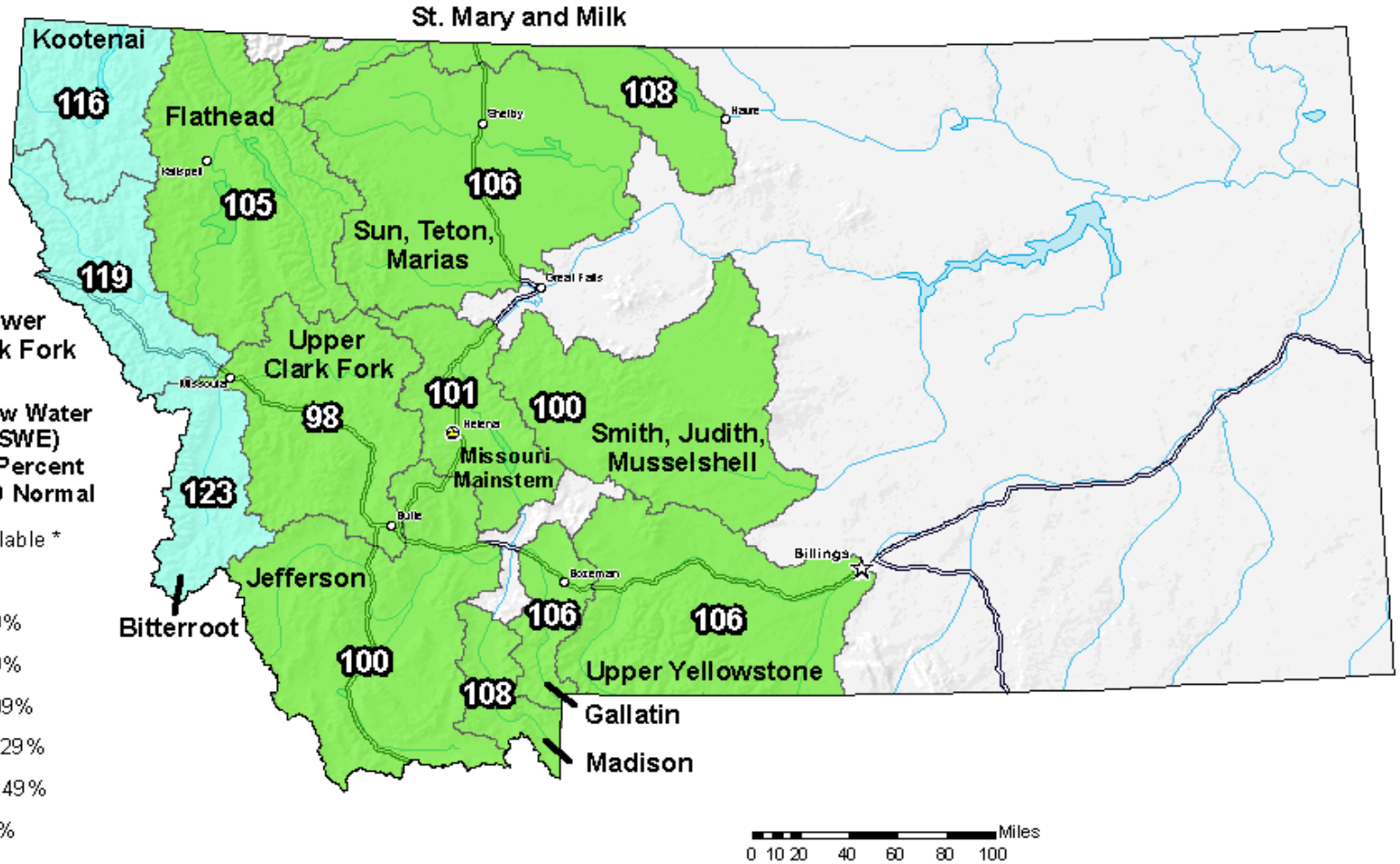
PCA	
# comp	1
% var	0.882

ArcMap
MILE CREEK
SMUGGLER MINE
LOWER TWIN
FOUR MILE
LOWER TWIN
MIDDLE MILE CREEK
6041000
NORRIS MADISON PH
GALLATIN GATEWAY 7 S
6043500
6050000
NEW WORLD
LICK CREEK
LICK CREEK
GALLATIN GATEWAY 10 ESE
6049450
DILLON AIRPORT
DILLON WMCE
VIRGINIA CITY
ARCH FALLS
SHOWER FALLS
SHOWER FALLS
BEAR BASIN
LITTLE PARK
LONE MOUNTAIN
LONE MOUNTAIN
LONE MOUNTAIN
6016000
6019500
CARTER CREEK
DILLON 9 SSE
6015400
CLARK CANYON DAM
CALL ROAD
ALDER 17 S
EAGLEHEAD
ROCK CREEK MEADOW
CLOVER MEADOW
CLOVER MEADOW
CLOVER MEADOW
SHORT CREEK
SHORT CREEK
6191500
NOTCH
CARROT BASIN
CARROT BASIN
BEAVER CREEK
BEAVER CREEK
POTOMAGETON PARK
DIVIDE
DIVIDE
DIVIDE
LAKE CREEK
HEBGEN DAM
6038500
HEBGEN DAM
TWENTY-ONE MILE
TEPEE CREEK
TEPEE CREEK
TEPEE CREEK
WHITE PINE RIDGE
LIME
6012500
WEST YELLOWSTONE 0 NNW
MOUNT HOLMES LOOKOUT
NORRIS BASIN
LAKEVIEW RIDGE
LAKEVIEW RIDGE
LAKEVIEW RIDGE
LAKEVIEW CANYON
VALLEY VIEW
WEST YELLOWSTONE
WEST YELLOWSTONE
WEST YELLOWSTONE
WEST YELLOWSTONE
IRVING CREEK
WHISKEY CREEK
WHISKEY CREEK
WHISKEY CREEK
MADISON PLATEAU
MADISON PLATEAU
CAMP CREEK
WHITE ELEPHANT
WHITE ELEPHANT
WEBBER CREEK
CRAB CREEK
CRAB CREEK
KILGORE
BIG SPRINGS
LUCKY DOG
LATHAM SPRINGS
BLACK BEAR
ISLAND PARK
ISLAND PARK
ISLAND PARK
OLD FAITHFUL
OLD FAITHFUL

ArcGIS Explorer

Montana SNOTEL Current Snow Water Equivalent (SWE) % of Normal

Apr 01, 2008



Current Snow Water Equivalent (SWE) Basin-wide Percent of 1971-2000 Normal

- unavailable *
- <50%
- 50 - 69%
- 70 - 89%
- 90 - 109%
- 110 - 129%
- 130 - 149%
- >=150%

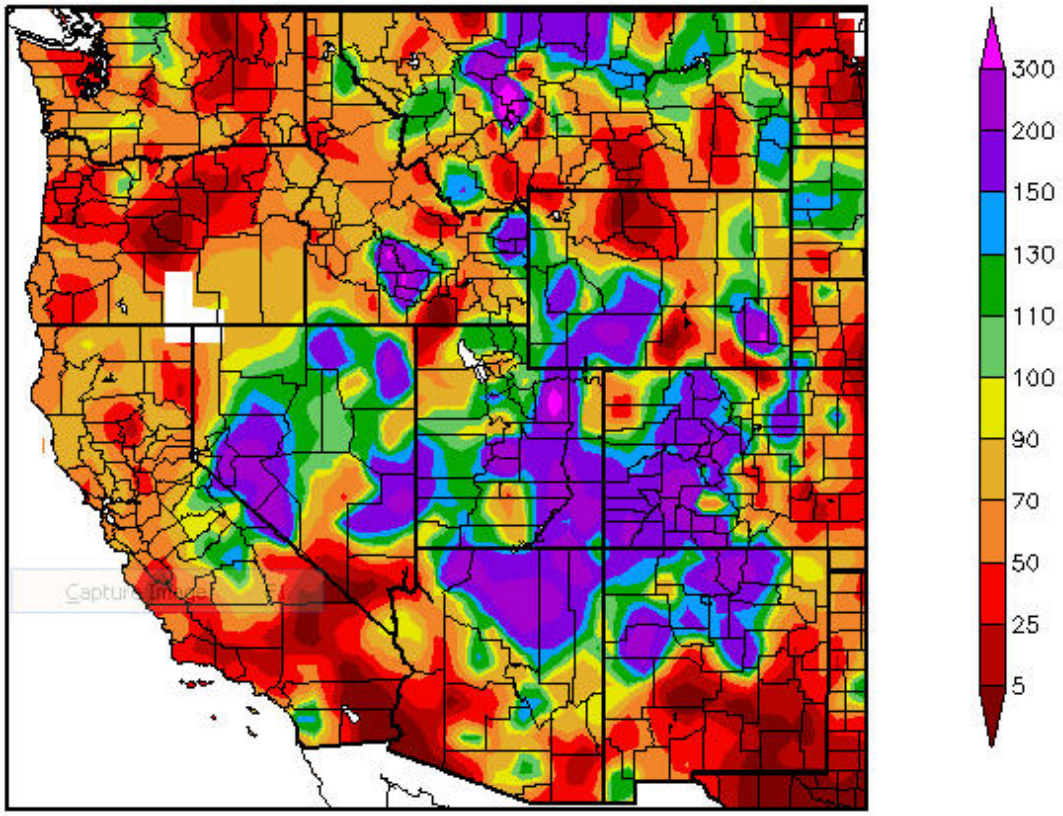
* *Data unavailable at time of posting or measurement is not representative at this time of year*

Provisional Data
Subject to Revision

The snow water equivalent percent of normal represents the current snowwater equivalent found at selected SNOTEL sites in or near the basin compared to the average value for those sites on this day. Data based on the first reading of the day (typically 00:00).

Prepared by the USDA/NRCS National Water and Climate Center
Portland, Oregon http://www.wcc.nrcs.usda.gov/gis/
Based on data from http://www.wcc.nrcs.usda.gov/reports/
Science contact: Tom.Pagano@por.usda.gov 503 414 3010

Percent of Normal Precipitation (%)
2/1/2008 – 2/29/2008



Generated 3/3/2008 at HPRCC using provisional data.

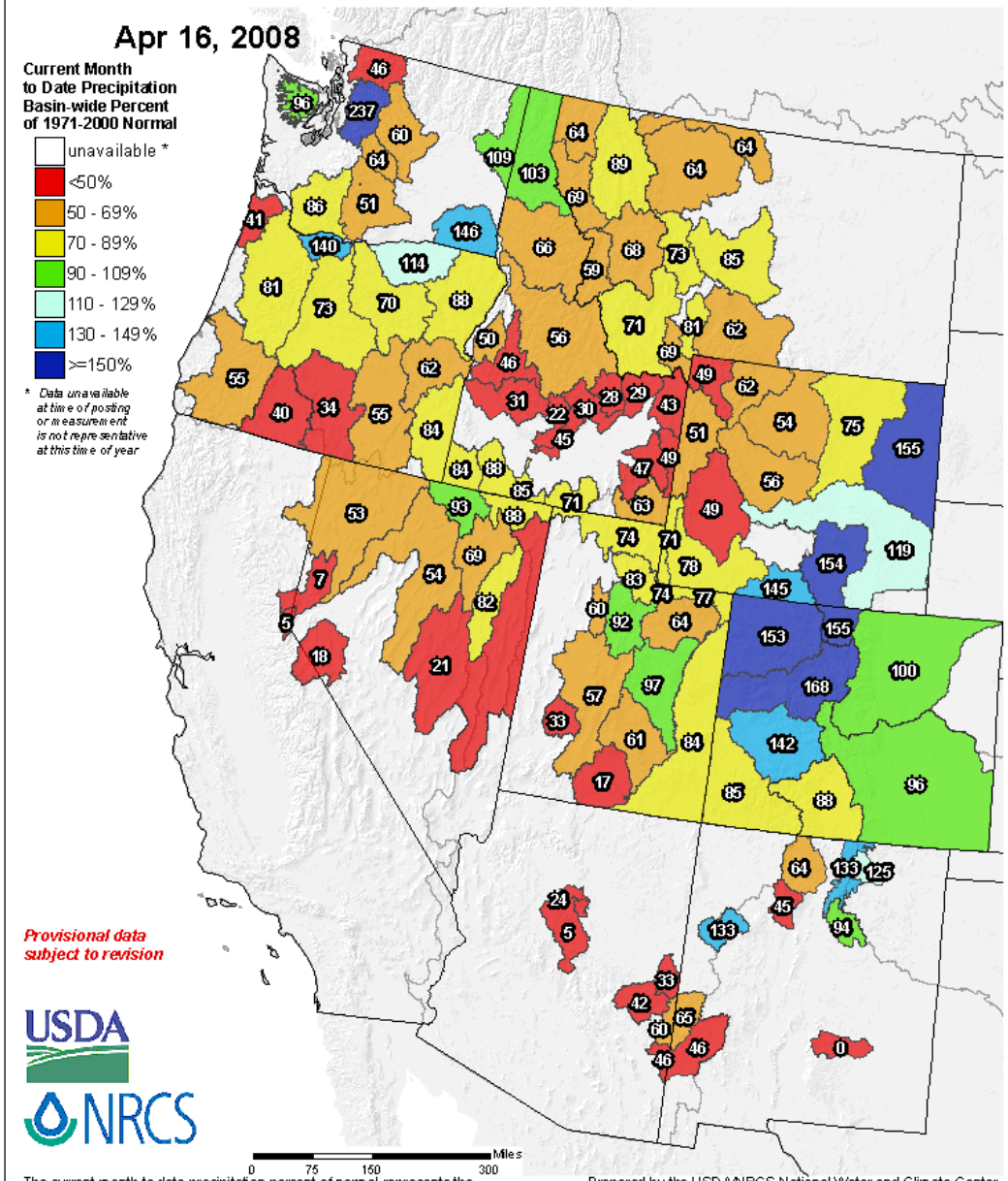
NOAA Regional Climate Centers

The Current Climate Summary Maps are produced daily using data from the Applied Climate Information System (ACIS). Stations used are from the National Weather Service Cooperative Observer Network (COOP), and the Automated Weather Data Network (AWDN). All near-real-time data are considered preliminary and should be used responsibly.

Normal refers to the 1971-2000 Climate Normal for the selected product.

ACIS

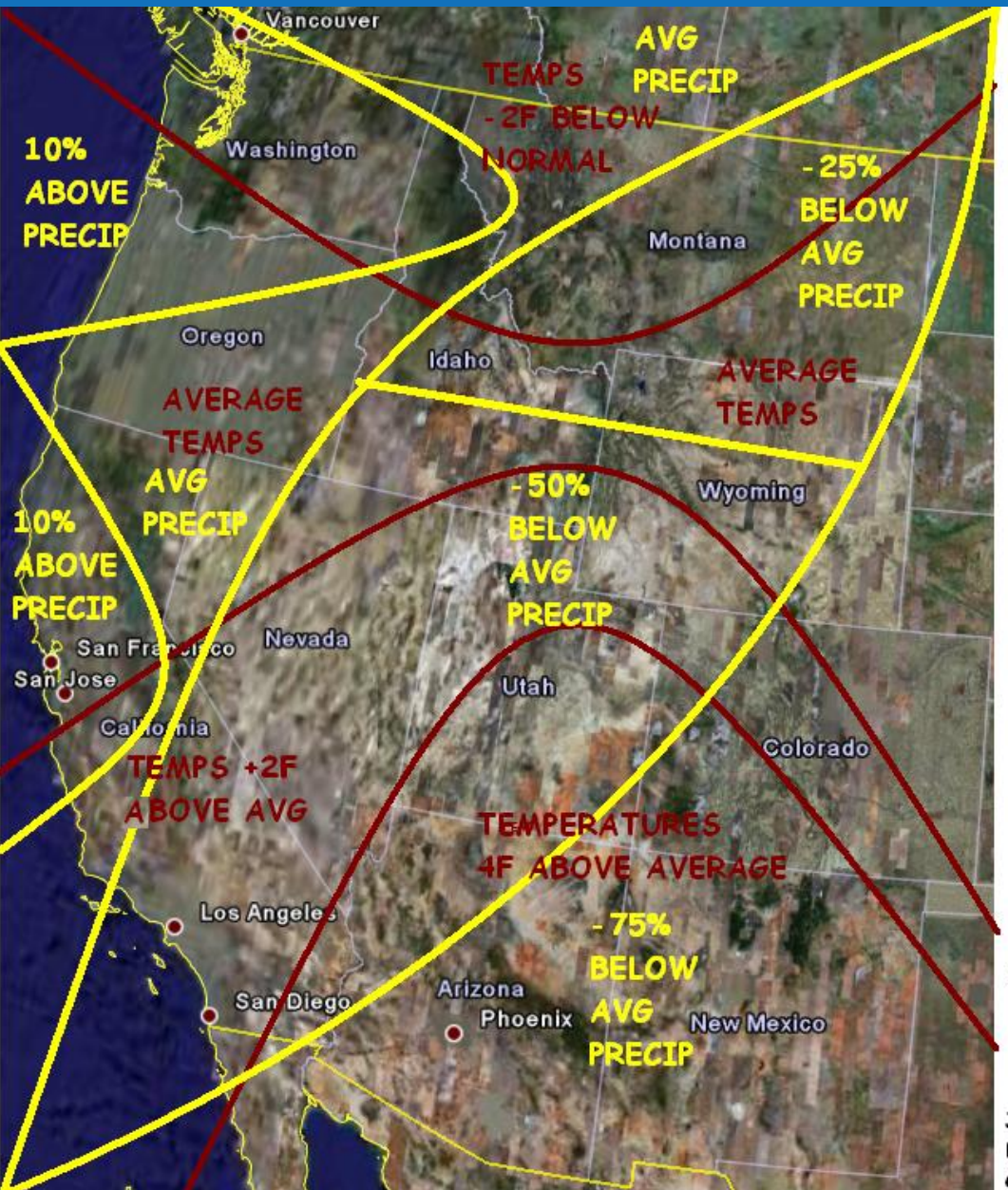
Westwide SNOTEL Current Month to Date Precipitation % of Normal

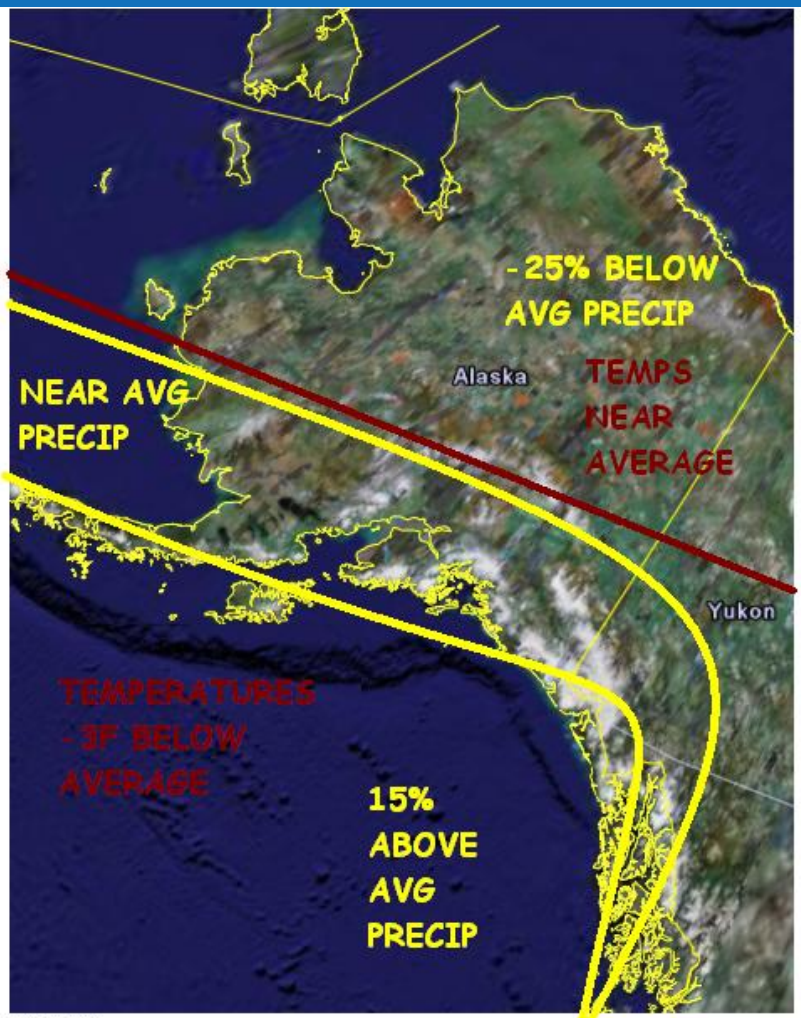


The current month to date precipitation percent of normal represents the accumulated precipitation found at selected SNOTEL sites in or near the basin compared to the average value for those sites on this day. Data based on the first reading of the day (typically 00:00).

Prepared by the USDA/NRCS National Water and Climate Center
Portland, Oregon http://www.wcc.nrcs.usda.gov/gis/
Based on data from http://www.wcc.nrcs.usda.gov/reports/
Science contact: Tom.Pagano@por.usda.gov 503 414 3010

USDA NRCS





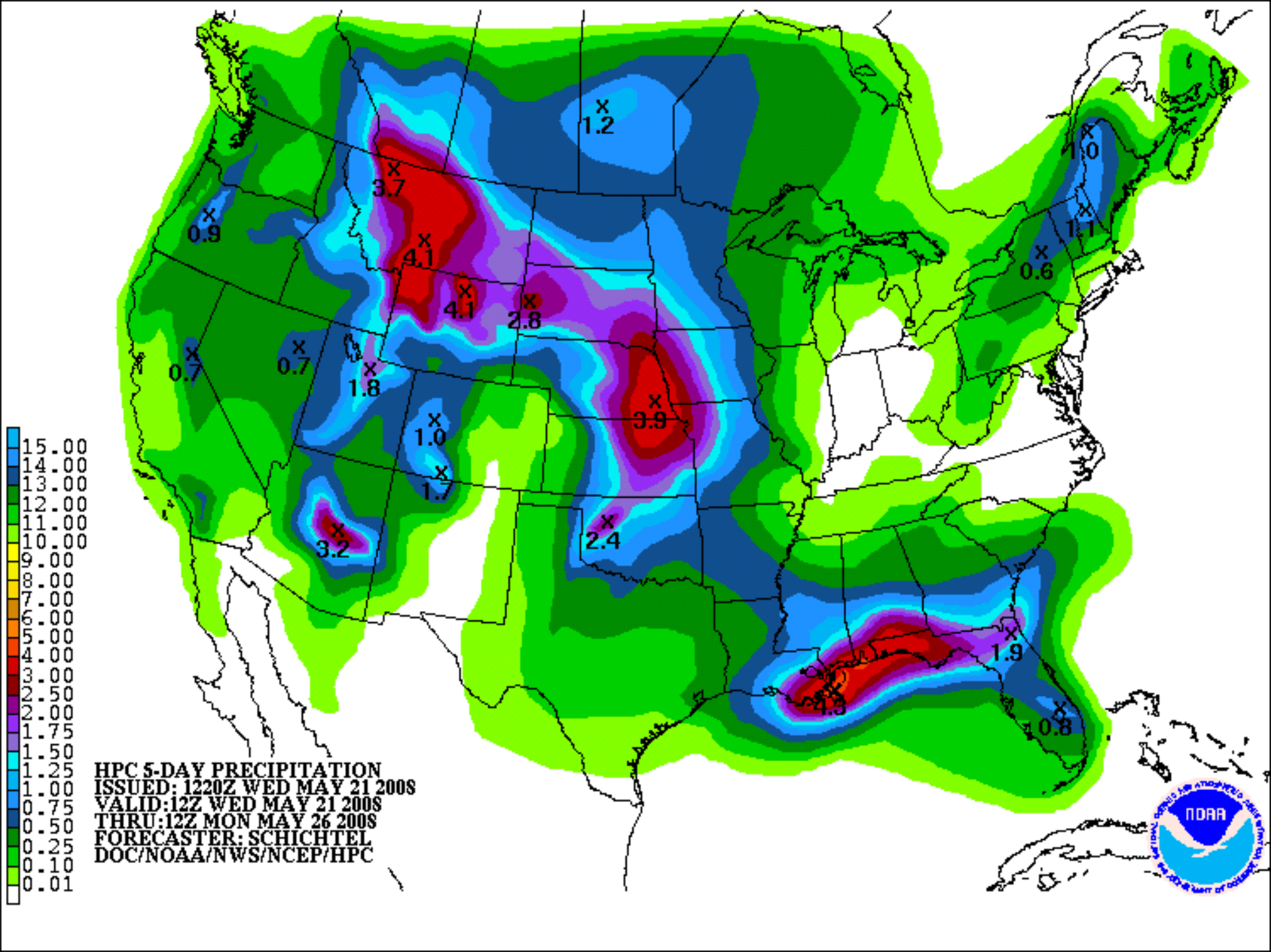
prepared 1 April 2008

Weather Outlook:

1 -15 April 2008

Jan Curtis
National Water & Climate Center, Portland, OR

United States Department of Agriculture
Natural Resources Conservation Service

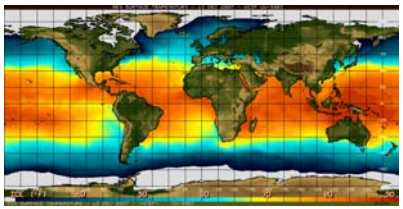

HPC 5-DAY PRECIPITATION
ISSUED: 1220Z WED MAY 21 2008
VALID:12Z WED MAY 21 2008
THRU:12Z MON MAY 26 2008
FORECASTER: SCHICHTEL
DOC/NOAA/NWS/NCEP/HPC
15.00
14.00
13.00
12.00
11.00
10.00
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.50
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.10
0.01
1.2
3.7
0.9
4.1
4.1
2.8
0.7
0.7
1.8
1.0
1.7
3.9
3.2
2.4
1.0
1.1
0.6
1.9
4.3
0.8
NOAA

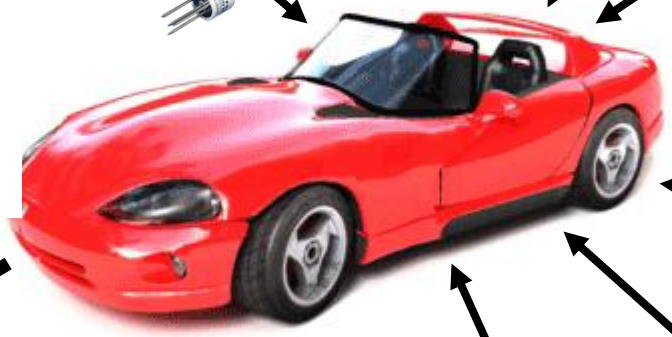
Autumn Precipitation

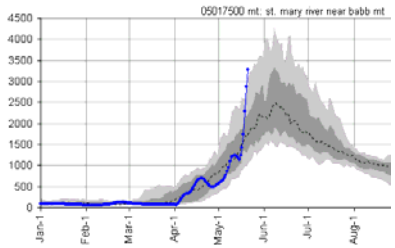
Current Snowpack

Soil Moisture

Teleconnection Indices



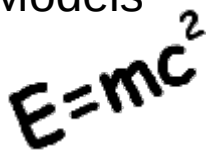




Streamflow

...Seasonal Runoff Volumes…
with Error Bounds

Regression Models



Interagency Coordination

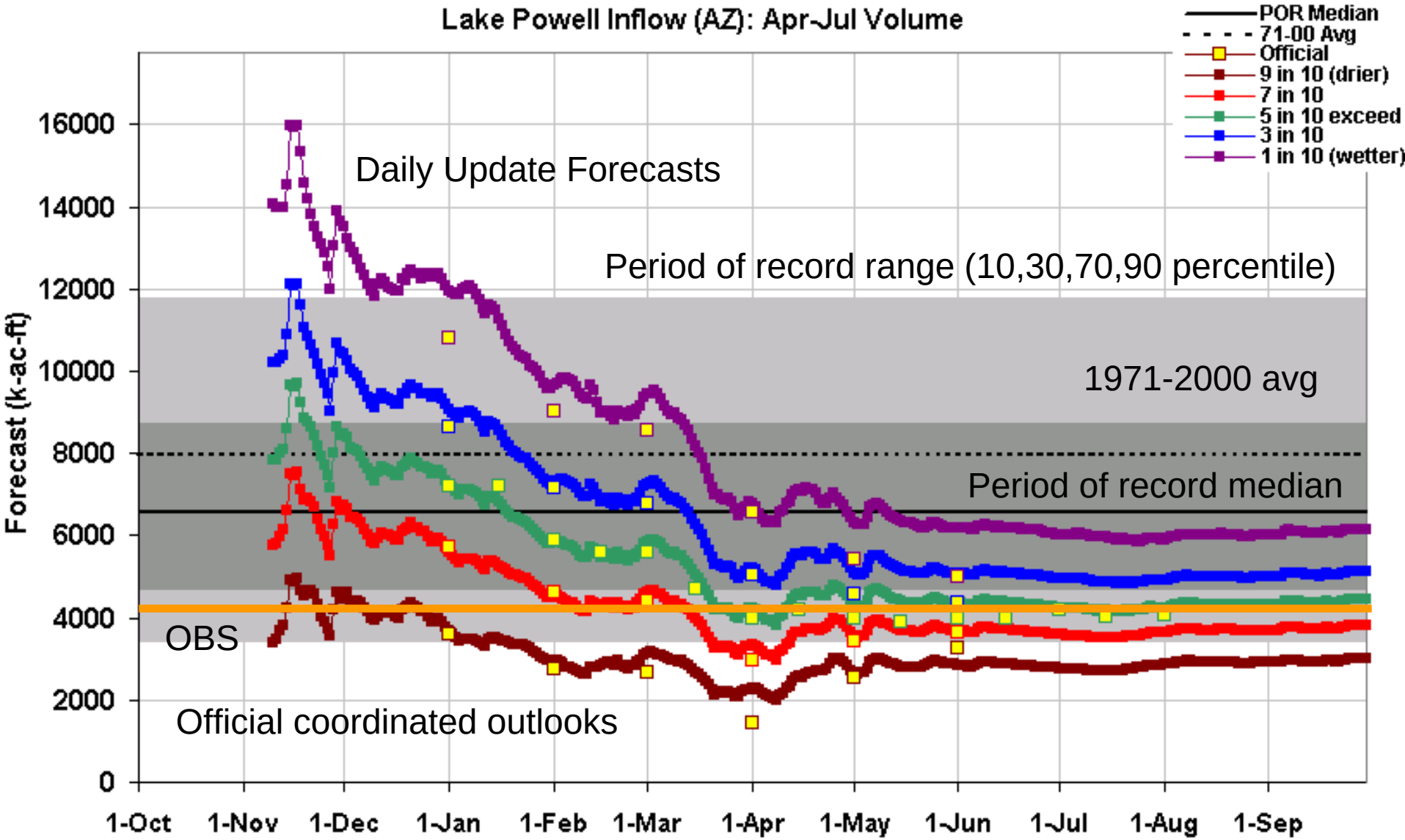
Daily Update Guidance

http://www.wcc.nrcs.usda.gov/wsf/daily_forecasts.html

Began October 2005 at request of Rio Grande water users.
Now, 148 forecast points running daily.
(several in Montana & Wyoming)

Like the official seasonal forecasts except totally objective, updated daily, and uses only snotel data.

Calibration highly automated and fast
Forecaster specifies input/output stations



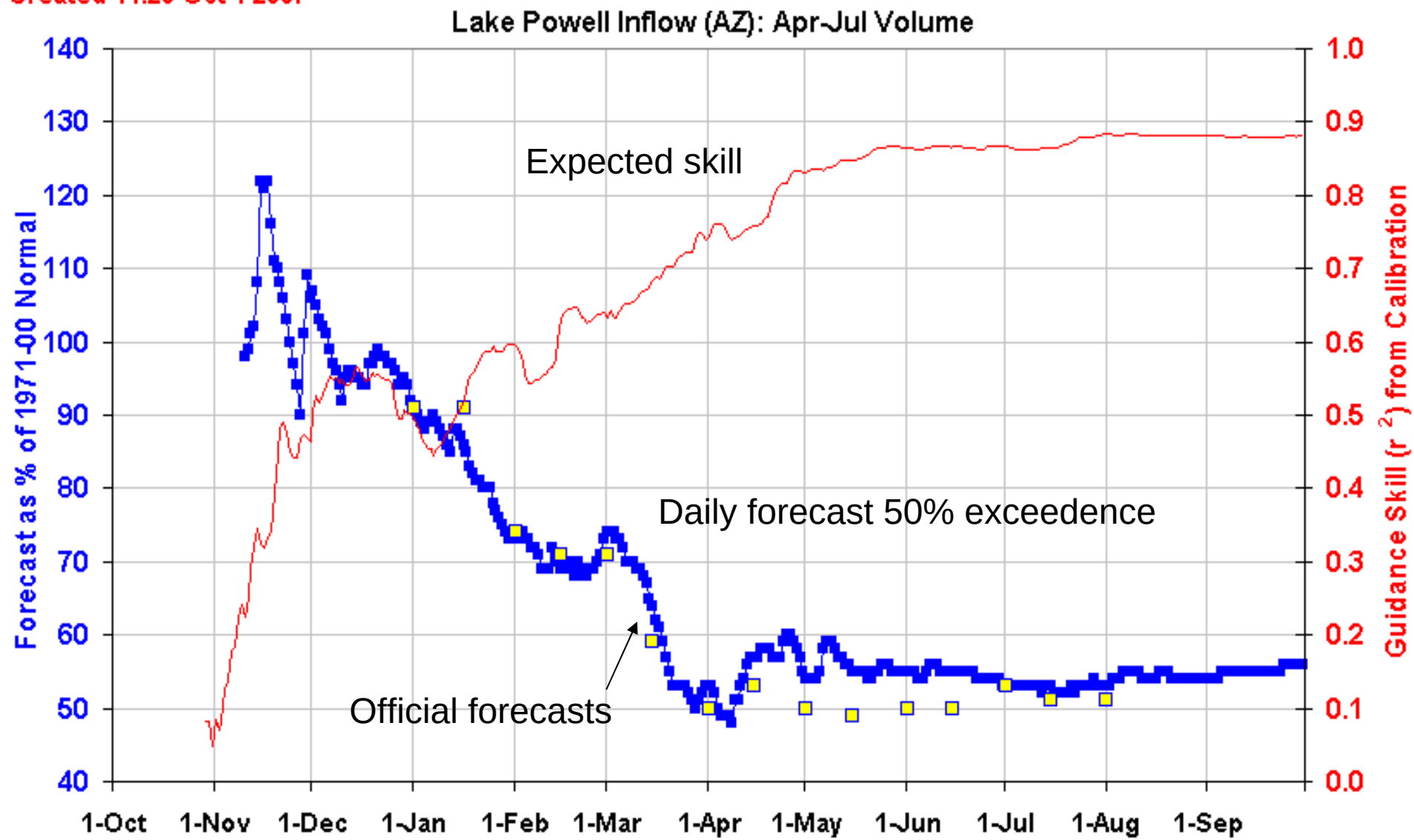
Created 11:31 Oct 1 2007

NRCS

This is an automated product based solely on SNOTEL data, provisional data are subject to change. This product is a statistically based guidance forecast combining indices of snowpack and precipitation. **Yellow squares** are the official outlooks. **Gray background** is the historical period of record variability. This product does not consider climate information such as El Nino or short range weather forecasts, or a variety of other factors considered in the official forecasts. This product is not meant to replace or supercede the official forecasts produced in coordination with the National Weather Service.

Science Contact: Tom.Pagano@por.usda.gov 503 414 3010 www.wcc.nrcs.usda.gov/wsf/daily_forecasts.html

Created 11:20 Oct 1 2007

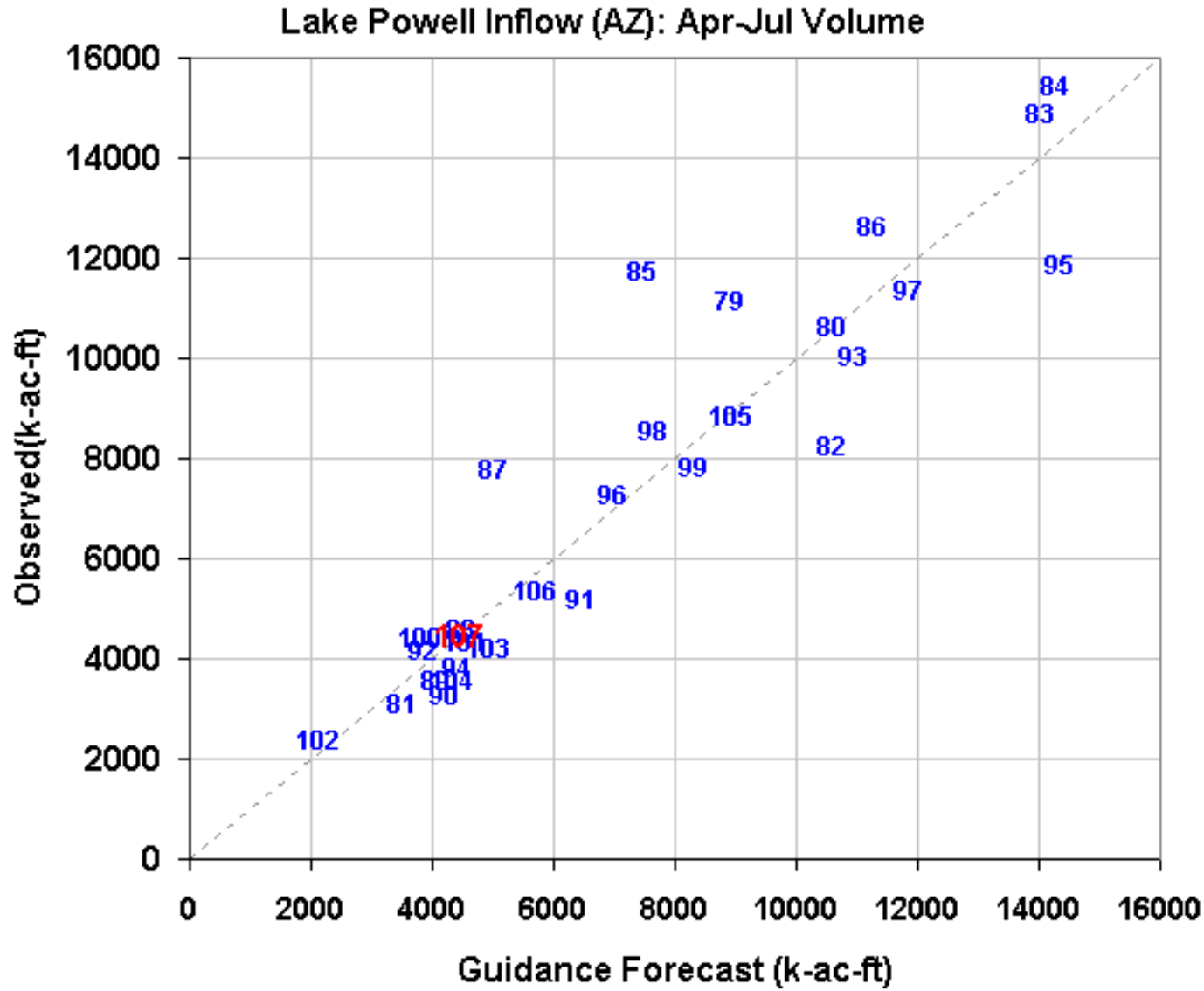


Guidance fcst % norm
Official fcst % norm
Guidance Skill (r2)

This is an automated product based solely on SNOTEL data, provisional data are subject to change. This product is a statistically based guidance forecast combining indices of snowpack and precipitation. Skill is defined as the correlation (squared) between the guidance and observed during calibration. This product does not consider climate information such as El Nino or short range weather forecasts, or a variety of other factors considered in the official forecasts. This product is not meant to replace or supercede the official forecasts produced in coordination with the National Weather Service.

Science Contact: Tom.Pagano@por.usda.gov 503 414 3010 www.wcc.nrcs.usda.gov/wsf/daily_forecasts.html

NRCS



This is an automated product based solely on SNOTEL data, provisional data are subject to change. **Each number** (e.g. "99", "104") represents an individual year (e.g. "1999","2004") from the calibration set of the guidance forecast equation for today's date. The most recent guidance forecast is shown in red. This product is not meant to replace or supercede the official forecasts produced in coordination with the National Weather Service and should only be used for planning purposes.

Science Contact: Tom.Pagano@por.usda.gov 503 414 3010

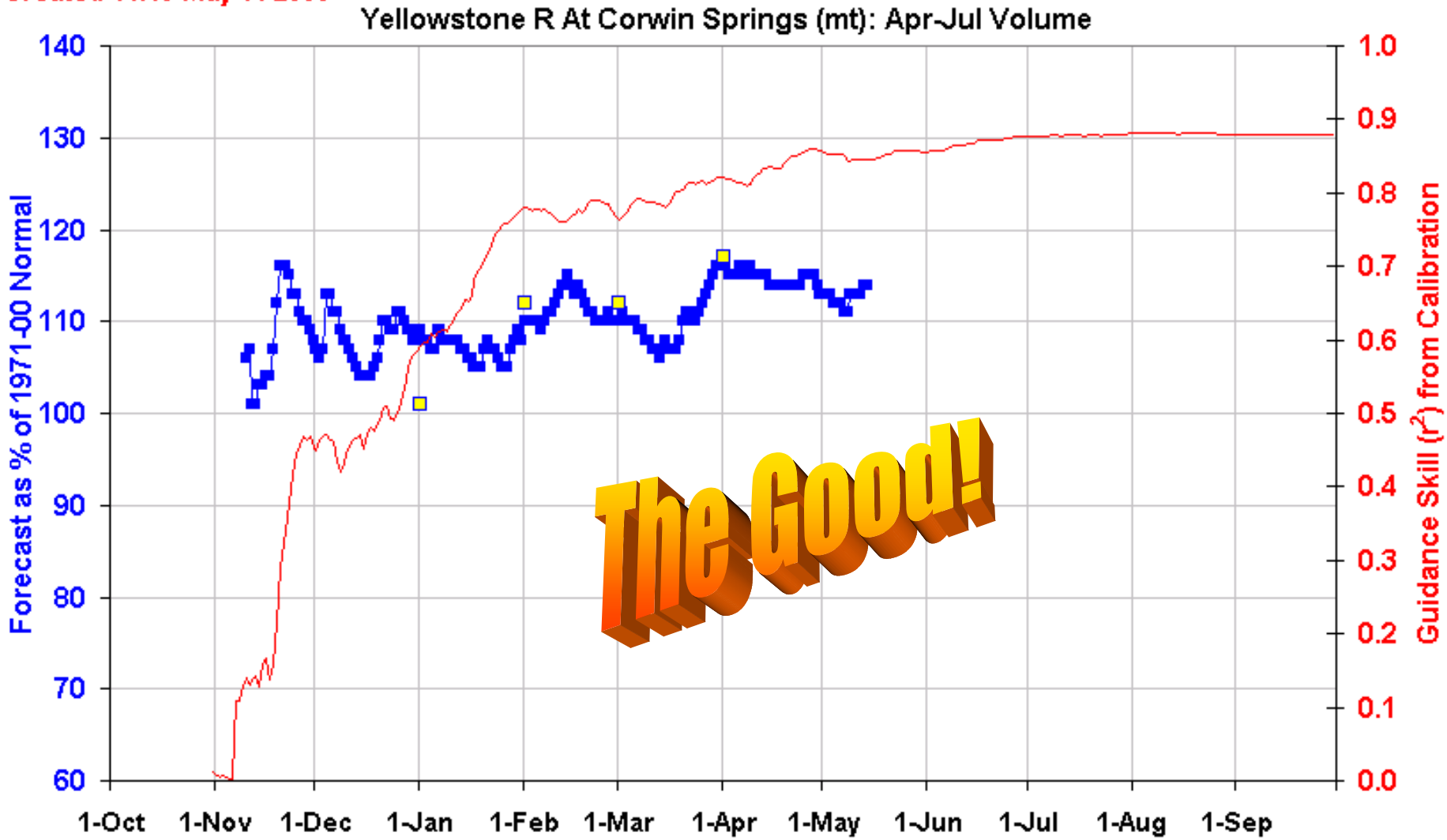
http://www.wcc.nrcs.usda.gov/wsf/daily_forecasts.html

Generated 11:38 Oct 1 2007

USDA NRCS

Calibration dataset of daily equation

Created 11:48 May 14 2008



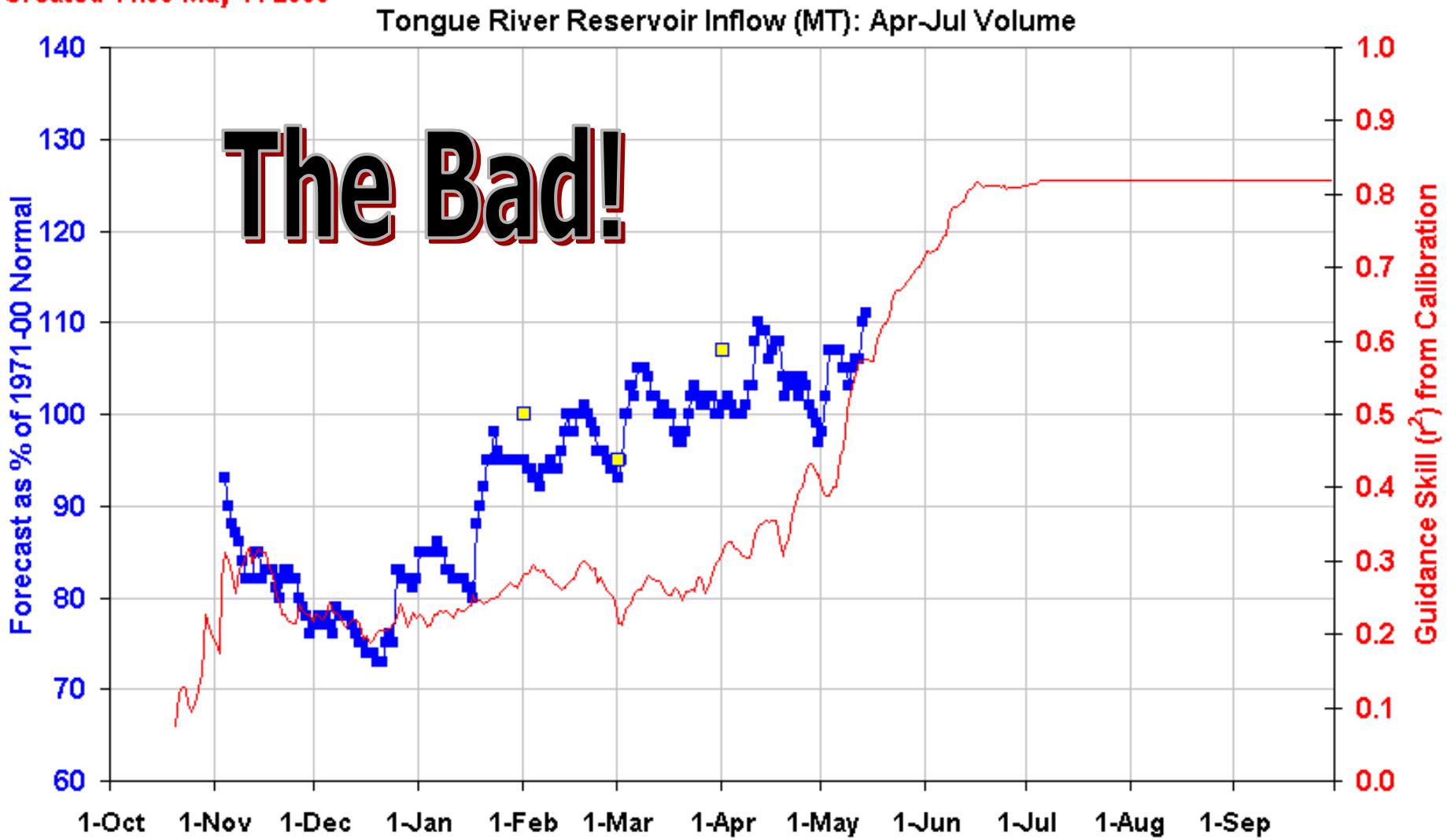
Guidance fcst % norm
Official fcst % norm
Guidance Skill (r2)

This is an automated product based solely on SNOTEL data, provisional data are subject to change. This product is a statistically based guidance forecast combining indices of snowpack and precipitation. Skill is defined as the correlation (squared) between the guidance and observed during calibration. This product does not consider climate information such as El Nino or short range weather forecasts, or a variety of other factors considered in the official forecasts. This product is not meant to replace or supercede the official forecasts produced in coordination with the National Weather Service.

Science Contact: Tom.Pagano@por.usda.gov 503 414 3010 www.wcc.nrcs.usda.gov/wsf/daily_forecasts.html

NRCS

Created 11:50 May 14 2008



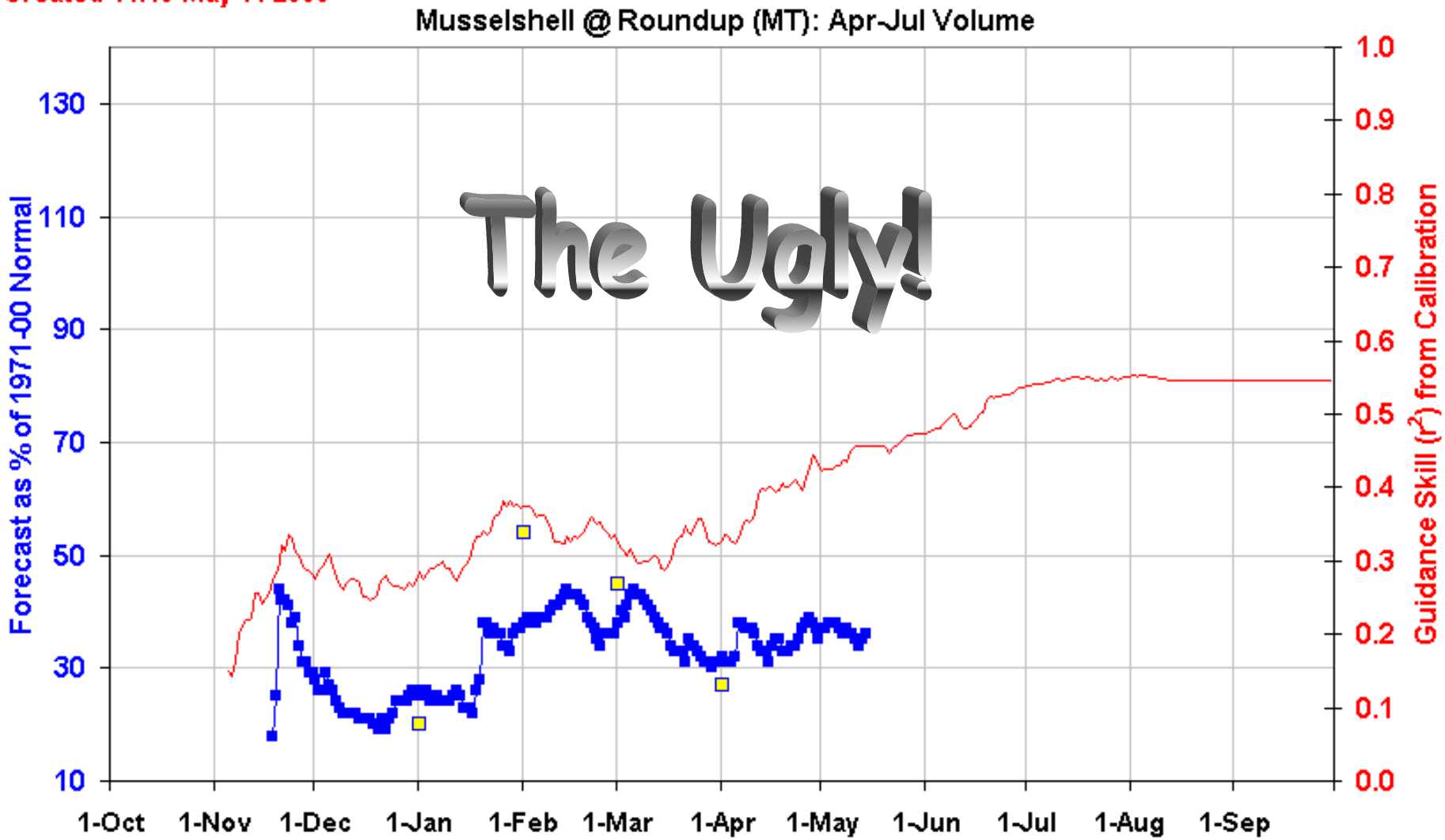
Guidance fcst % norm
Official fcst % norm
Guidance Skill (r2)

This is an automated product based solely on SNOTEL data, provisional data are subject to change. This product is a statistically based guidance forecast combining indices of snowpack and precipitation. Skill is defined as the correlation (squared) between the guidance and observed during calibration. This product does not consider climate information such as El Nino or short range weather forecasts, or a variety of other factors considered in the official forecasts. This product is not meant to replace or supercede the official forecasts produced in coordination with the National Weather Service.

Science Contact: Tom.Pagano@por.usda.gov 503 414 3010 www.wcc.nrcs.usda.gov/wsf/daily_forecasts.html

NRCS

Created 11:48 May 14 2008



Guidance fcst % norm
Official fcst % norm
Guidance Skill (r2)

This is an automated product based solely on SNOTEL data, provisional data are subject to change.
This product is a statistically based guidance forecast combining indices of snowpack and precipitation.
Skill is defined as the correlation (squared) between the guidance and observed during calibration.
This product does not consider climate information such as El Nino or short range weather forecasts, or a variety of other factors considered in the official forecasts. This product is not meant to replace or supercede the official forecasts produced in coordination with the National Weather Service.
Science Contact: Tom.Pagano@por.usda.gov 503 414 3010 www.wcc.nrcs.usda.gov/wsf/daily_forecasts.html

NRCS

Daily Update Guidance

http://www.wcc.nrcs.usda.gov/wsf/daily_forecasts.html

Skill and performance:

"Reliability" excellent
(e.g. when it says there's a 70% chance of exceeding,
it exceeds 70% of the time)

Skill comparable with official outlooks…
Westwide April 1 2007 forecast error as % 1971-2000 normal:

	Average error	Median error
Official Forecasts	**14%**	**11%**
Daily Forecasts	**12%**	**10%**

Automated daily forecasts are comparable to the official, coordinated forecasts in accuracy

Daily Update Guidance

http://www.wcc.nrcs.usda.gov/wsf/daily_forecasts.html

Requests for future development:

Include more information about inputs in final outputs

(i.e. a multi-station basin snow/precip time series plot)

Use Canadian data

Use antecedent fall stream flow data/climate indices

Forecast residual (use observed flow to date)

ftp://ftp.wcc.nrcs.usda.gov/wcs_info/MontanaHydrologyConference.ppt

USDA Natural Resources Conservation Service
National Water & Climate Center

tom.perkins@por.usda.gov 503.414.3059

So long folks!

United States Department of Agriculture
Natural Resources Conservation Service

Discussing Water Rights, A Western Pastime