US Army Corps
of Engineers
Northwestern Division

Improved Water Supply Forecasts for the Kootenay Basin

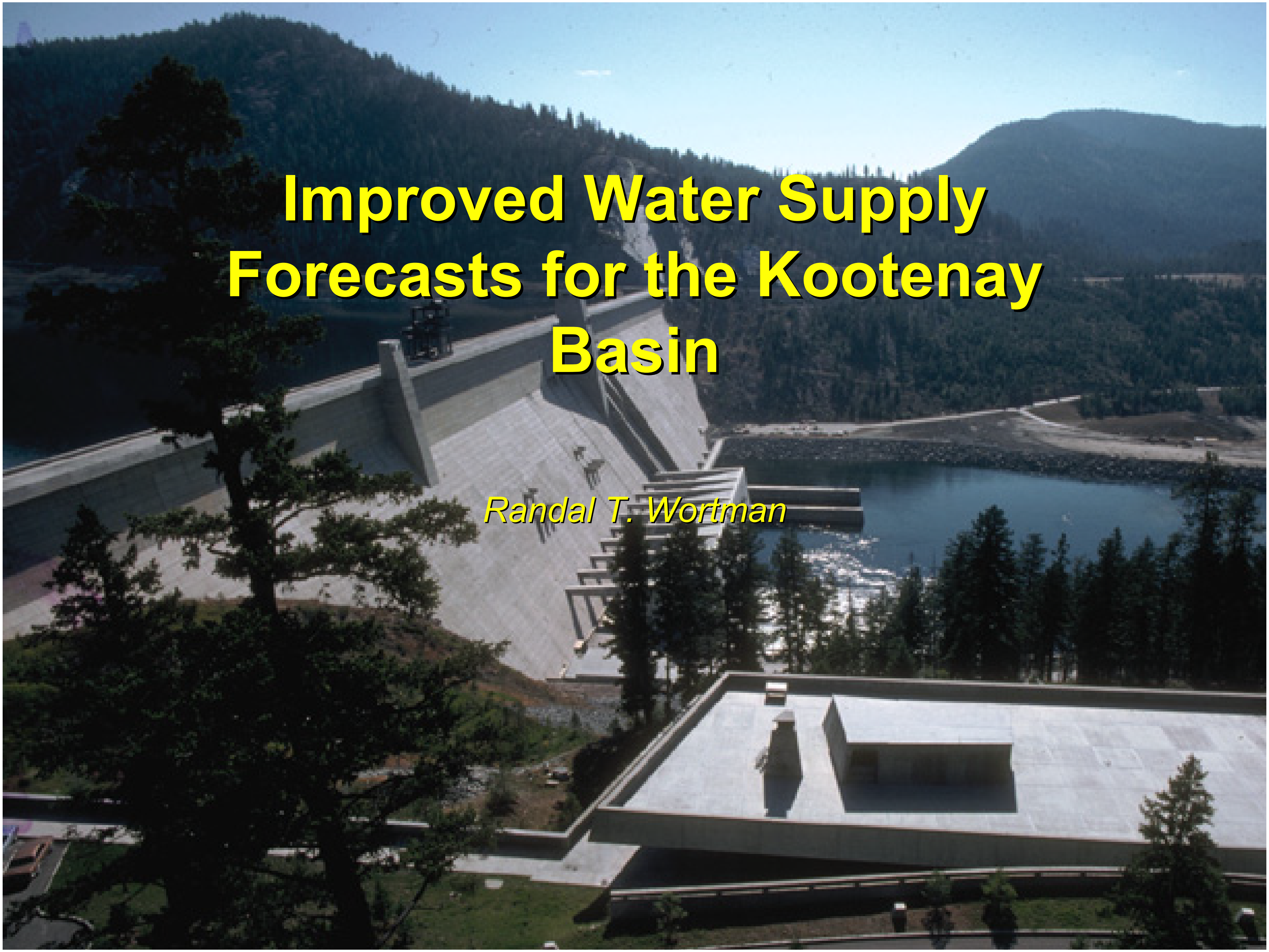
Randal T. Wortman
Hydraulic Engineer

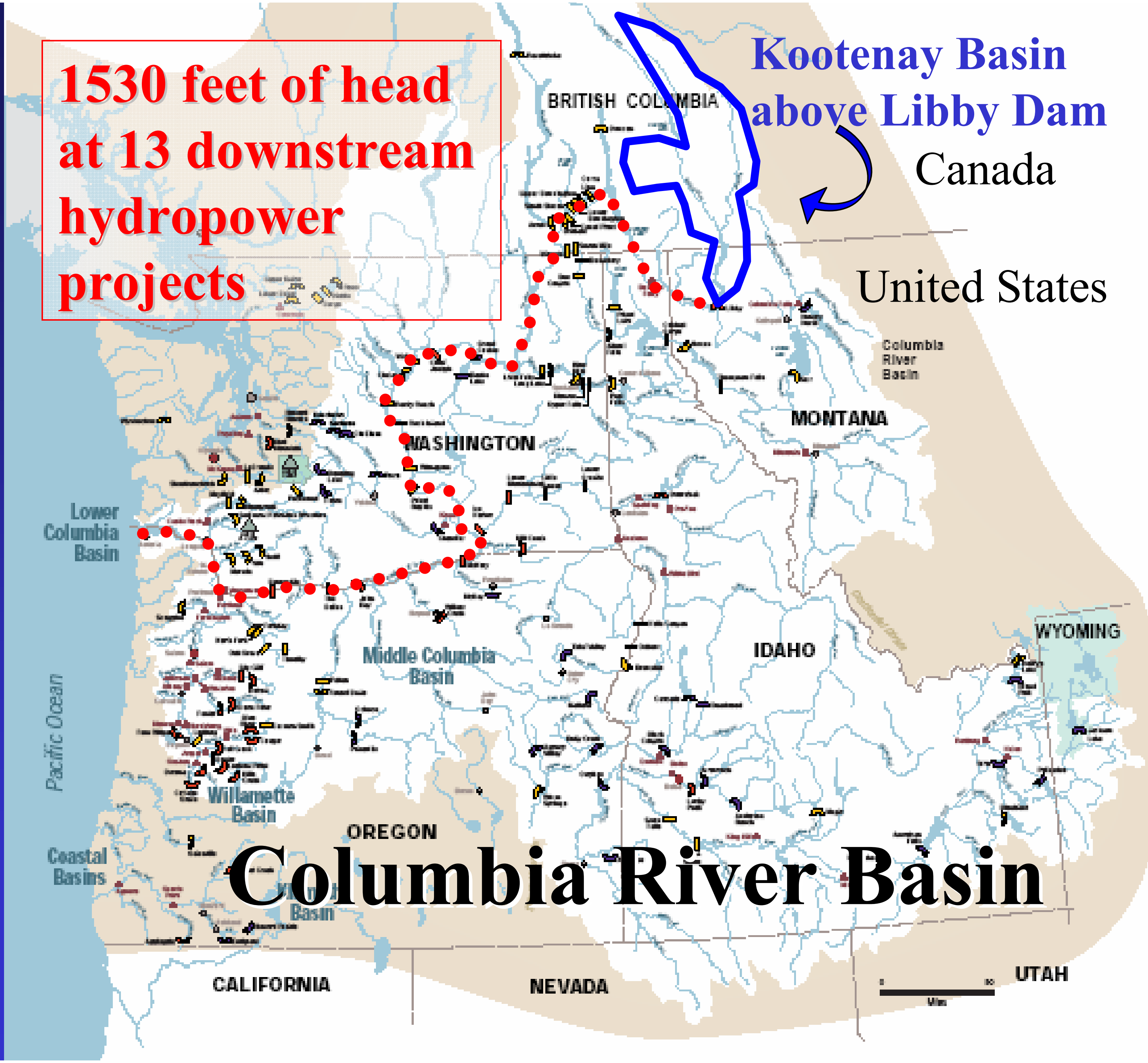
August 4, 2005

Columbia Basin Water Management Division
Northwestern Division, U.S. Army Corps of Engineers
PO Box 2870, Portland, OR 97208-2870
Ph: 503 808-3957 Email: randal.t.wortman@usace.army.mil

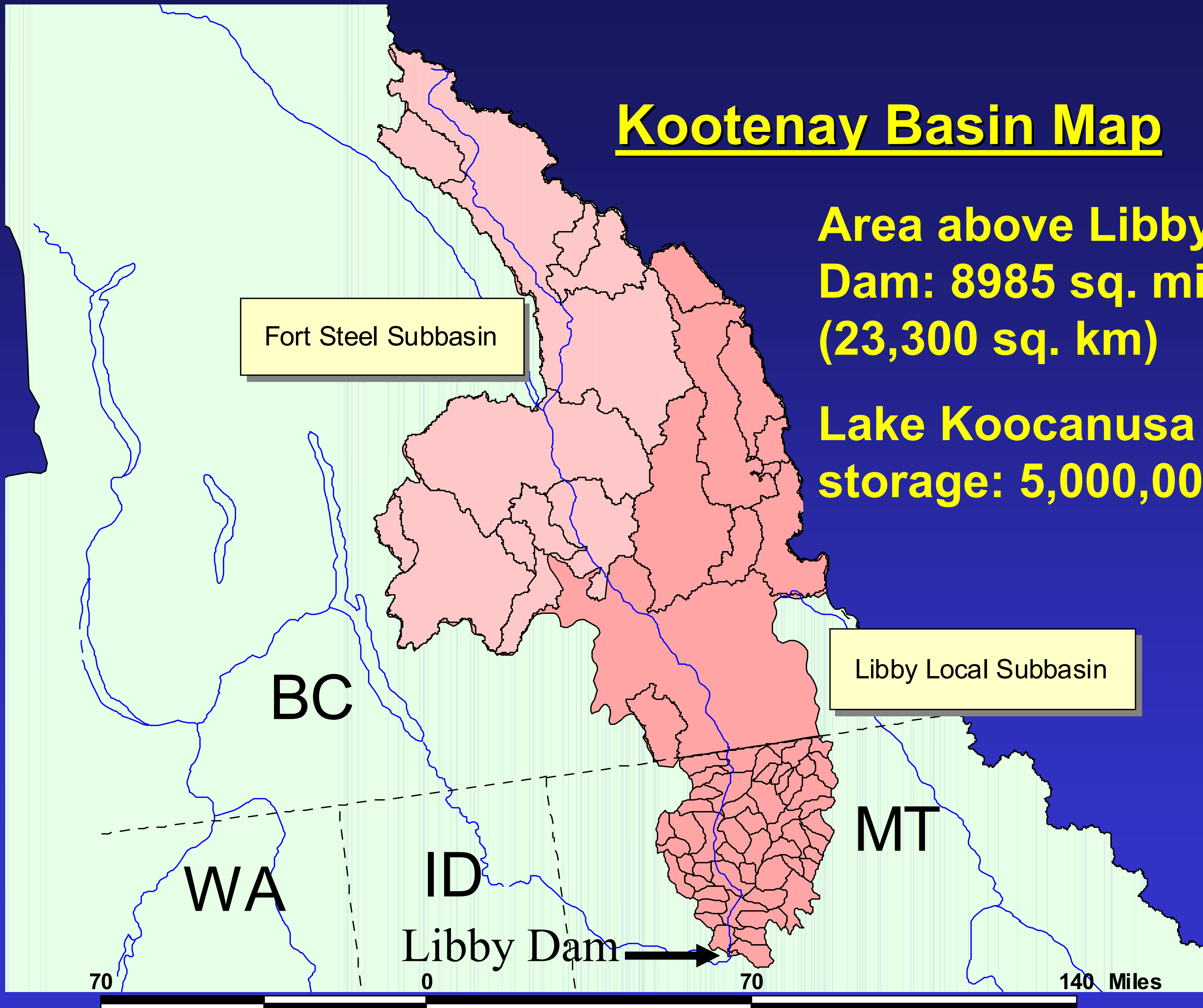
Improved Water Supply Forecasts for the Kootenay Basin

Randal T. Wortman




1530 feet of head at 13 downstream hydropower projects
Kootenay Basin above Libby Dam
Canada
United States
BRITISH COLUMBIA
Columbia River Basin
MONTANA
WASHINGTON
Lower Columbia Basin
Middle Columbia Basin
IDAHO
WYOMING
Pacific Ocean
Willamette Basin
OREGON
Coastal Basins
Columbia River Basin
CALIFORNIA
NEVADA
UTAH

Kootenay Basin Map

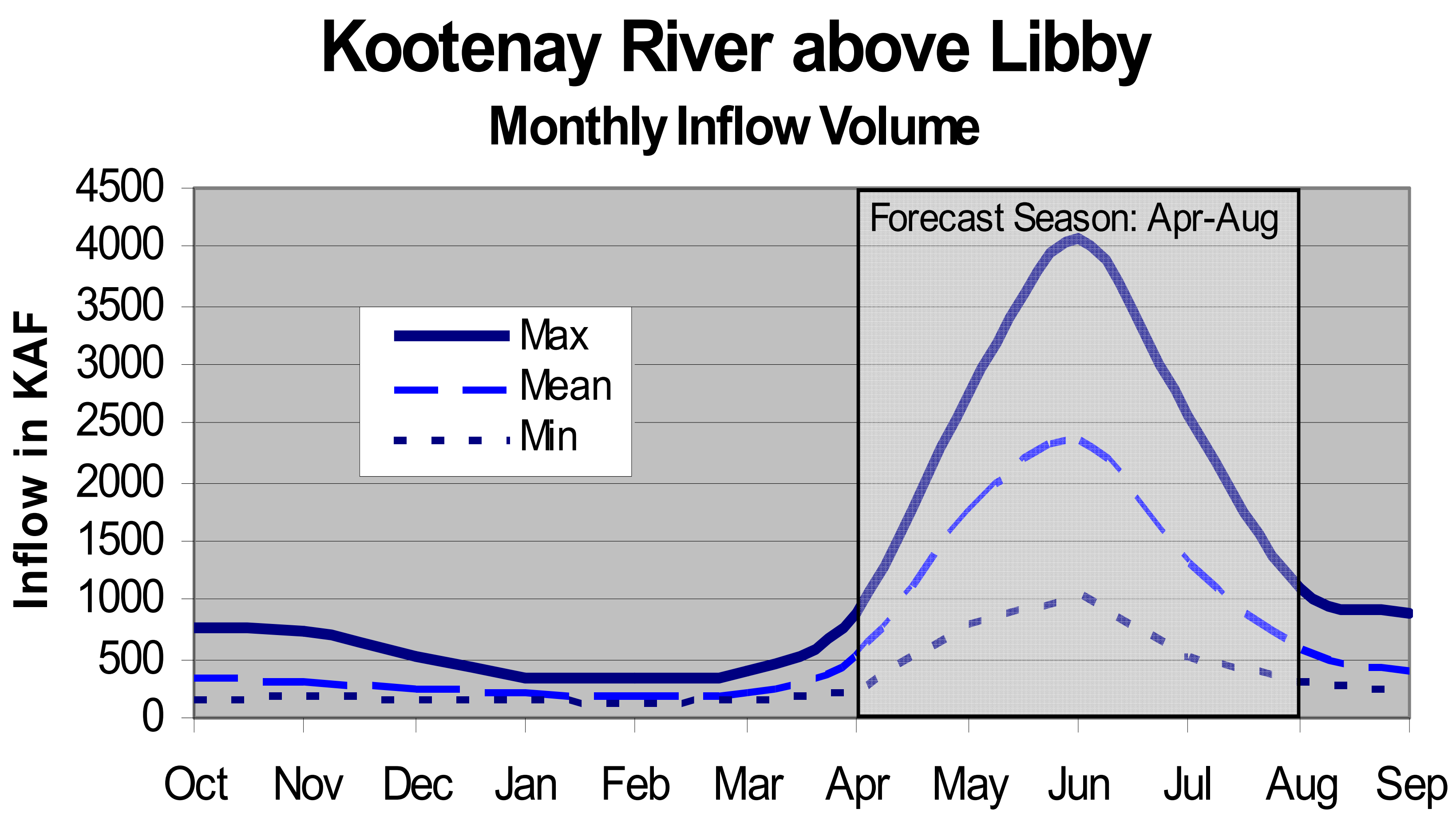


Area above Libby Dam: 8985 sq. miles (23,300 sq. km)

Lake Koocanusa storage: 5,000,000 AF

US Army Corps of Engineers
Northwestern Division

Kootenay Inflow above Libby Dam



US Army Corps
of Engineers
Northwestern Division

Standard Multiple-Variable Regression in Water Supply Forecasting

- *The dependent variable is a seasonal inflow volume, e.g. April-August runoff in thousand-acre-feet (KAF)*
- *Predictor variables are pseudo-variables created from weighted combinations of similar stations (e.g. sum or average of three snow stations)*

US Army Corps
of Engineers
Northwestern Division

Original Libby Forecast "Split-Basin" Regression Equations

Variable	Ft Steel Basin	Libby Local Basin
1 April Snow Water Equivalent (SWE)	Σ MILB, MORB, KGHB, SUMB, MBCB, GRPB, NFRB	Σ SUMB, NFRB, RMTM, KIMB, WSLM, 0.5*MORB
Winter (Oct--Mar) Precipitation (WP)	Σ Oct, Nov, Dec, Jan, Feb, Mar Σ ELKB, BABB, GRPB, BRIB, KASB	Σ Oct, Nov, Dec, Jan, Feb, Mar Σ ELKB, FENB, FTIM, LRSM, BONI, **POLM**
Spring (Apr--Aug) Precipitation (SP)	Σ Apr, May, .8 Jun, .5 Jul, .2 Aug Σ BRIB, KASB, PTHI, **WASB**, CRSB	Σ Apr, May, .8 Jun, .5 Jul, .2 Aug Σ FTIM, PTHI, KASB, WHFM
Fall Runoff (FRO)	Σ Oct, Nov Ft Steele basin runoff	Σ Oct, Nov Libby Local basin runoff

US Army Corps
of Engineers
Northwestern Division

Original Libby Forecast "Split-Basin" Regression Equations

Fort Steele Regression Model

1.309 FRO + 0.067 SWE + 0.068 WP + 0.167 SP – 5.114

R^2=.914   Sept 1 Forecast Std Error=213 KAF

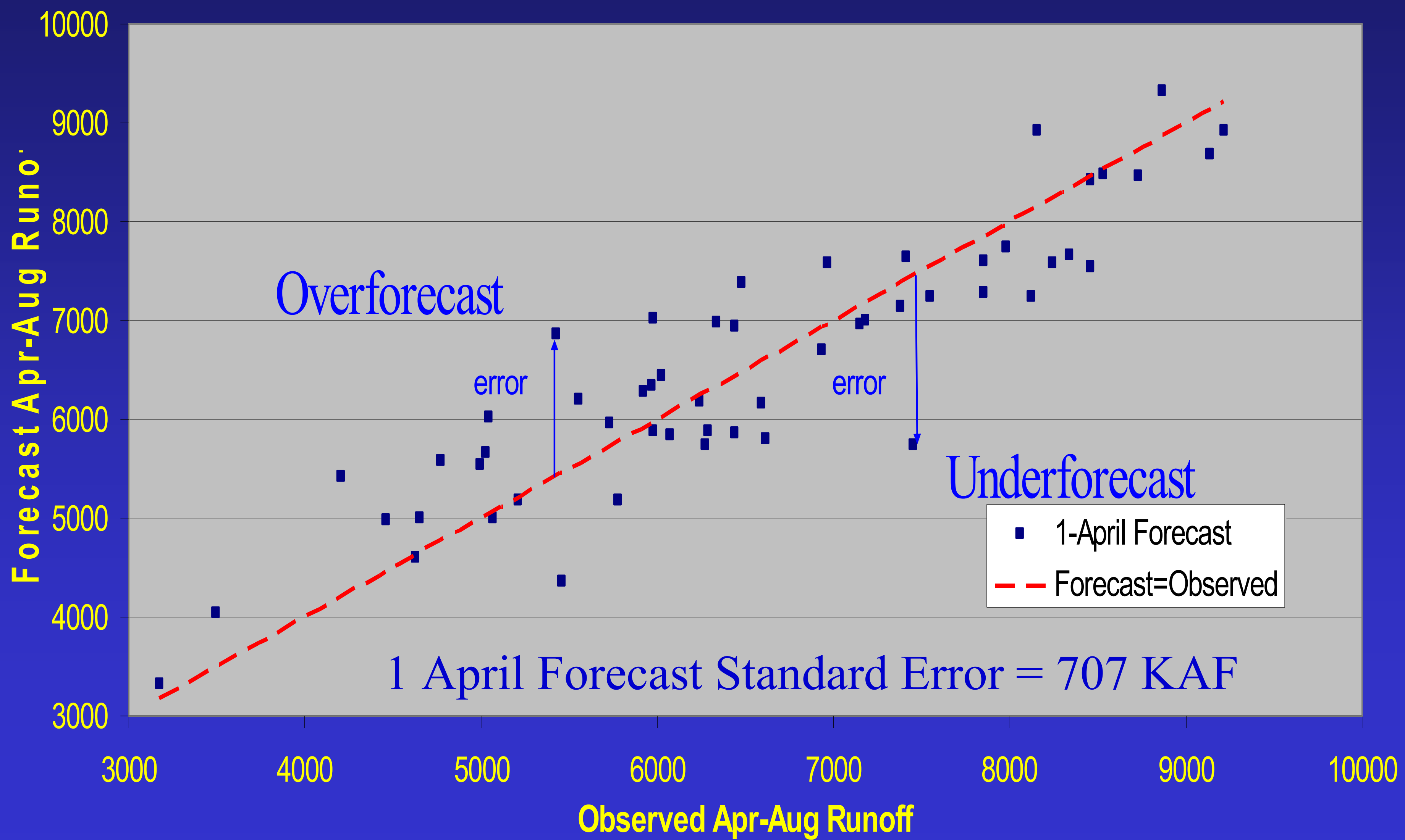
Libby Local Regression Model

0.921 FRO + 0.046 SWE + 0.086 WP + 0.152 SP – 4.183

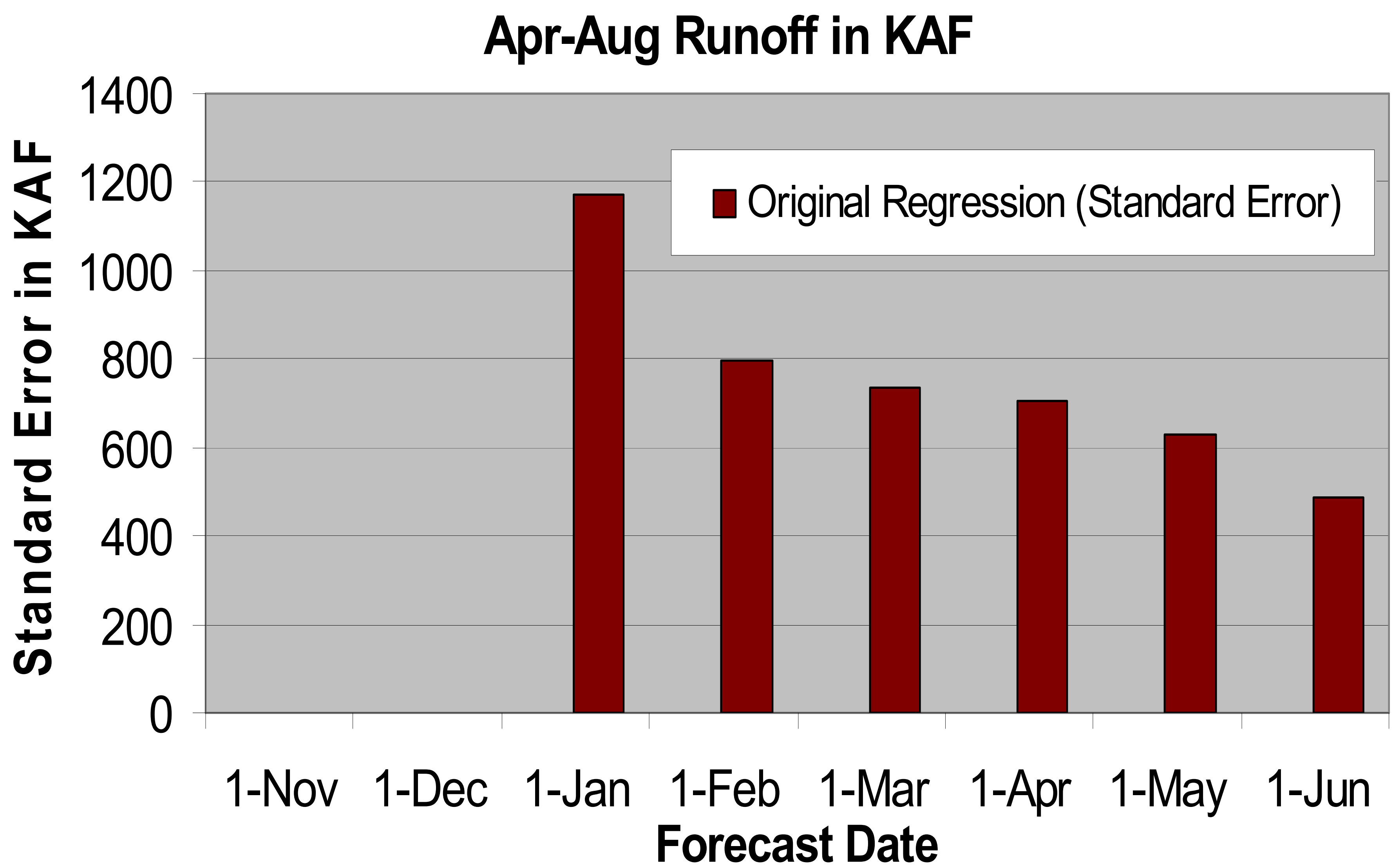
R^2=.874   Sept 1 Forecast Std Error=262 KAF

US Army Corps of Engineers
Northwestern Division

Libby Forecast Performance – 1 April Split-Basin Regression



Libby Water Supply Forecast using "Split-Basin" Standard Regression



US Army Corps
of Engineers
Northwestern Division

Concerns with Traditional Regression Models

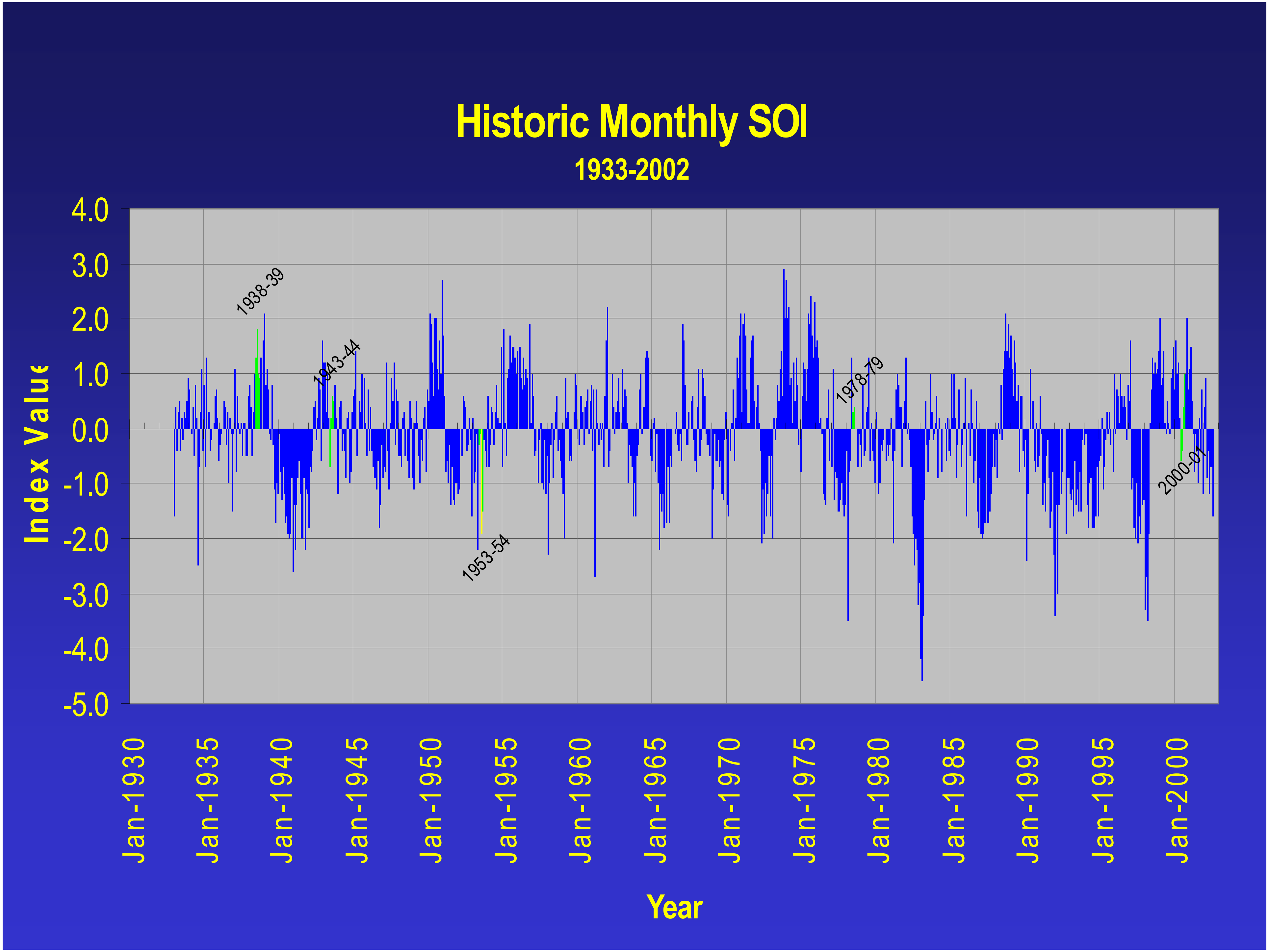
- *Subjective station selection*
- *Subjective station weighting/aggregating*
- *Use of "normal subsequent" variable as a surrogate for a "future value" variable (Avoid)*
- <u>*Predictor variables are frequently highly intercorrelated. Intercorrelated variables produce interactions and problems with the regression coefficients and goodness-of-fit model statistics.*</u>

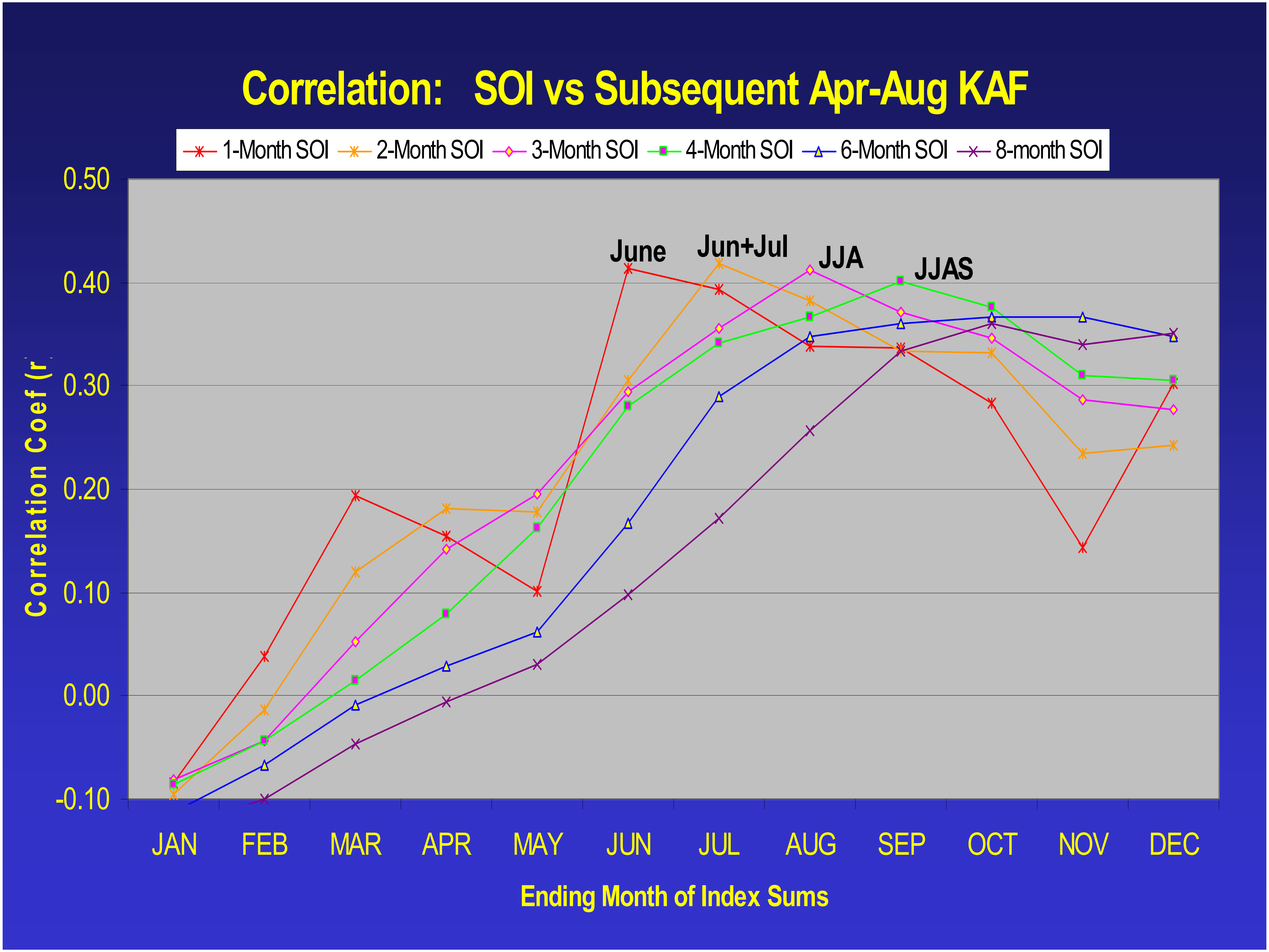
US Army Corps
of Engineers
Northwestern Division

Objectives for New Forecast Equations

- Single regression model for the entire basin
  - New equation for each month
  - Choose models to maintain month-to-month consistency of variables
- Investigate climate variables
  - SOI and PDOI
- Eliminate intercorrelation between predictor variables
- Optimize variable weighting and selection procedures
- Model evaluation/selection utilizes cross-validation statistics

=>Regression variables selected and fitted utilizing NRCS Principal Components regression procedure


Historic Monthly SOI
1933-2002
Index Value
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
1938-39
1943-44
1953-54
1978-79
2000-01
Jan-1930
Jan-1935
Jan-1940
Jan-1945
Jan-1950
Jan-1955
Jan-1960
Jan-1965
Jan-1970
Jan-1975
Jan-1980
Jan-1985
Jan-1990
Jan-1995
Jan-2000
Year


Correlation: SOI vs Subsequent Apr-Aug KAF
1-Month SOI
2-Month SOI
3-Month SOI
4-Month SOI
6-Month SOI
8-month SOI
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
Correlation Coef (r)
June
Jun+Jul
JJA
JJAS
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC
Ending Month of Index Sums

US Army Corps
of Engineers
Northwestern Division

Principal Components

Observed Data

$X1_1$	$X2_1$	$X3_1$
$X1_2$	$X2_2$	$X3_2$
$X1_3$	$X2_3$	$X3_3$
$X1_4$	$X2_4$	$X3_4$
$X1_5$	$X2_5$	$X3_5$
$X1_6$	$X2_6$	$X3_6$
…	…	…
$X1_N$	$X2_N$	$X3_N$

Principal Components

$PC1_1$	$PC2_1$	$PC3_1$
$PC1_2$	$PC2_2$	$PC3_2$
$PC1_3$	$PC2_3$	$PC3_3$
$PC1_4$	$PC2_4$	$PC3_4$
$PC1_5$	$PC2_5$	$PC3_5$
$PC1_6$	$PC2_6$	$PC3_6$
…	…	…
$PC1_N$	$PC2_N$	$PC3_N$

US Army Corps
of Engineers
Northwestern Division

Principal Components

- *Creates surrogate variables (principal components) that are a weighted combination of the original variables.*
- *The principal components have the property of being fully independent of one another (zero intercorrelation)*
- *Most of the "variability" in the predictor variables is loaded into the first one or two components.*
- *Eigenvalues reflect the proportion of the variability in original variables loaded into each component*
- *Note: PCs combine the information within the predictor variables, but "have no knowledge" of the dependent variable, the variable to be forecasted.*

US Army Corps
of Engineers
Northwestern Division

Principal Components Regression

Observed Data

$X1_1$	$X2_1$	$X3_1$
$X1_2$	$X2_2$	$X3_2$
$X1_3$	$X2_3$	$X3_3$
$X1_4$	$X2_4$	$X3_4$
$X1_5$	$X2_5$	$X3_5$
$X1_6$	$X2_6$	$X3_6$
...	...	...
$X1_N$	$X2_N$	$X3_N$

Traditional Regression Model

$$Y = \beta_0 + \beta_1 * X1 + \beta_2 * X2 + \beta_3 * X3$$

Principal Components

$PC1_1$	$PC2_1$	$PC3_1$
$PC1_2$	$PC2_2$	$PC3_2$
$PC1_3$	$PC2_3$	$PC3_3$
$PC1_4$	$PC2_4$	$PC3_4$
$PC1_5$	$PC2_5$	$PC3_5$
$PC1_6$	$PC2_6$	$PC3_6$
...	...	...
$PC1_N$	$PC2_N$	$PC3_N$

Principal Component Regression Model

$$Y = \beta_0 + \beta_1 * PC1 + \beta_2 * PC2 + \beta_3 * PC3$$

US Army Corps
of Engineers
Northwestern Division

US ARMY CORPS OF ENGINEERS
NORTHWESTERN DIVISION

Principal Components Regression

Example using SOI, 2 Precip & 4 snow variables

- ***Properties of the Principal Component Regression Model with all "P" Components ("p"= # of original variables)***
  - Component R-squared values
  - Component R-squared loading
  - Eigenvalue loading
- **7 original variables -> 7 principal components:**

PC Analysis	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7
R-Square	0.84144	0.00383	0.00040	0.00003	0.00495	0.00336	0.00068
R-Square %	98.5%	0.4%	0.0%	0.0%	0.6%	0.4%	0.1%
Cumul R-Sqr	0.8414	0.8453	0.8457	0.8457	0.8506	0.8540	0.8547
Eigenvalues	4.41	0.80	0.74	0.55	0.36	0.09	0.04

US Army Corps
of Engineers
Northwestern Division

Principal Components Regression

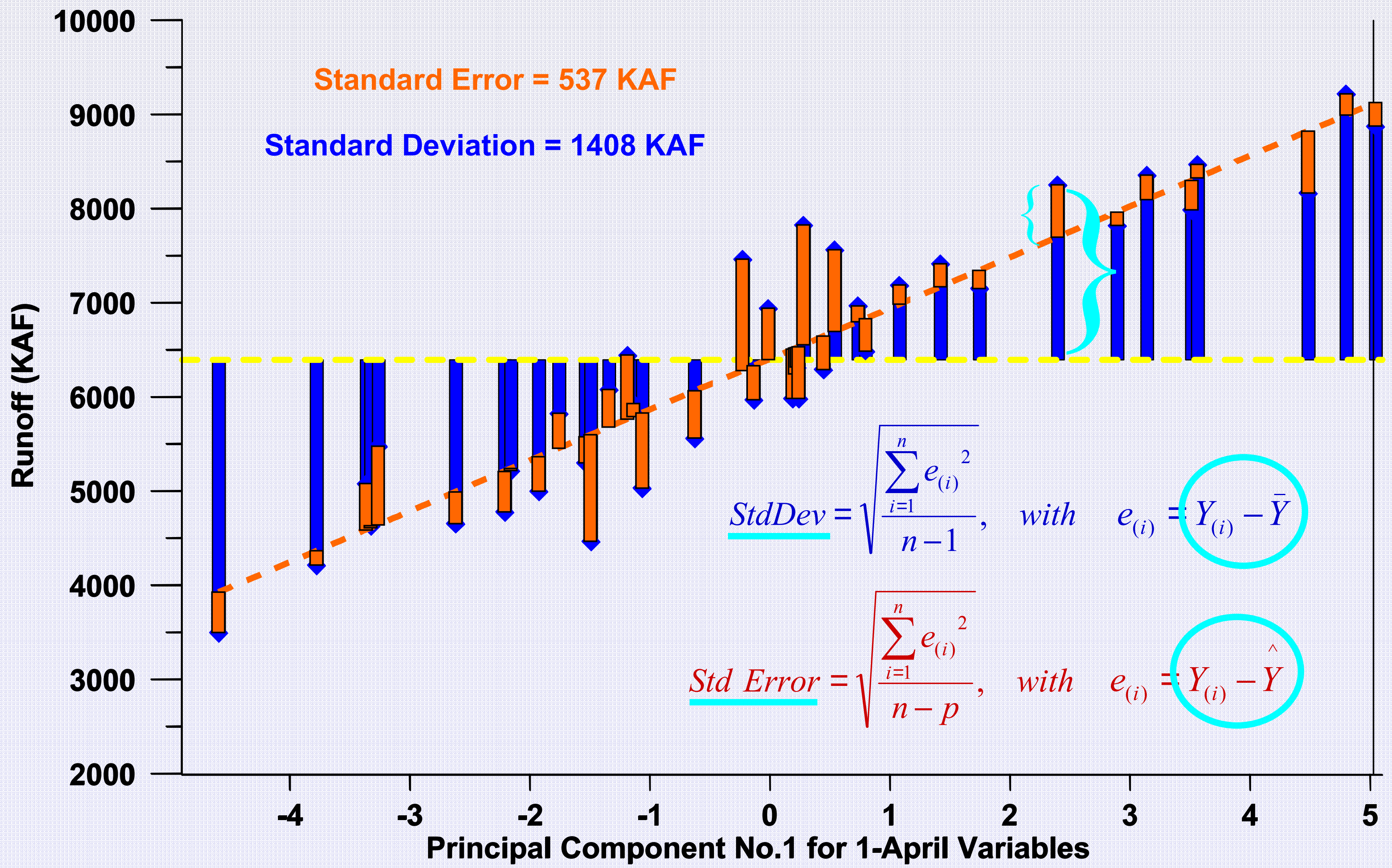
- ***Variable Selection (Which components do I keep?)***
  - **Component retention criteria:**
    - **Eigenvalue: provides the proportion of variability of X variables contained in each PC**
    - **significant R-squared: indicates the variability in the Y variable explained by this PC in a linear model, i.e. the usefulness of this PC in predicting the Y variable**
    - **significant *Beta:* reject this component when the regression coefficient is indistinguishable from zero.**
    - **sign of *Beta*: be wary of this component if the sign is negative (applies to water supply forecasting)**

US Army Corps
of Engineers
Northwestern Division

Comparing and Evaluating Models

- Be cautious of statistical models that have too many variables in comparison to the number of observations (years of data).
- Fitting too many variables leads to a model that is "overfit", i.e. is not parsimonious. Overfit models usually produce poor forecasts!
- Both the Adjusted R-square and Standard Error statistics are useful in comparing models, as they include a "degrees-of-freedom adjustment" to account for the number of coefficients used to fit the regression model to the data.

Standard Error



$$StdDev = \sqrt{\frac{\sum_{i=1}^{n} e_{(i)}^{2}}{n-1}}, \quad with \quad e_{(i)} = Y_{(i)} - \bar{Y}$$

$$Std\ Error = \sqrt{\frac{\sum_{i=1}^{n} e_{(i)}^{2}}{n-p}}, \quad with \quad e_{(i)} = Y_{(i)} - \hat{Y}$$

US Army Corps
of Engineers
Northwestern Division

Model Comparison: Validation Statistics

- Calibration statistics reflect the errors of the model optimized to fit to a given set of data.
- *Adjusted R-Square* and *Standard Error* are both statistics of the calibration model.
- Validation statistics reflect the errors of the calibrated model being applied to data not used in the calibration.
- Calibration statistics tend to be overly optimistic. Forecast models are best suited to be evaluated and compared based on validation statistics.

US Army Corps
of Engineers
Northwestern Division

Model validation

- **Split-Sample validation**

Calibration Data	Validation Data
Fit your model to this data	Compute your error statistics using this data

- **Leave-one-out validation**

X X X Calibration Data

est est est

US Army Corps
of Engineers
Northwestern Division

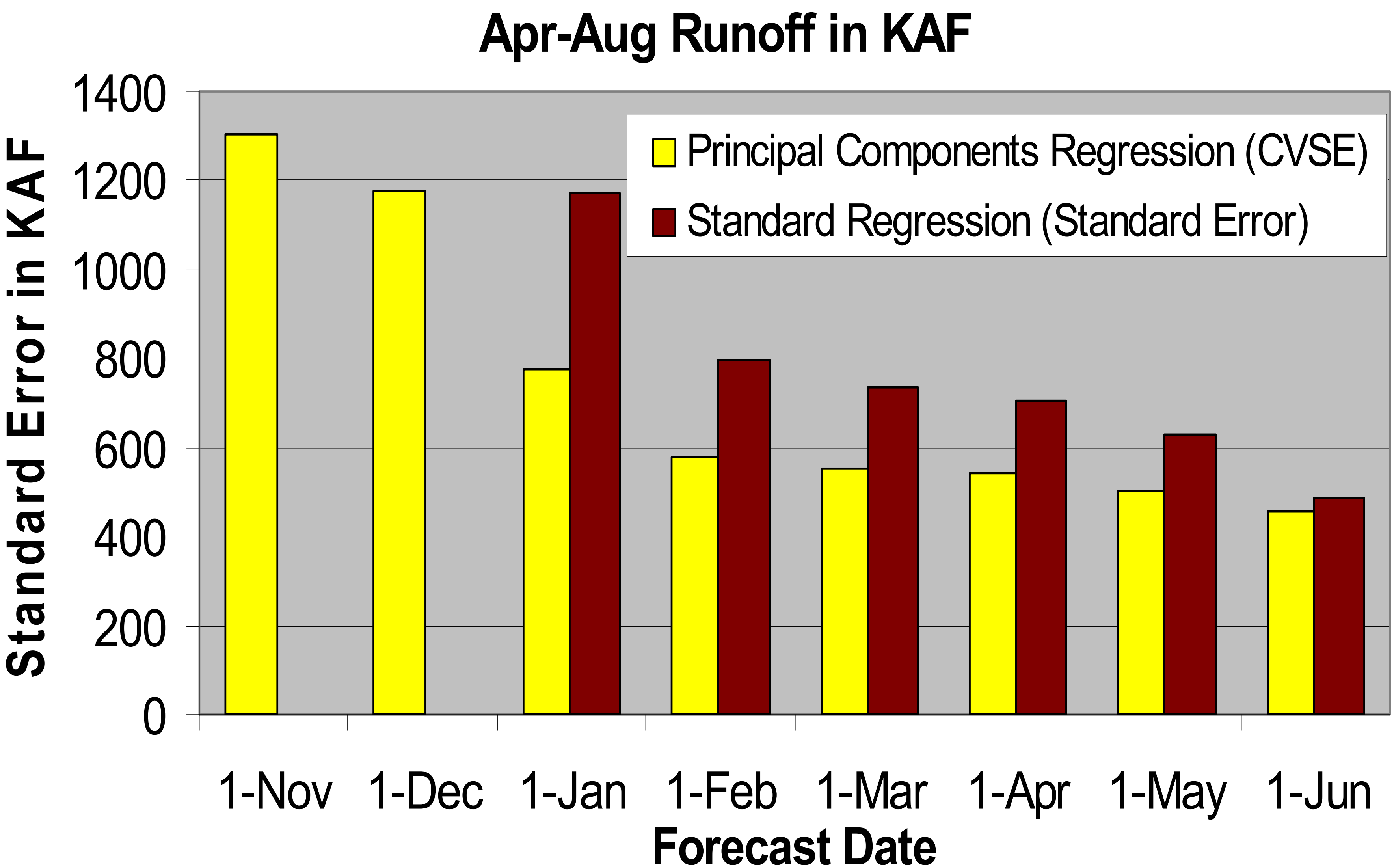
Cross-Validation Standard Error

$$CVSE = \sqrt{\frac{\sum e_{(i)}^{2}}{n-p}}$$

Where $e_{(i)}$ is the forecast error for the leave-one-out forecast of observation i

- Cross-validation standard error ("Jackknife" Std Err)
- CVSE supports model parsimony by including d-f adj.
- CVSE better indicator of how the model performs with data not used in the calibration, that it "hasn't seen yet"
- CVSE = PRESS statistic adjusted for degrees-of-freedom
- CVSE can be directly calculated from either Projection matrix or "Hat" matrix (calculated by NRCS PCREG)

Libby Water Supply Forecast using Principal Components Regression



US Army Corps
of Engineers
Northwestern Division

Endorsements

The following agencies use Principal Components regression in their Water Supply Forecasting procedures:

- *National Water and Climate Center, Natural Resources Conservation Service*
- *Northwestern Division, U.S. Army Corps of Engineers*
- *Northwest River Forecasting Center, National Oceanic and Atmospheric Administration*
- *Columbia River Treaty Operating Committee; Canadian and United States Entities*
- *Bonneville Power Administration, Dept of Energy*
- *BC Hydro*

US Army Corps
of Engineers
Northwestern Division

Questions?