Real-time Verification of Short-term Probabilistic Streamflow Forecasts

Is it possible?

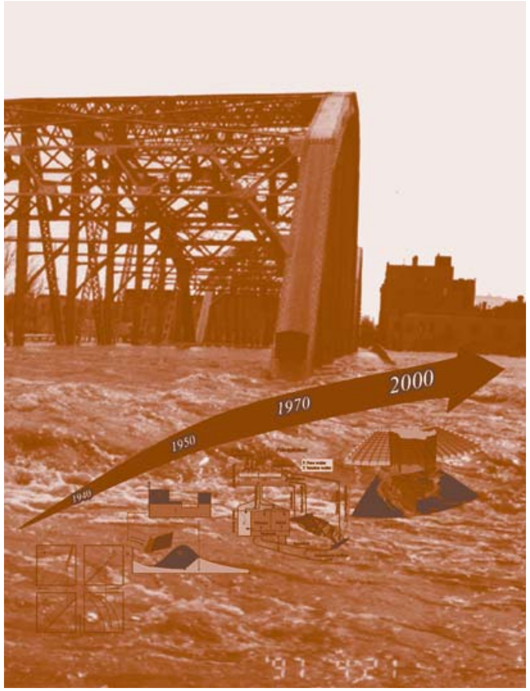
Bisher Imam
University of California, Irvine

Holly Hartmann
University of Arizona

Towards a New AHPS (NRC, 2006)

Limitation and Research Needs for ESP

“The AHPS approach to quantifying uncertainties in operational forecasts must be articulated”

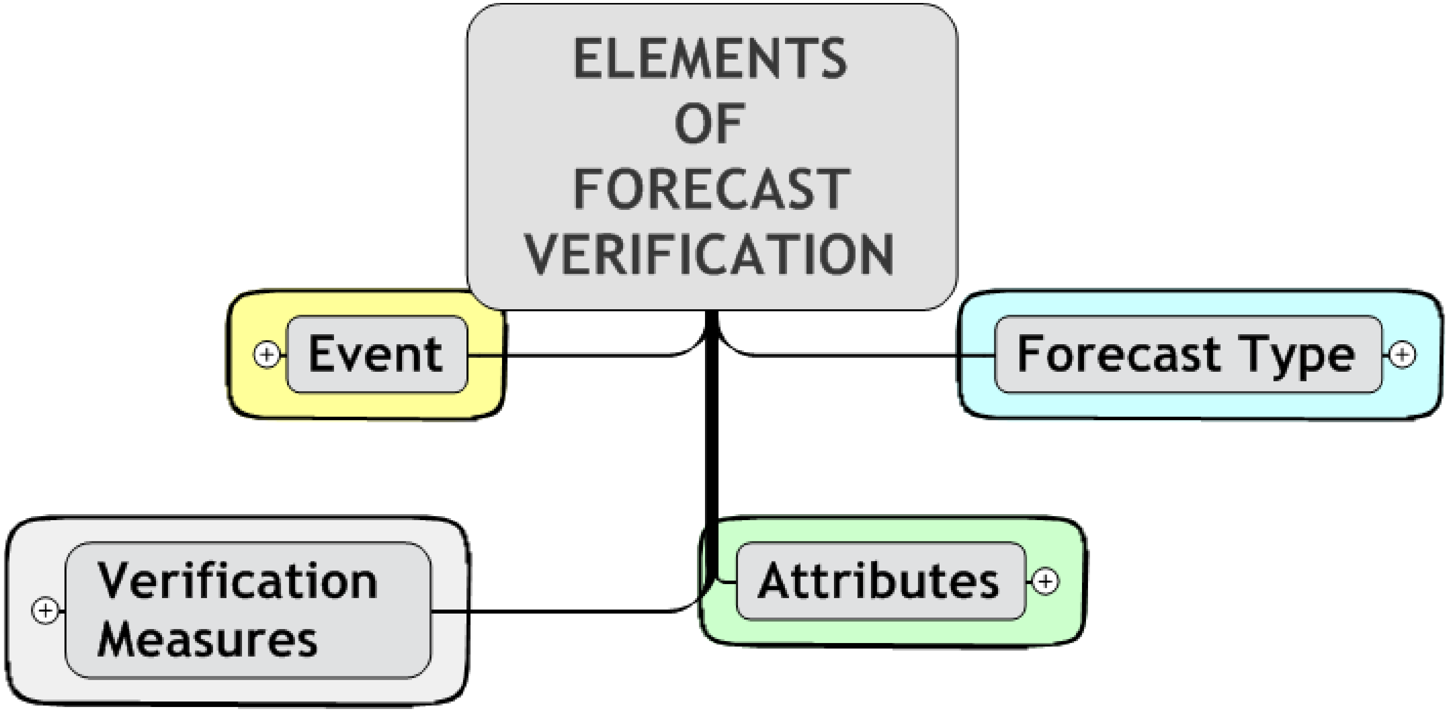


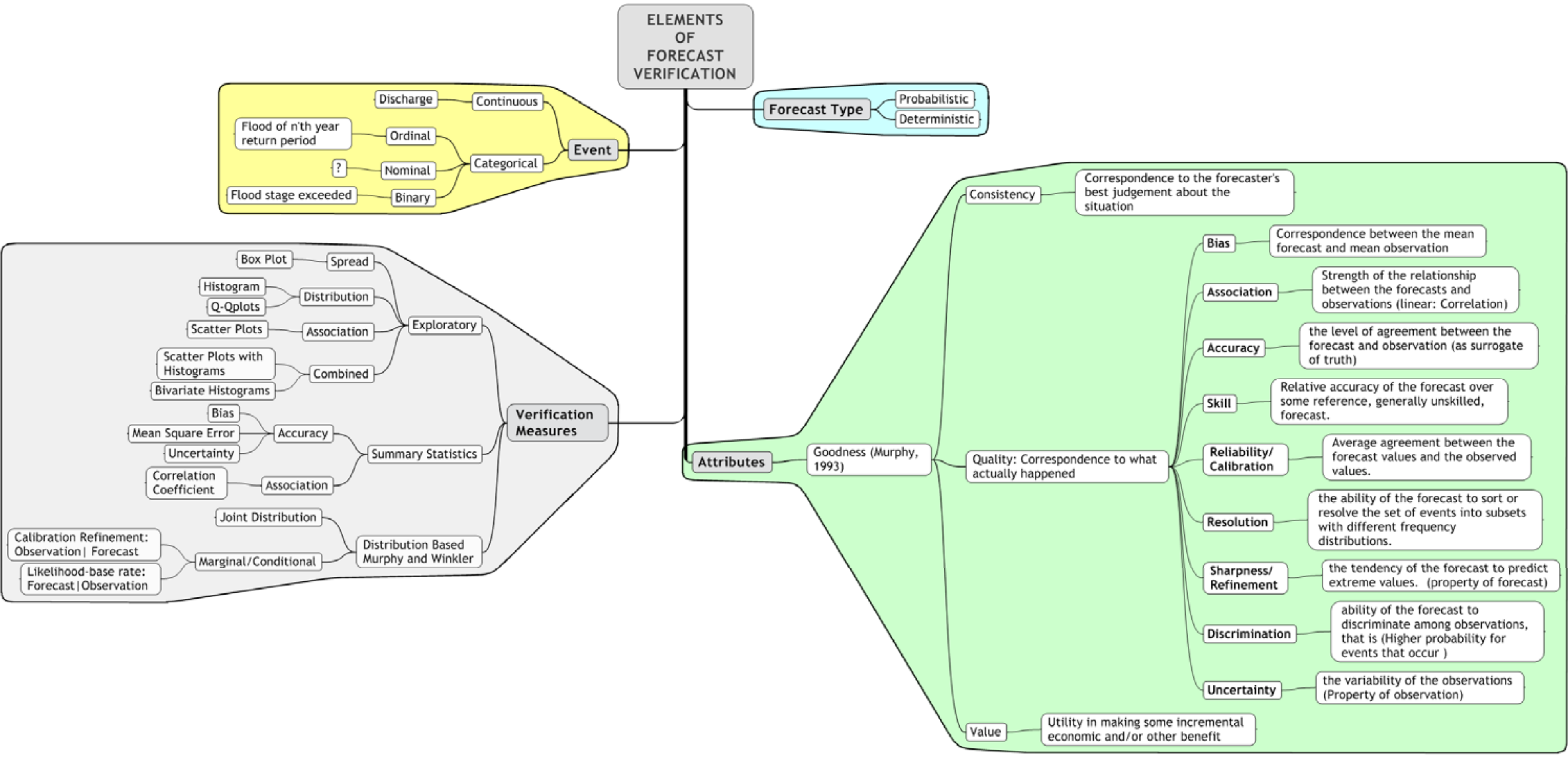
Limitations and Research Needs for Verification in the NWSRFS

“Unlike meteorological forecasts, little is known about hydrologic forecasts and actual river forecast skills”

Objectives

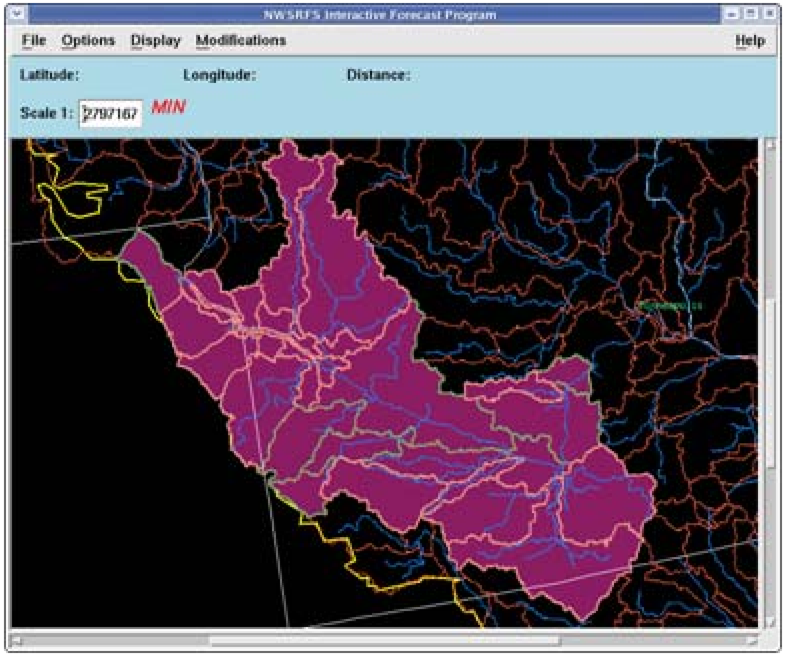
- Identify suitable measures for real-time ESP forecast verification
- Propose operational procedure for ESP forecast verification
- Propose examples of possible screens that can be integrated into the system


ELEMENTS OF FORECAST VERIFICATION
Event
Forecast Type
Verification Measures
Attributes

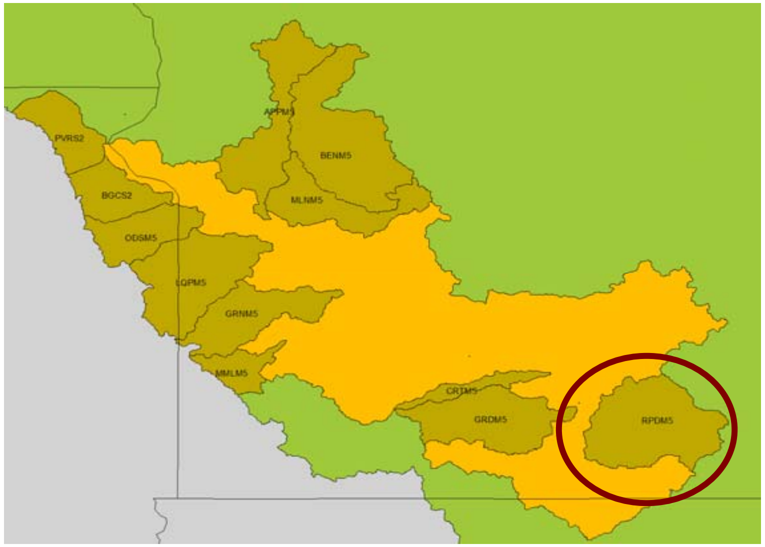


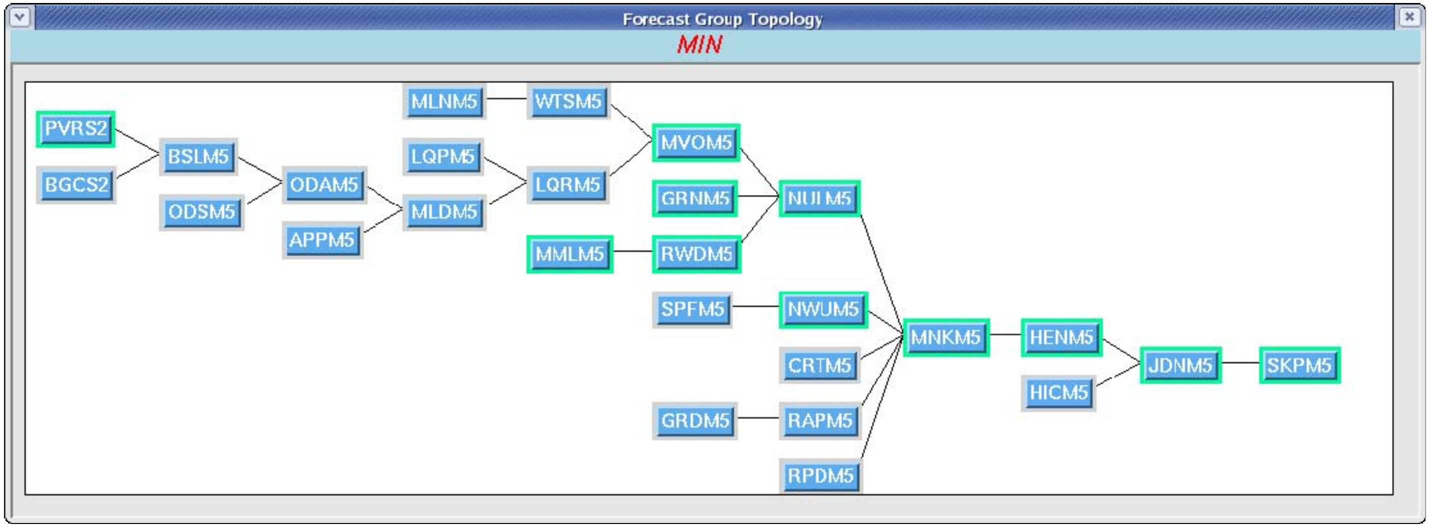
Focus on Forecasters

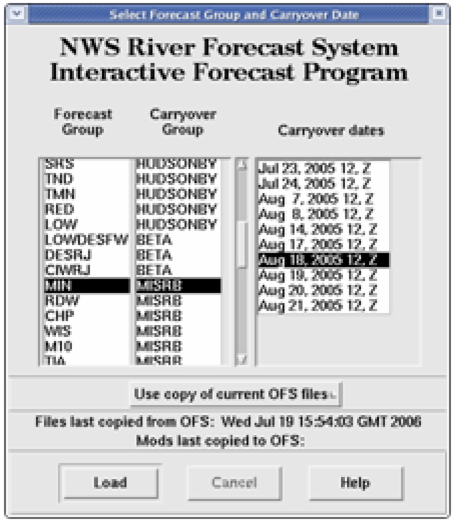
Examples Site



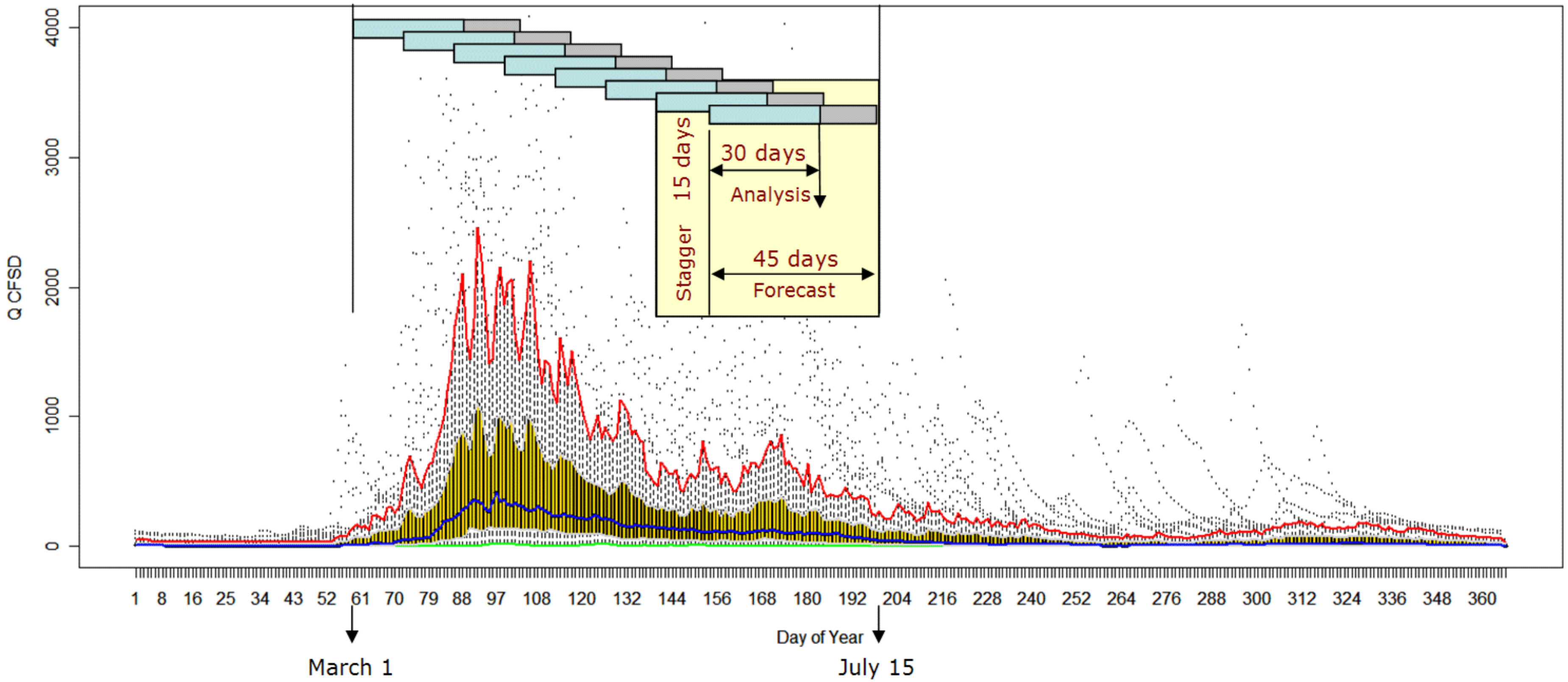
Le Sueur River Near Rapidan MN [RPDM5] (USGS station 05320500)





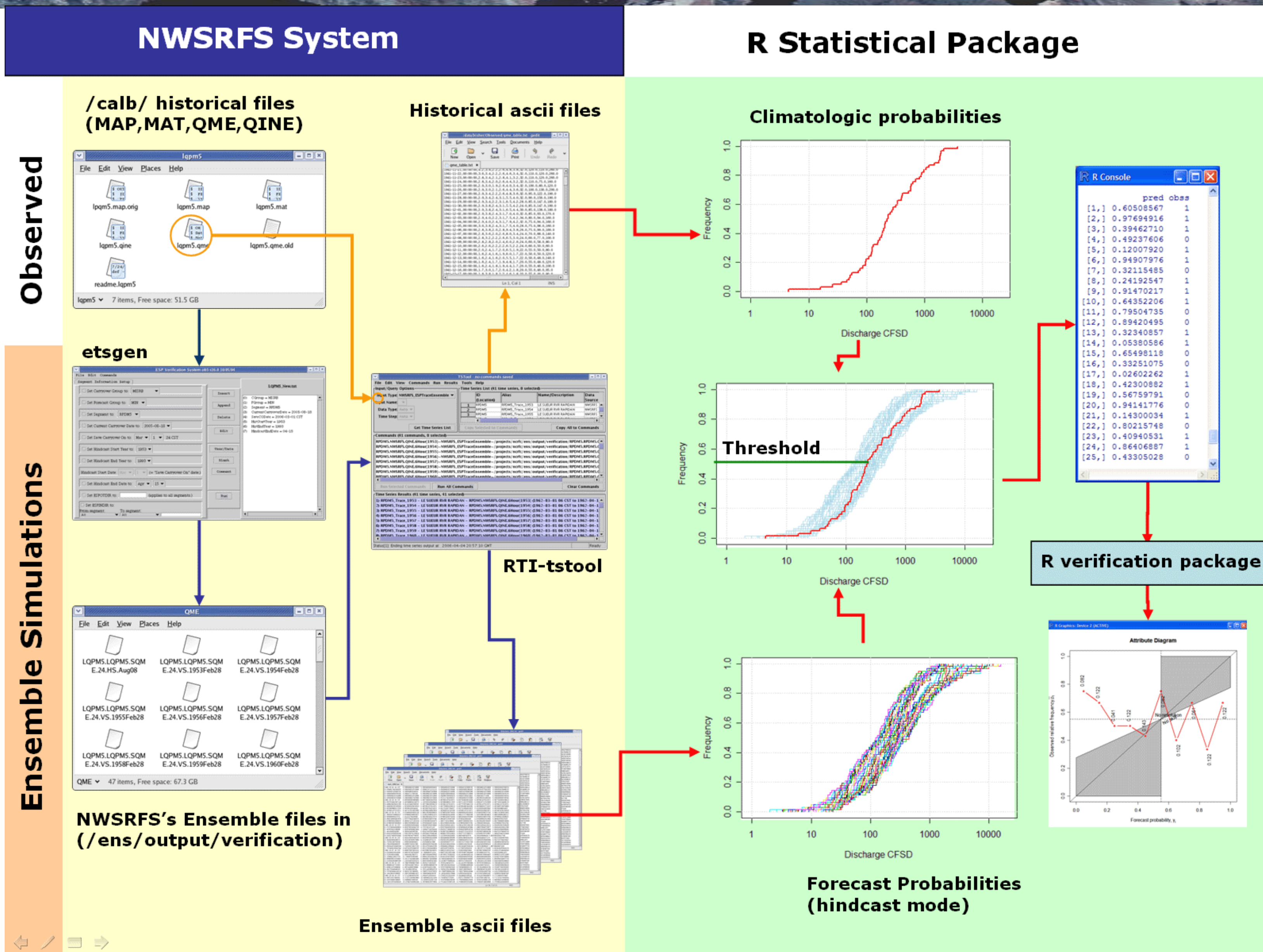


Forecasting Period and Windows

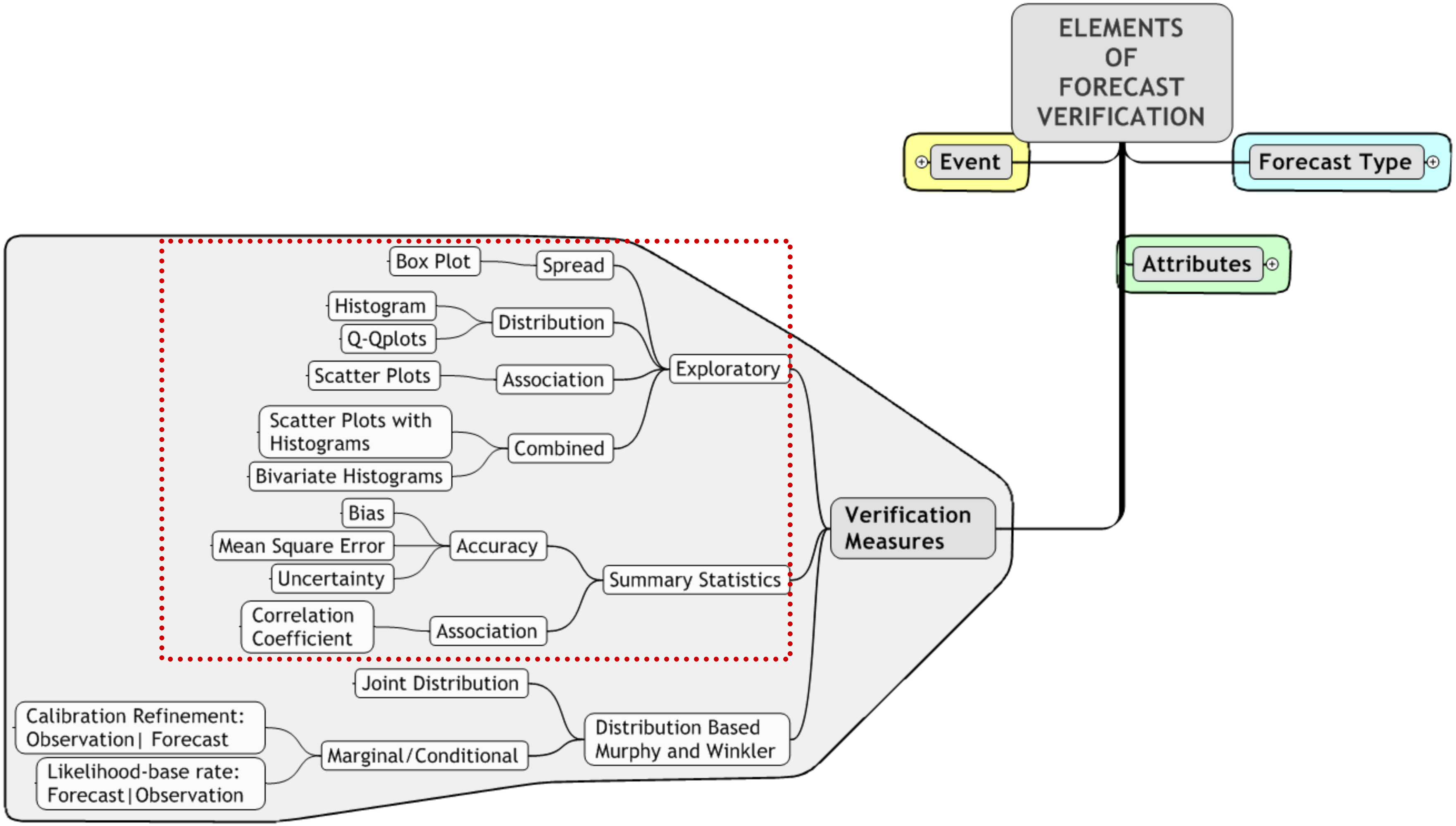


Hindcasts: 1953-1993→ 45 Years
Ensembles:1952-1953 → 46 Ensembles

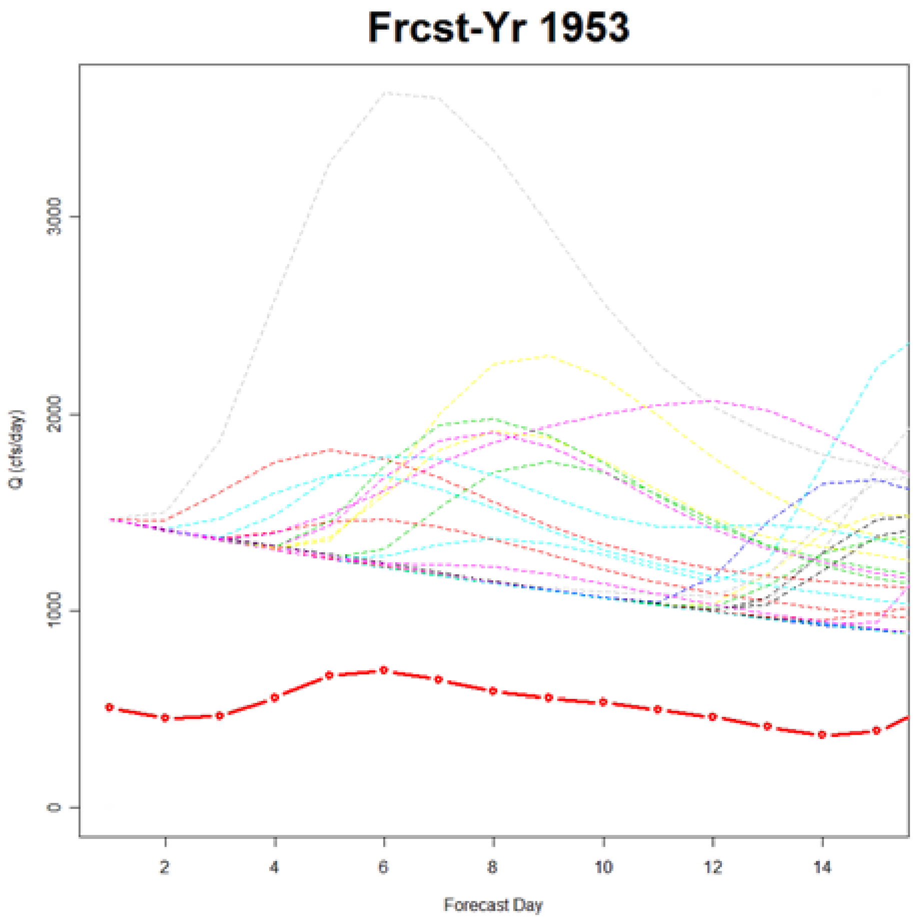
Procedure

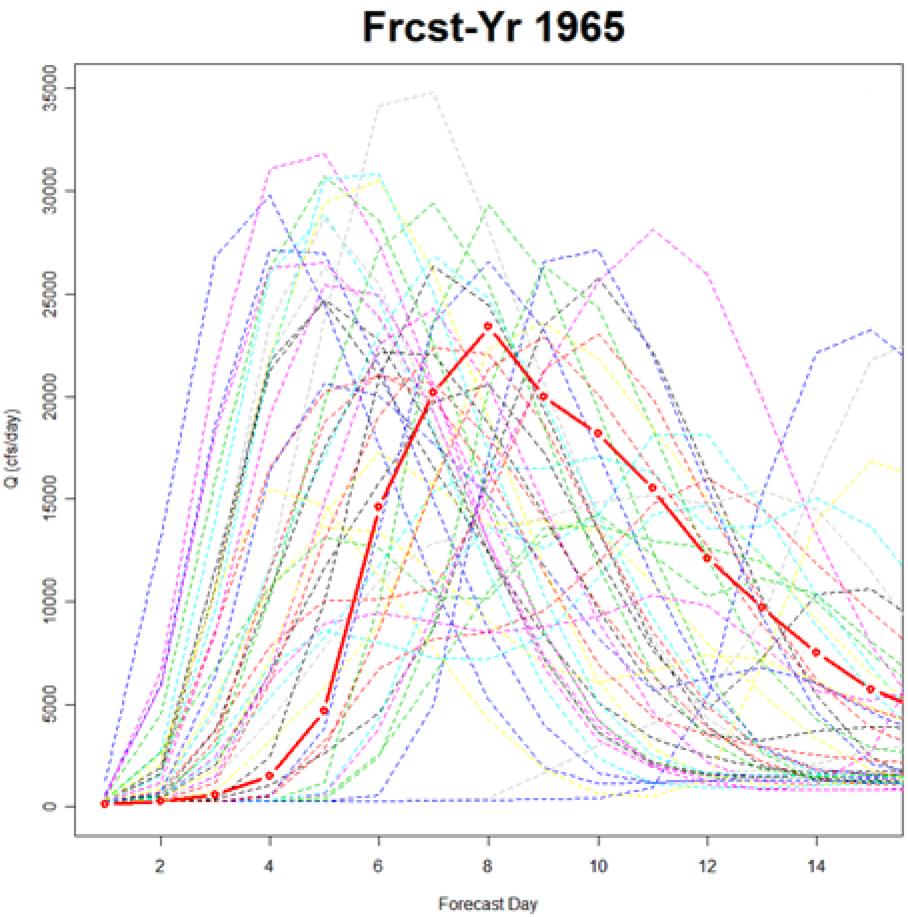


NWSRFS System
R Statistical Package
Observed
Ensemble Simulations
/calb/ historical files (MAP,MAT,QME,QINE)
Historical ascii files
Climatologic probabilities
Frequency
Discharge CFSD
etsgen
RTI-tstool
Threshold
R Console
pred obss
[1,] 0.60508567 1
[2,] 0.97694916 1
[3,] 0.39462710 1
[4,] 0.49237606 0
[5,] 0.12007920 1
[6,] 0.94907976 1
[7,] 0.32115485 0
[8,] 0.24192547 1
[9,] 0.91470217 1
[10,] 0.64352206 1
[11,] 0.79504735 0
[12,] 0.89420495 0
[13,] 0.32340857 1
[14,] 0.05380586 1
[15,] 0.65498118 0
[16,] 0.33251075 0
[17,] 0.02602262 1
[18,] 0.42300882 1
[19,] 0.56759791 0
[20,] 0.94141776 0
[21,] 0.14300034 1
[22,] 0.80215748 0
[23,] 0.40940531 1
[24,] 0.86406887 1
[25,] 0.43305028 0
R verification package
Attribute Diagram
NWSRFS's Ensemble files in (/ens/output/verification)
Ensemble ascii files
Forecast Probabilities (hindcast mode)

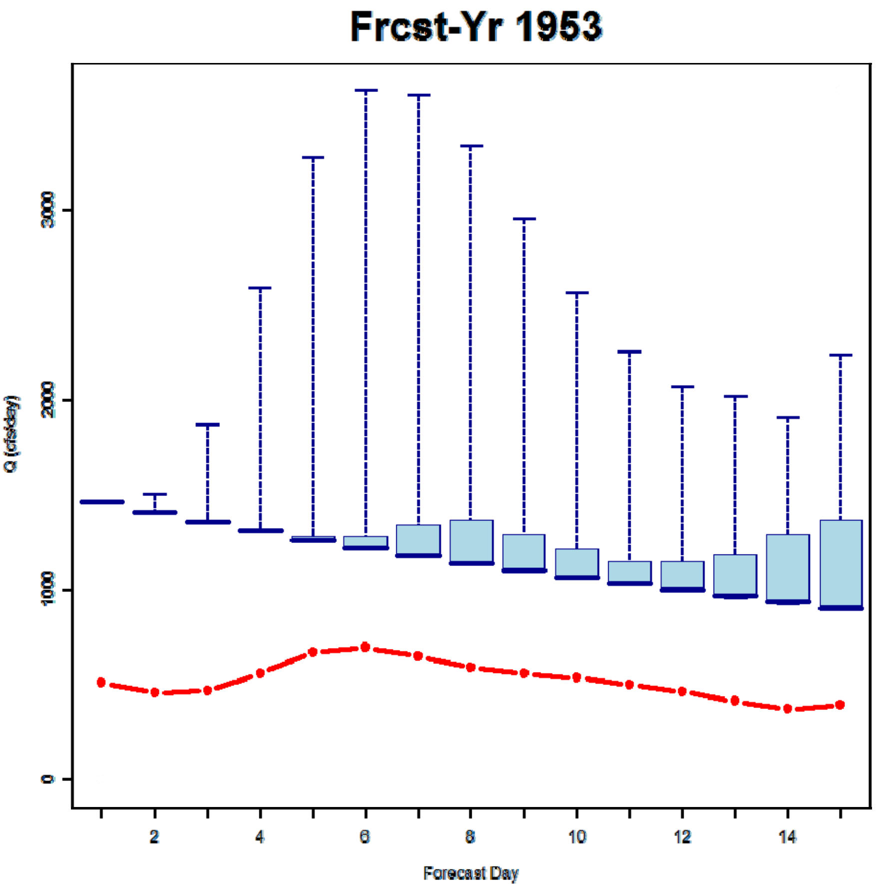

ELEMENTS OF FORECAST VERIFICATION
Event
Forecast Type
Attributes
Verification Measures
Exploratory
Spread
Box Plot
Distribution
Histogram
Q-Qplots
Association
Scatter Plots
Combined
Scatter Plots with Histograms
Bivariate Histograms
Summary Statistics
Accuracy
Bias
Mean Square Error
Uncertainty
Association
Correlation Coefficient
Distribution Based Murphy and Winkler
Joint Distribution
Marginal/Conditional
Calibration Refinement: Observation| Forecast
Likelihood-base rate: Forecast|Observation

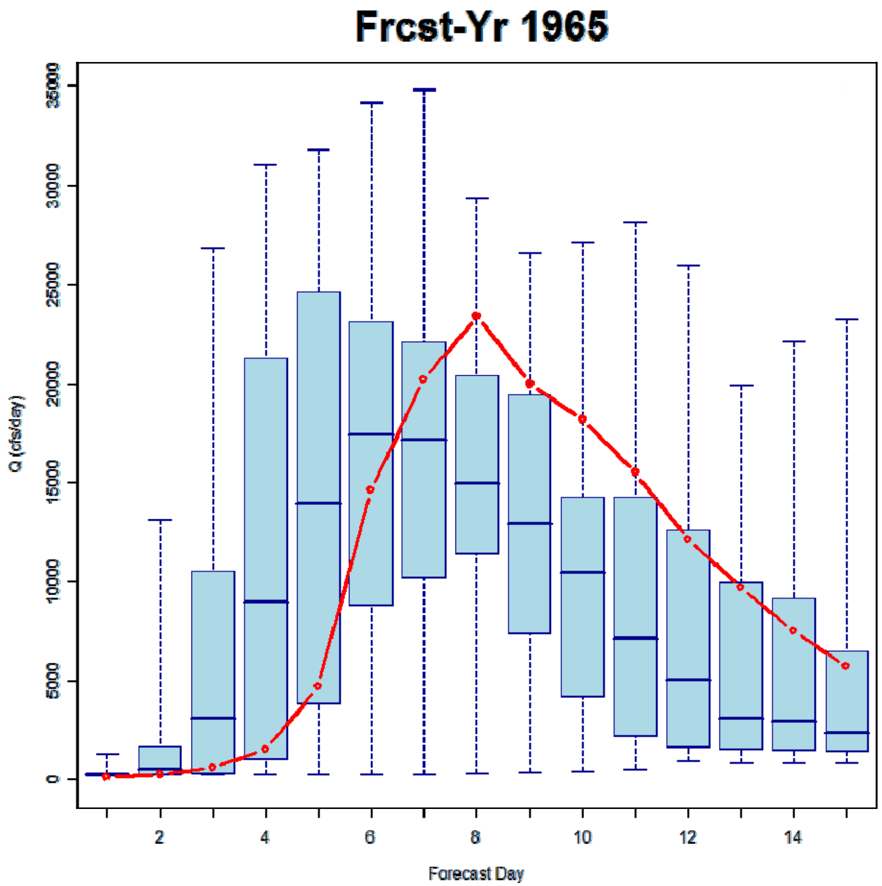
Exploratory Approaches; The Ensemble Plots



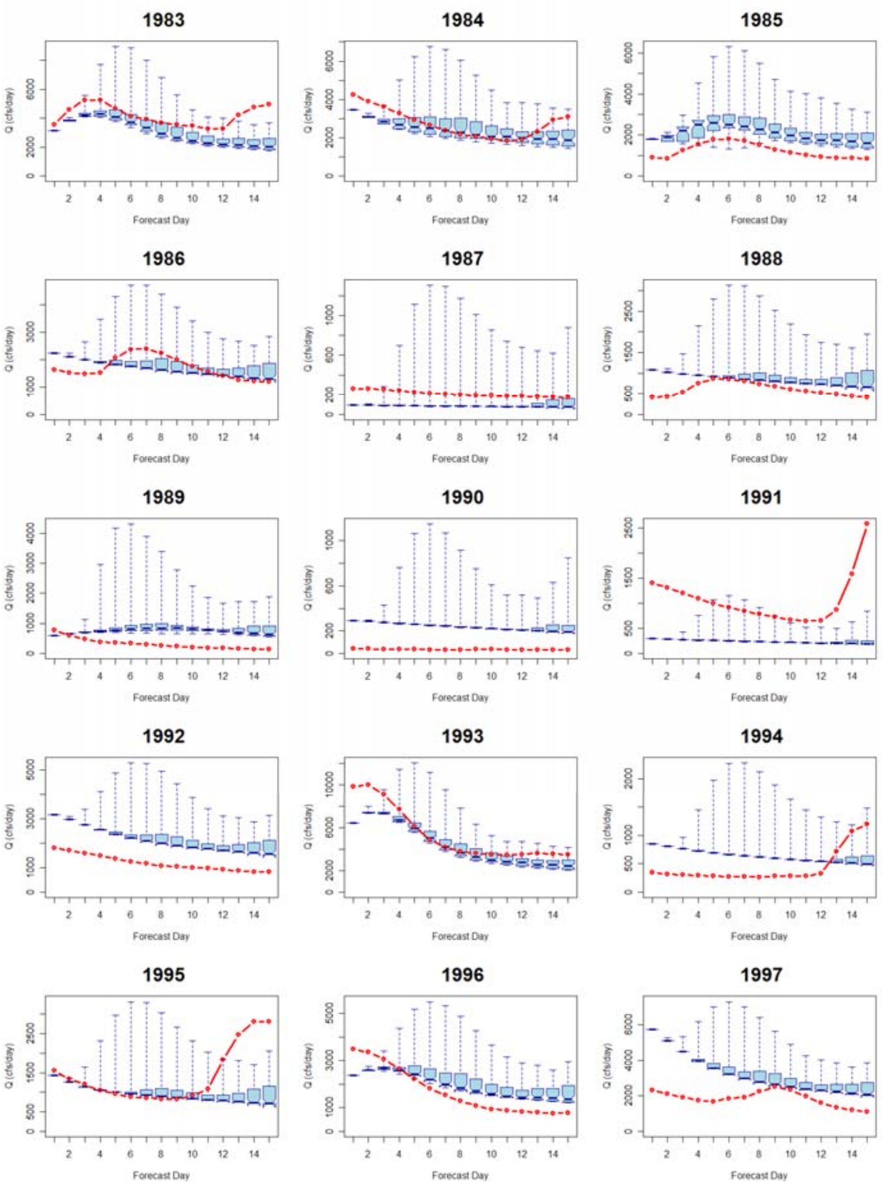


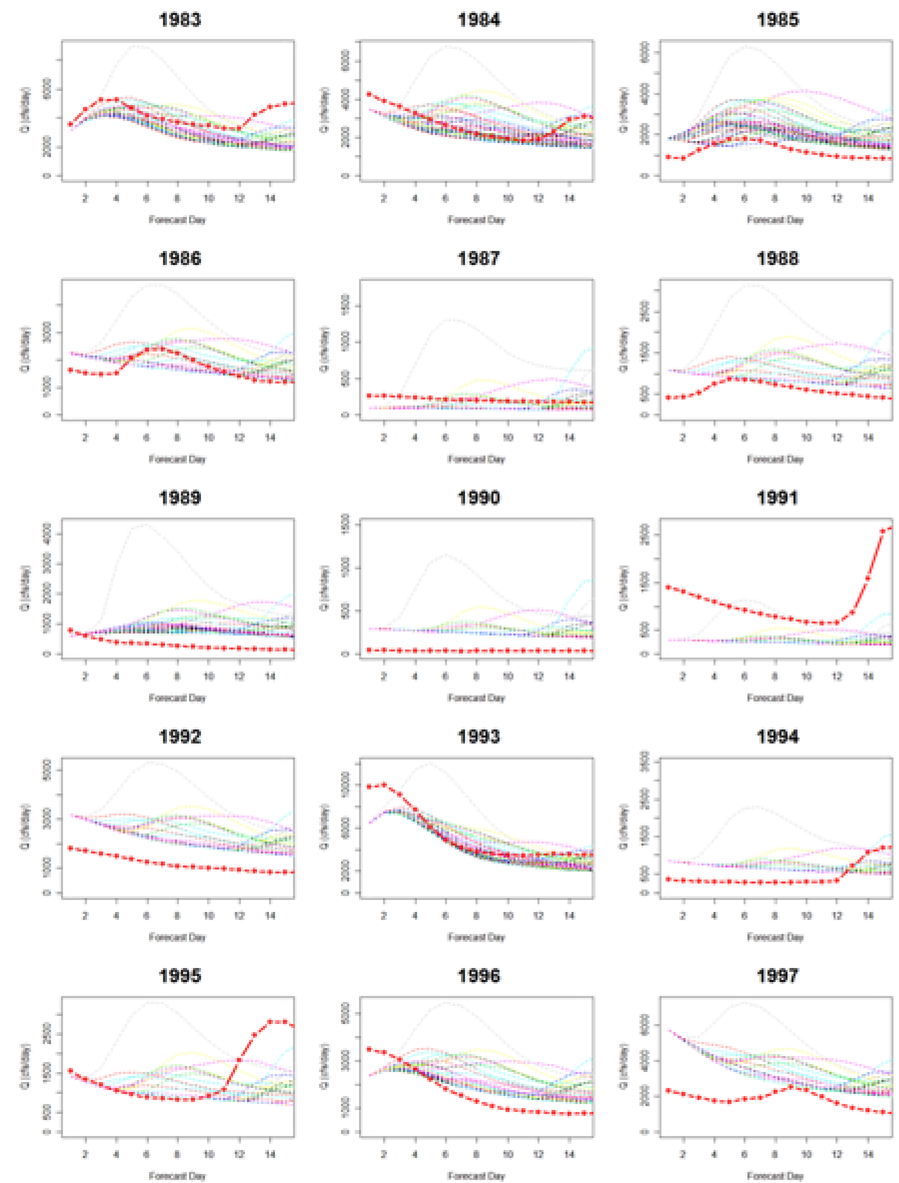
Multiple Forecasts



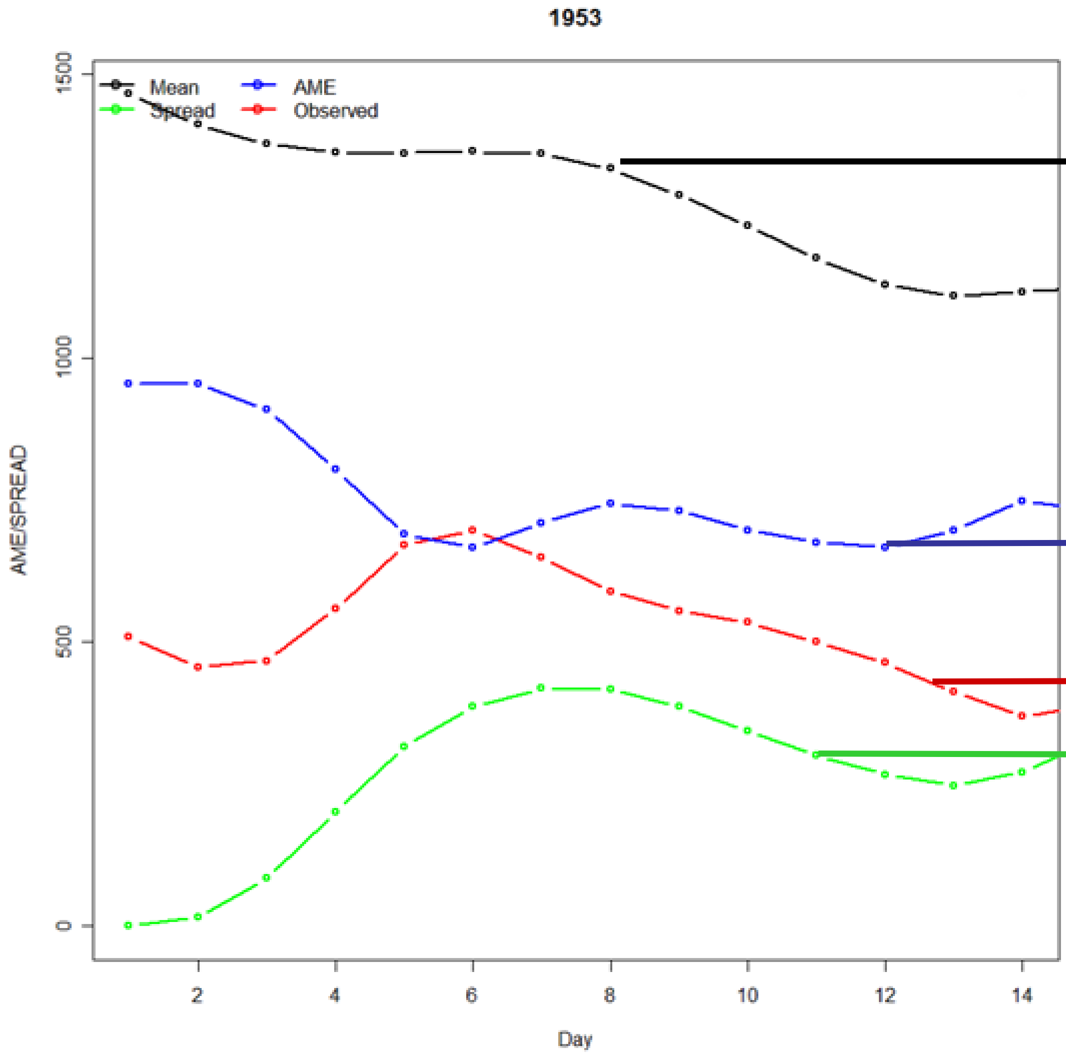


Multiple Forecasts





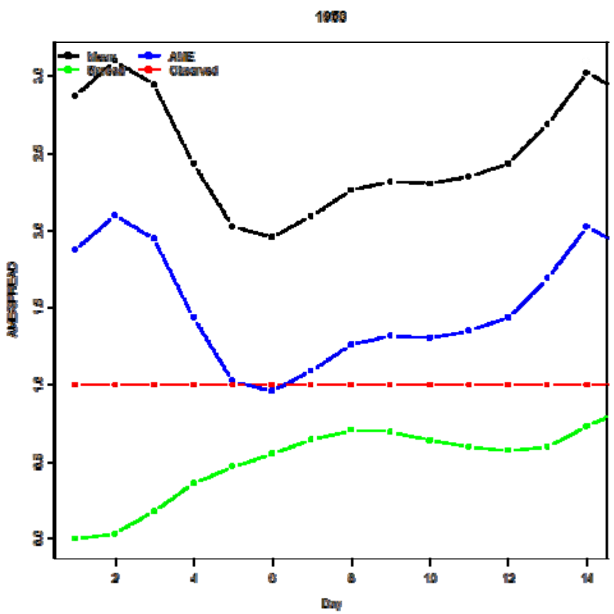
Consistency: Error and Spread

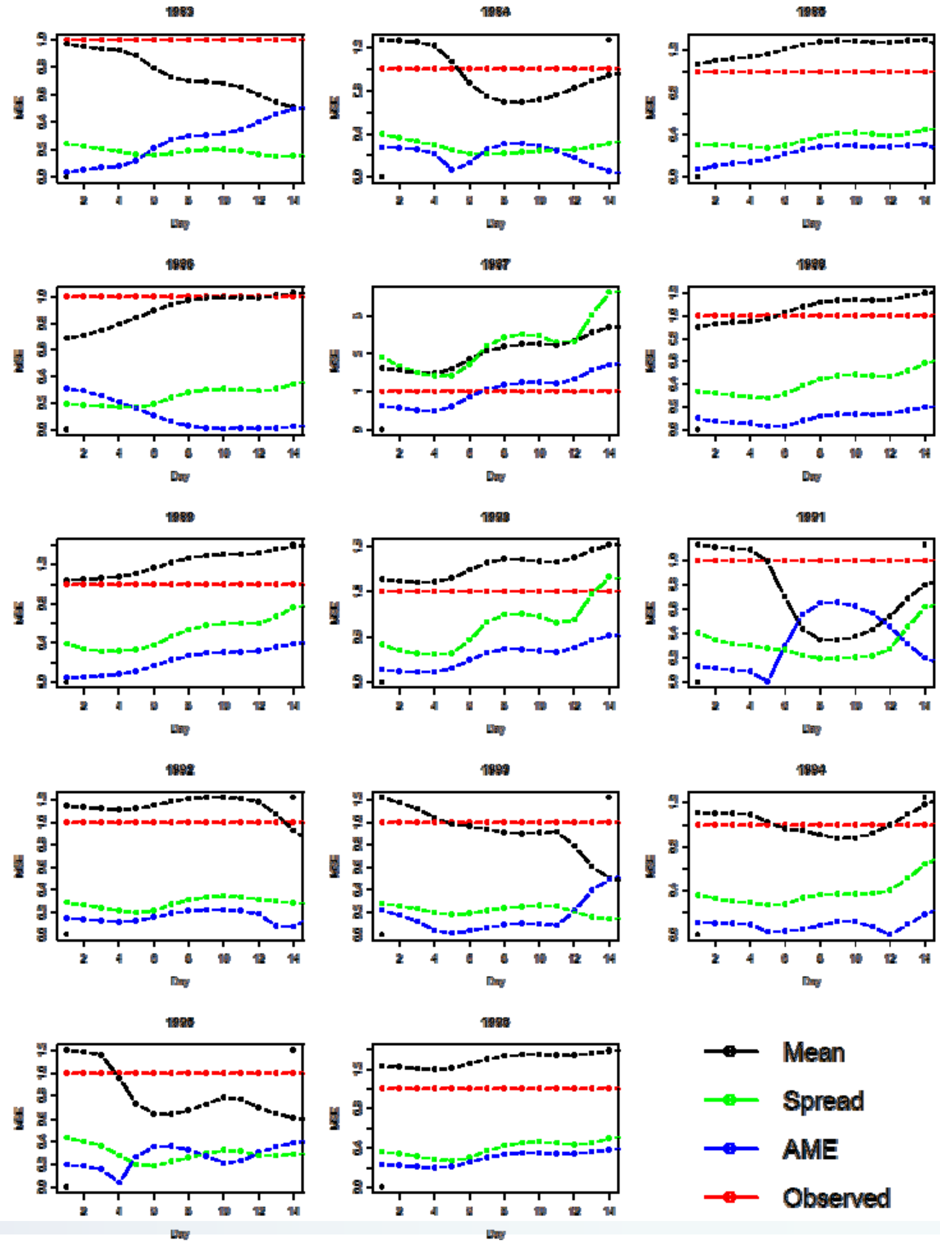


$$\left|\mu_{Ens} - Obs\right|$$

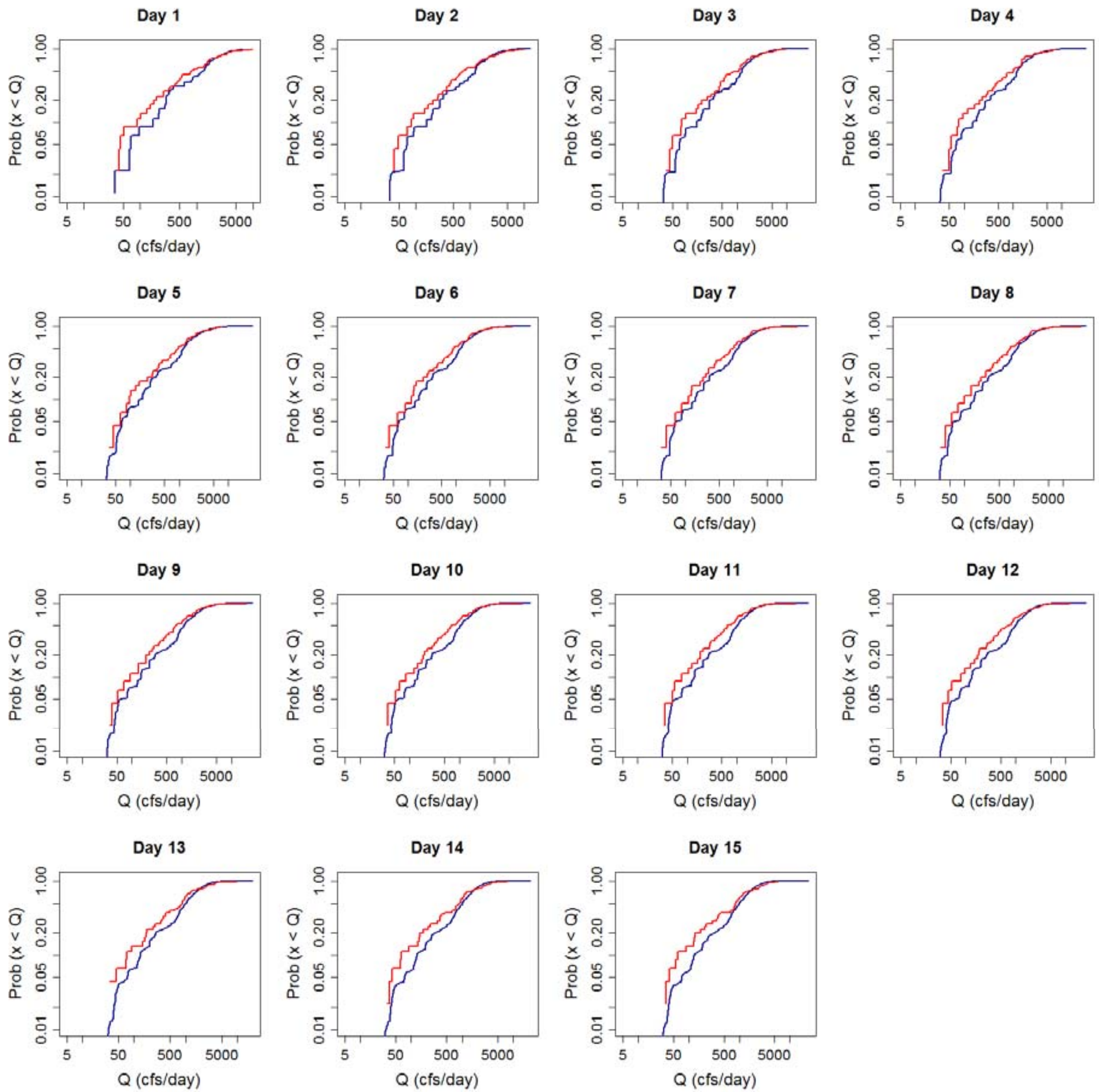
$$\sigma_{Ens}$$

Normalized Error/Spread Plot



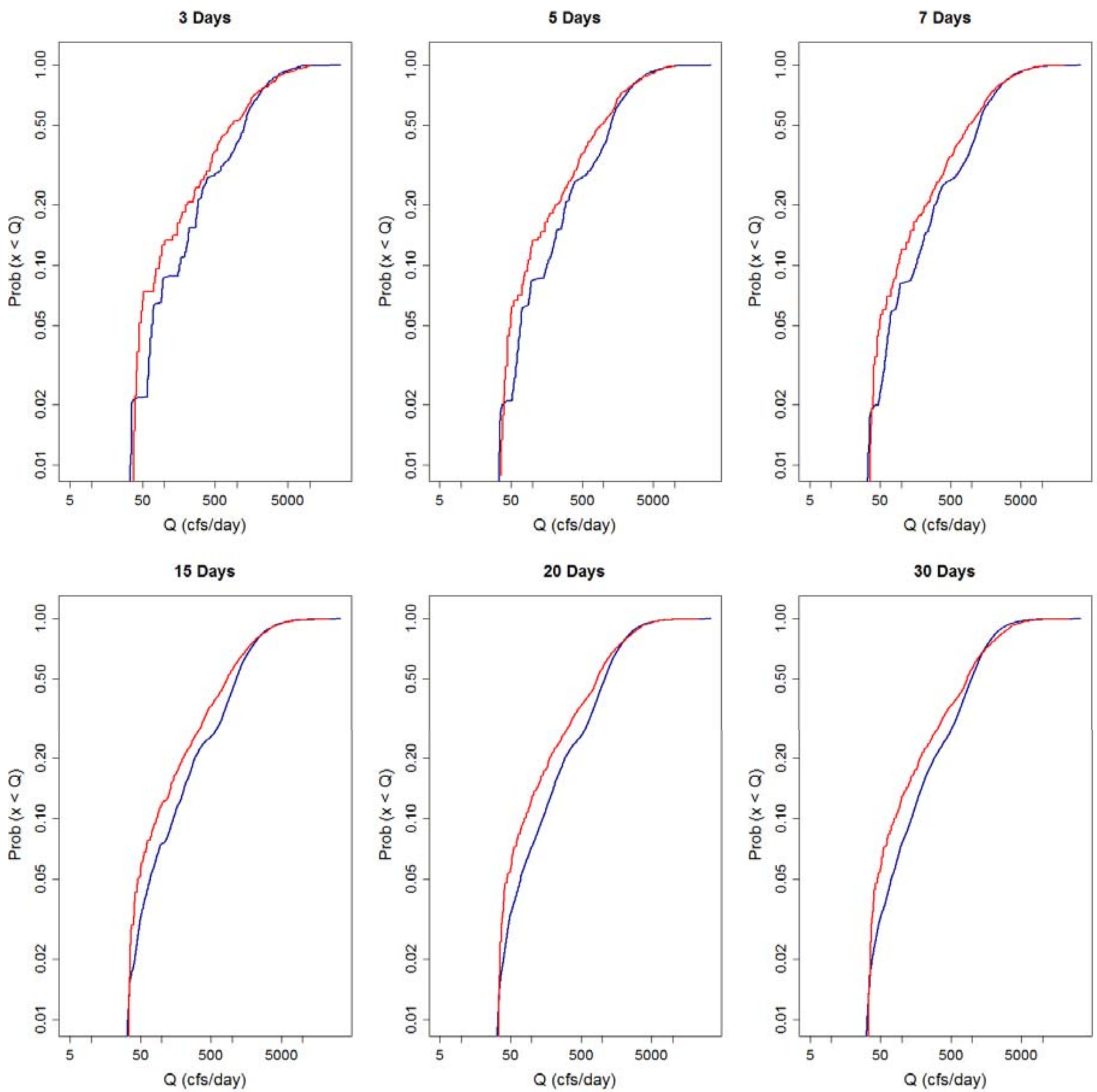


Larger Data Set: All Ensembles Considered



Verifying Data Augmentation: Multiple Days

What is a reasonable forecast variable



Distribution oriented measures

CR Factorization $h(f,o) = y(o \mid f)\,p(f)$

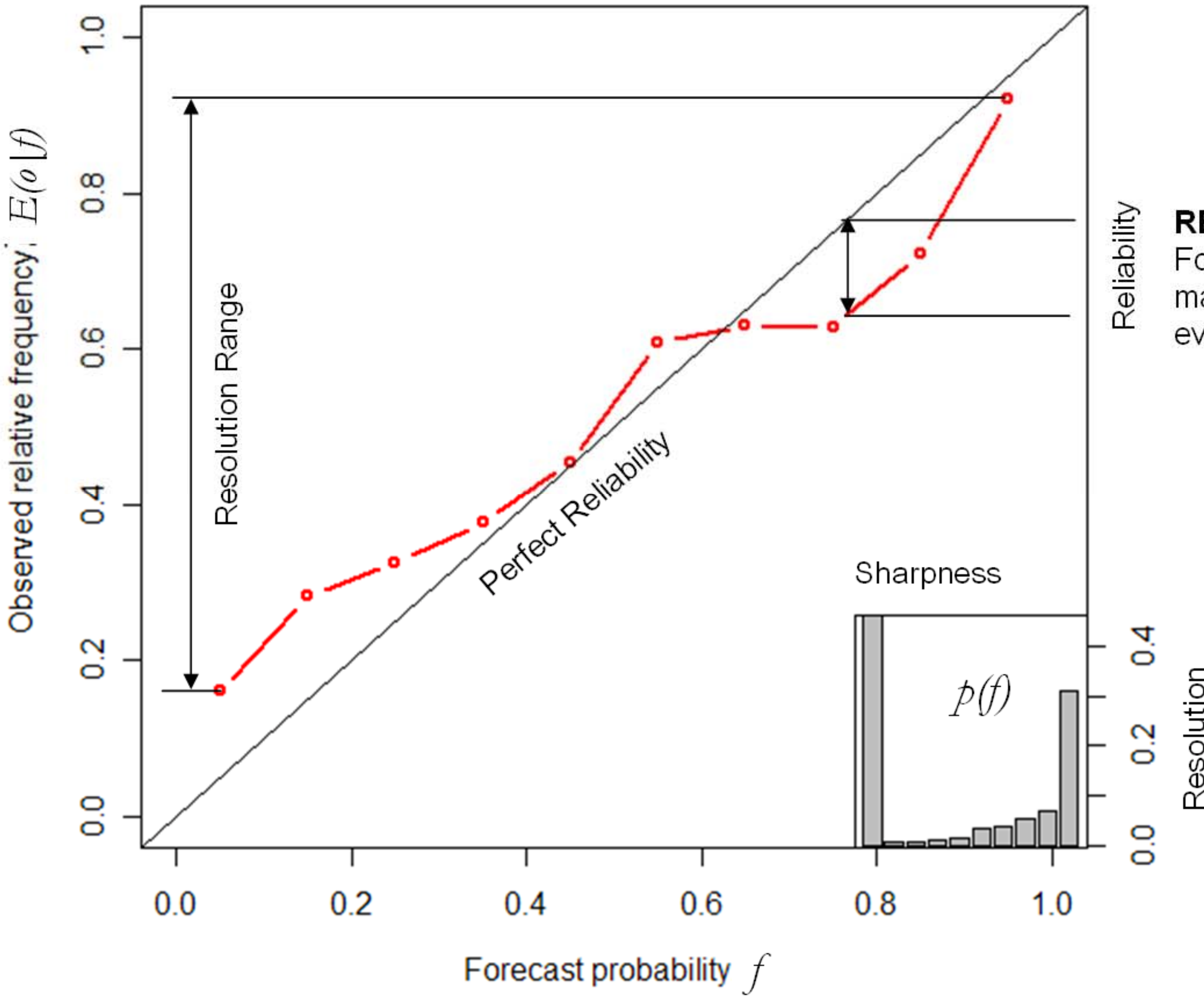
LBR Factorization $h(f,o) = r(f \mid o)\,t(o)$

$$f_i(q^*) = P\left\{ Q_i \le q^* \mid \alpha_i \right\}$$

where f_i is the probability forecast, and α_i *is the initial condition*.
observation variable $o_i(q^*)$ as

$$o_i(q^*) = \begin{cases} 1 & \text{if } Q_i \le q^* \\ 0 & \text{if } Q_i > q^* \end{cases}$$

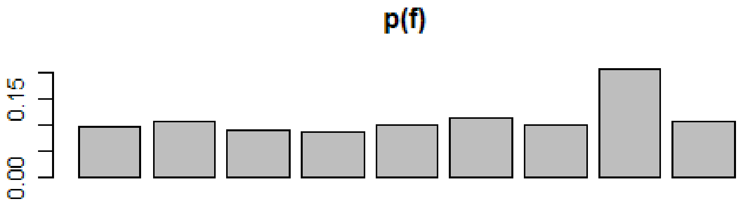
Reliability/Resolution Diagram



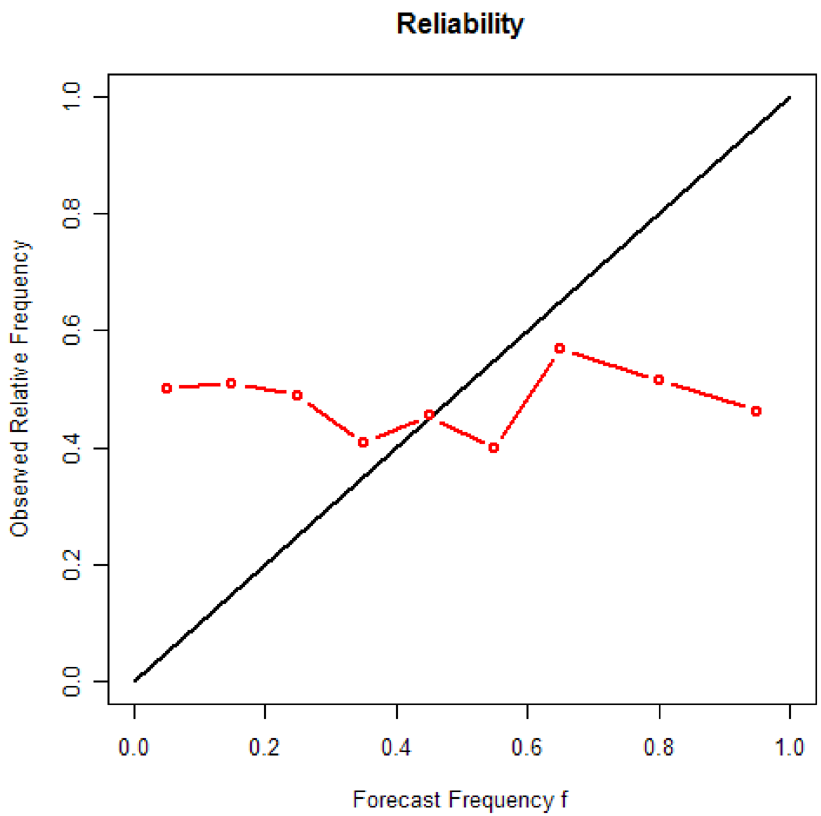
RELIABILITY:
Forecast probabilities for given event match observed frequencies of that event (with given prob. forecast)

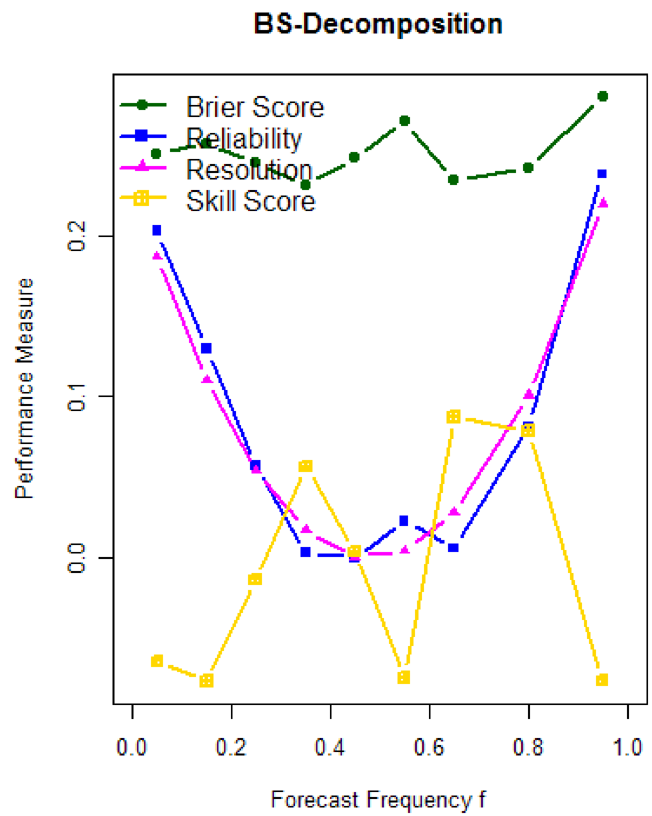
RESOLUTION:
Occurrence and non-occurrence of event is well resolved by forecast system

Random Forecast

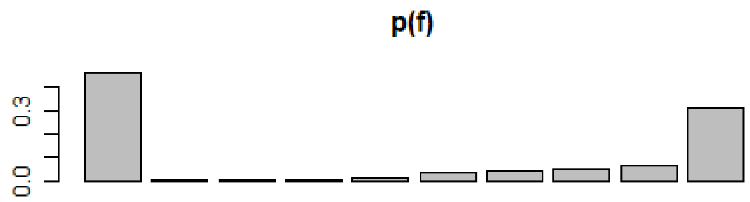


Overall Perfromance Measures		
Brier Score (BS)	=	0.3306
Brier Score - Baseline	=	0.2497
Skill Score	=	-0.324
Reliability	=	0.08339
Resolution	=	0.0025
Uncertaintity	=	0.2497

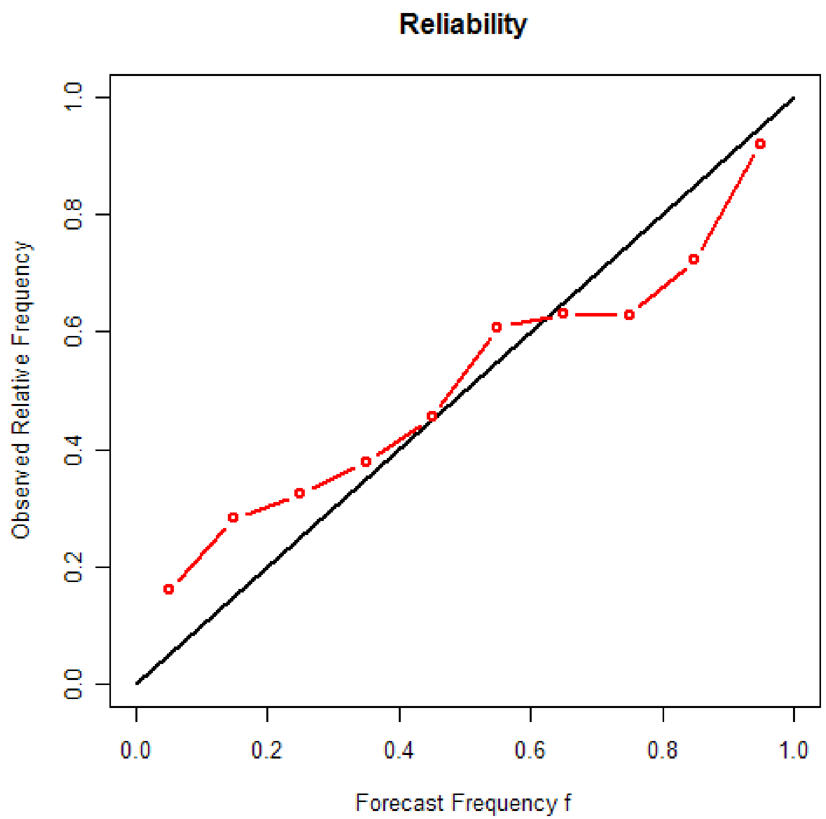


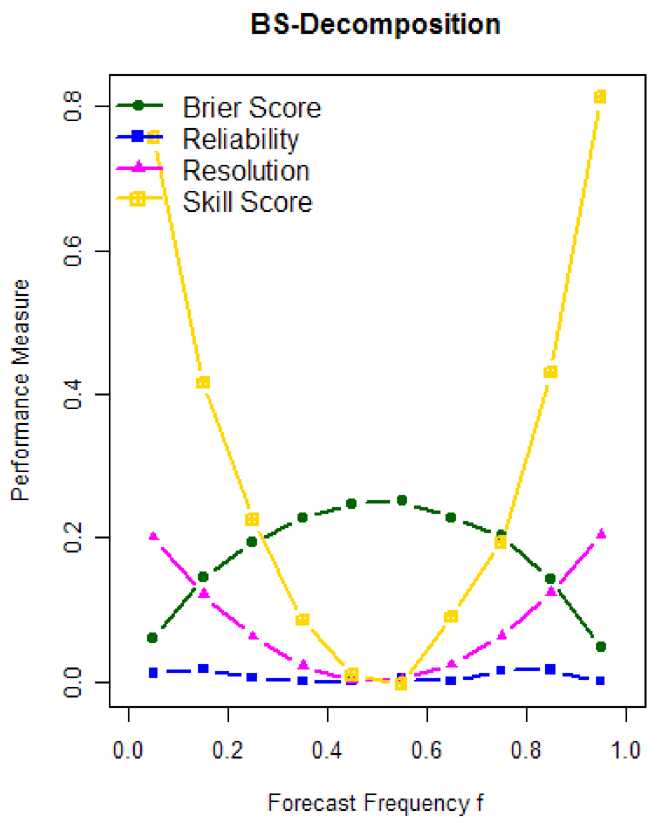


Real Forecast (Augmented 15 Days)

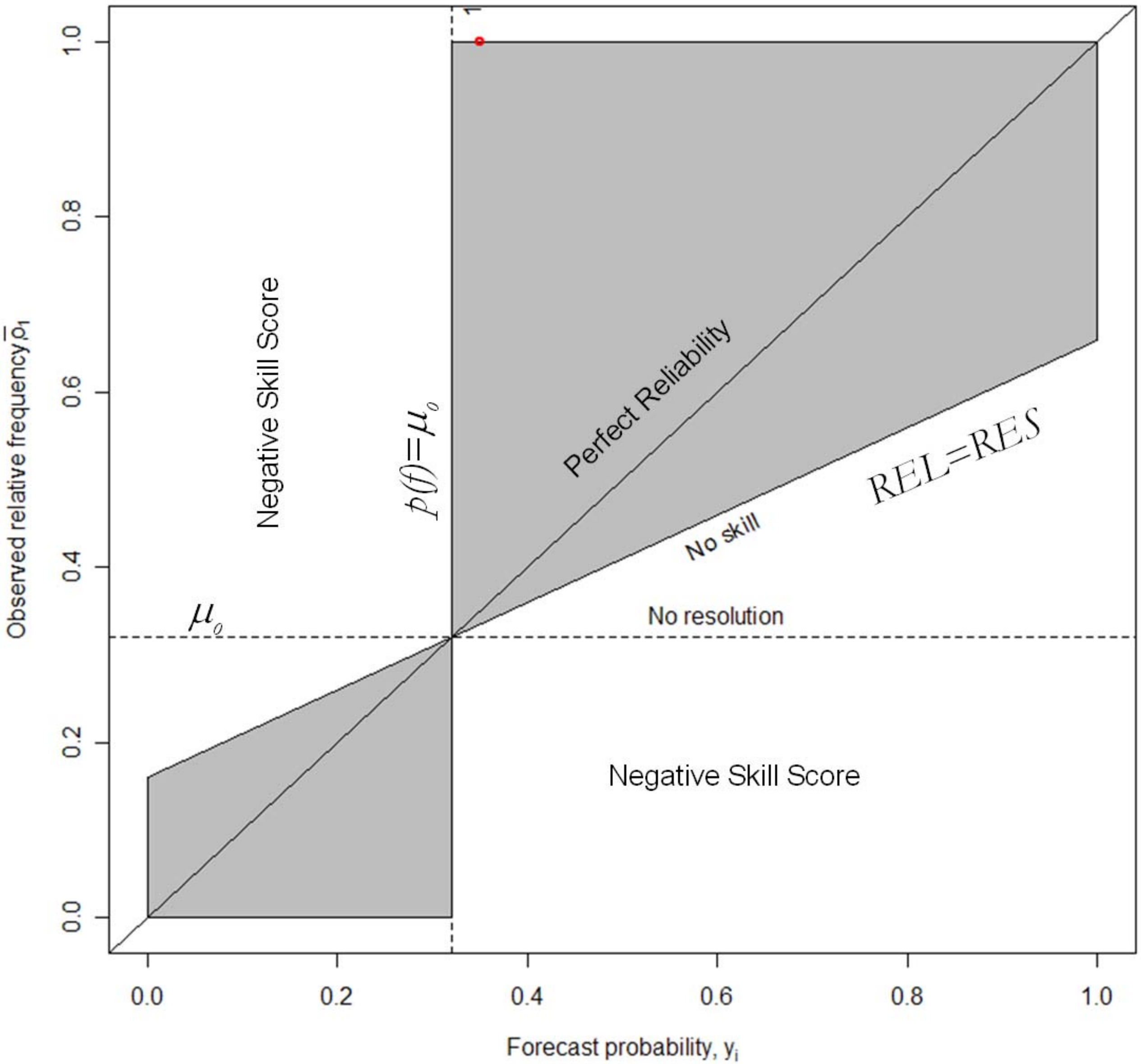


Overall Perfromance Measures		
Brier Score (BS)	=	0.1431
Brier Score - Baseline	=	0.25
Skill Score	=	0.4276
Reliability	=	0.007992
Resolution	=	0.1149
Uncertaintity	=	0.25

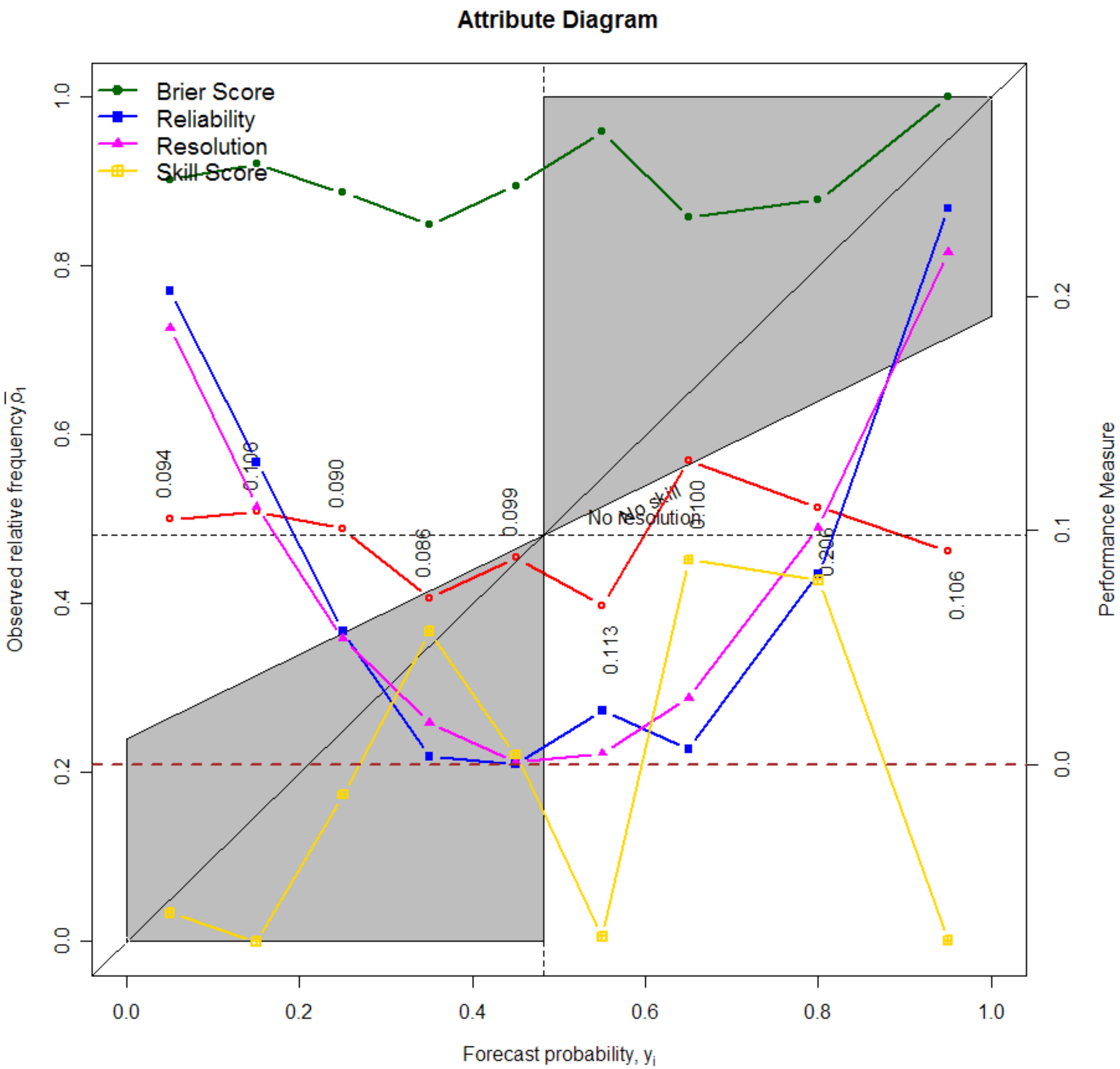




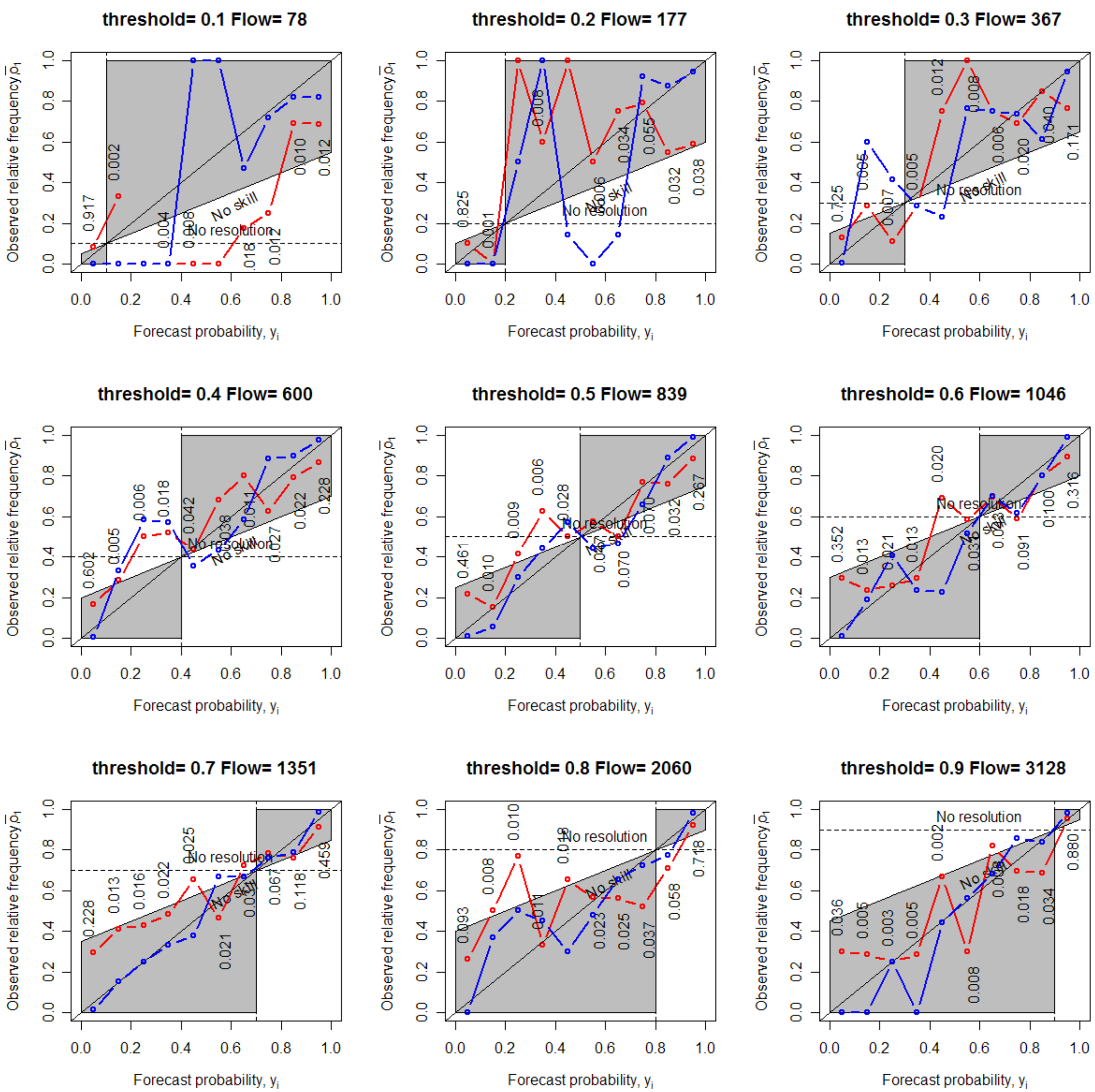
The Attribute Diagram



Attribute Diagram for Random Forecast

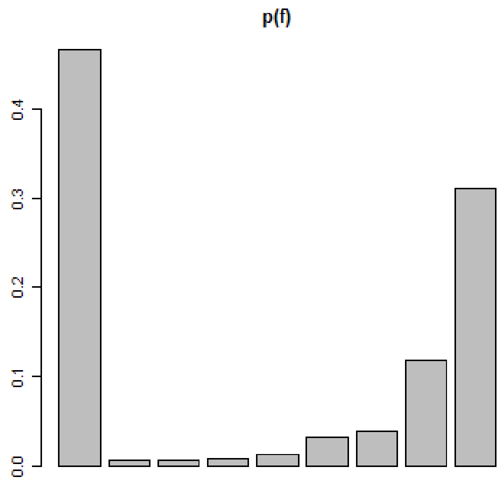


Thresholds and Probability Range Selection

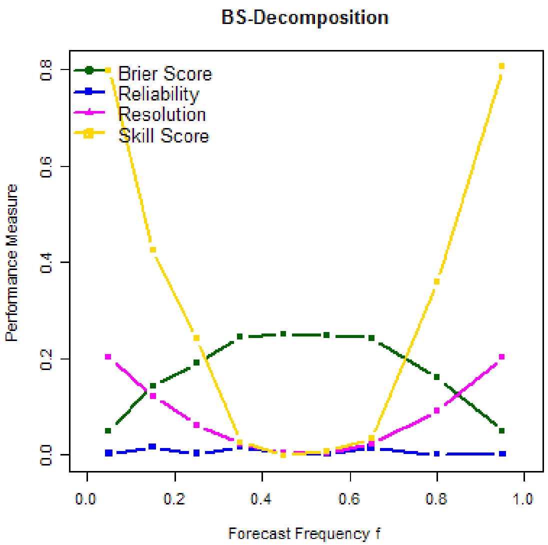


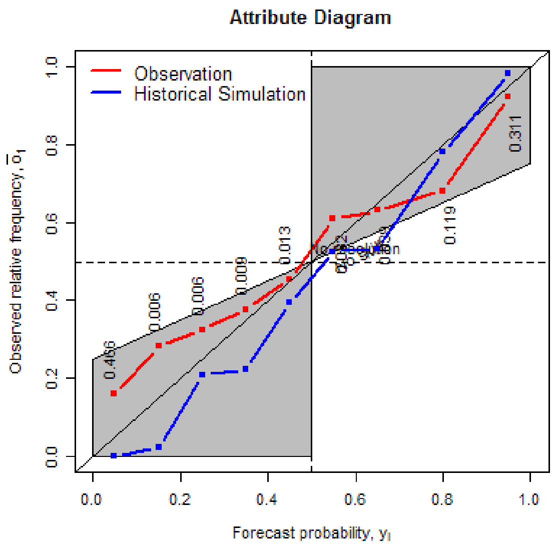
All thresholds together (Also applicable to individual thresholds)

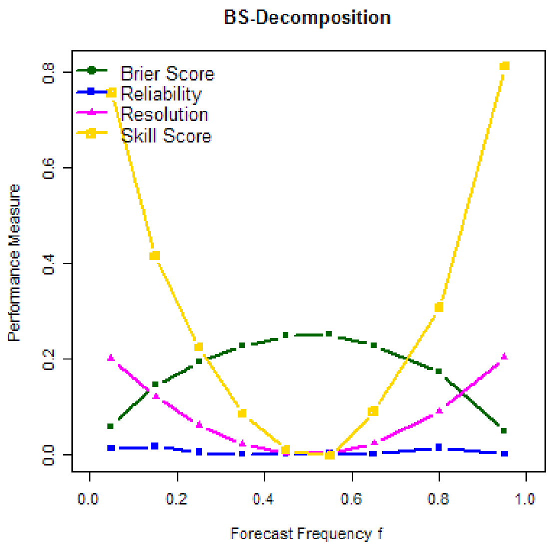
```
Overall Perfromance: Historical Simulation
Brier Score (BS)            =  0.05767
Brier Score - Baseline      =   0.25
Skill Score                 =  0.7693
Reliability                 =  0.002273
Resolution                  =  0.1946
Uncertaintity               =   0.25
```




```
Overall Perfromance: Observation
Brier Score (BS)            =  0.1432
Brier Score - Baseline      =   0.25
Skill Score                 =  0.4273
Reliability                 =  0.007799
Resolution                  =  0.1146
Uncertaintity               =   0.25
```



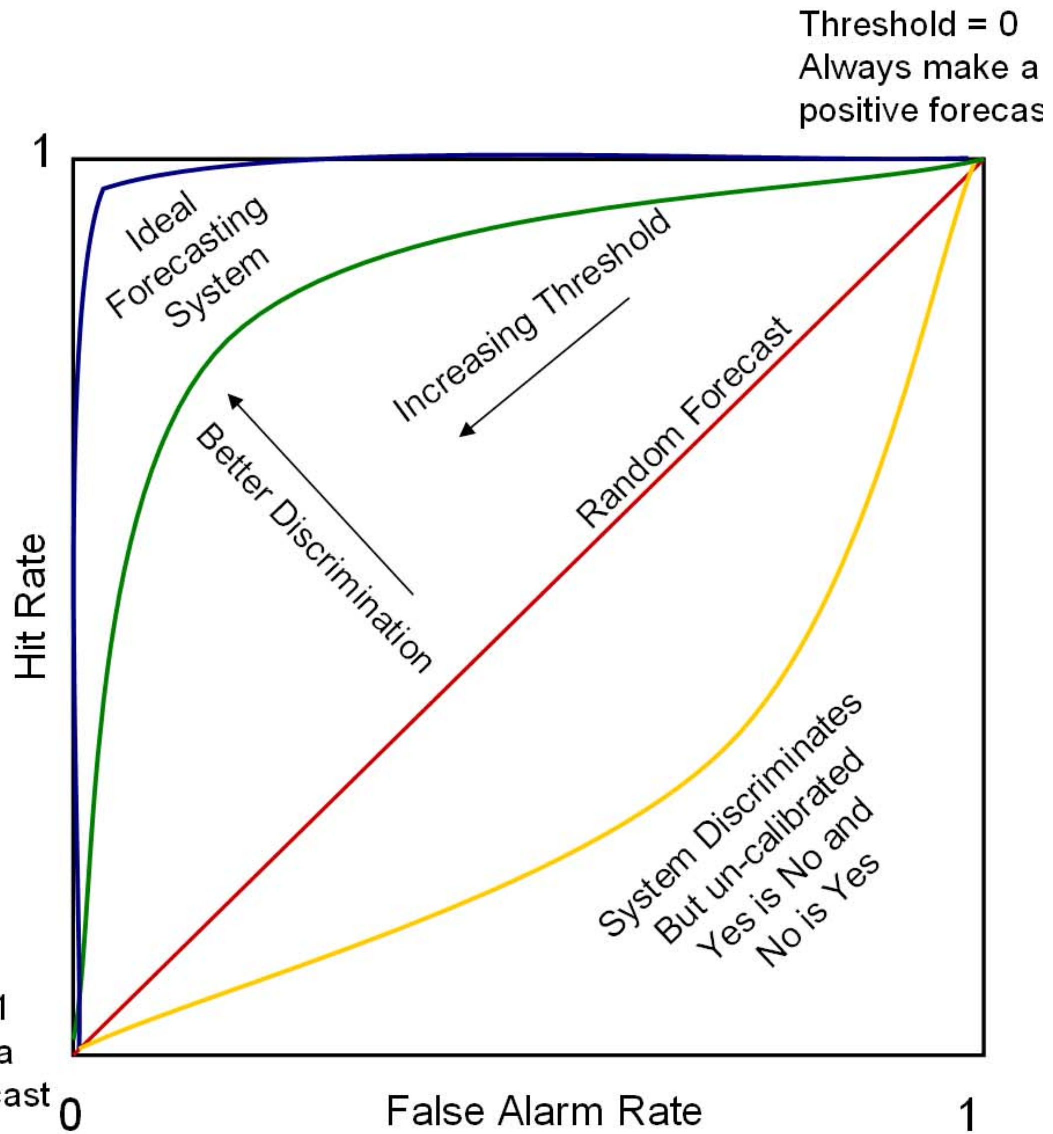




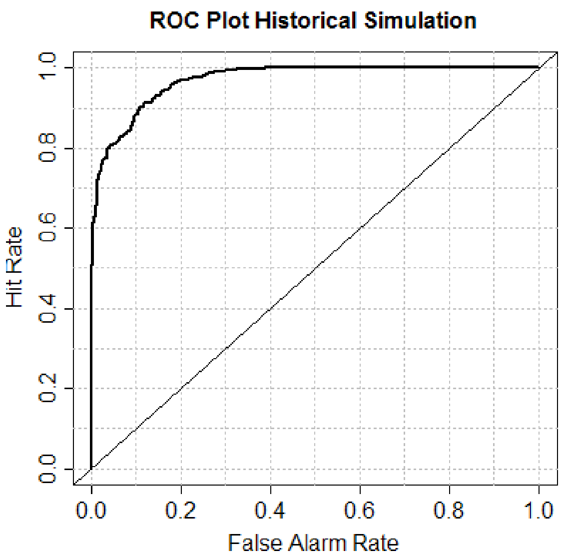
ROC Curve

	Observation	
Forecast	p	n
Y	**True Positive**	**False Positive**
N	**False Negative**	**True Negative**
Total	**P**	**N**

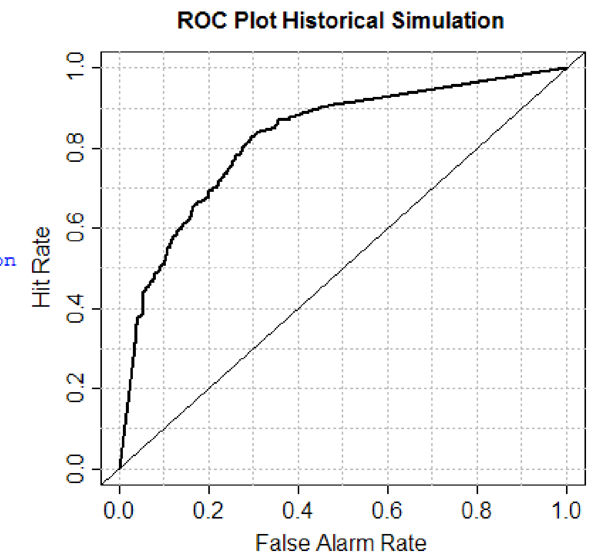
False Positive rate=FP/N = False Alarm Rate
True Positive Rate= TP/P = Hit Rate
Precision = TP/(TP+FP) = Positive Predictive value
Accuracy= (TP+TN)/(P+N)
Specificity = TN/(FP+TN)

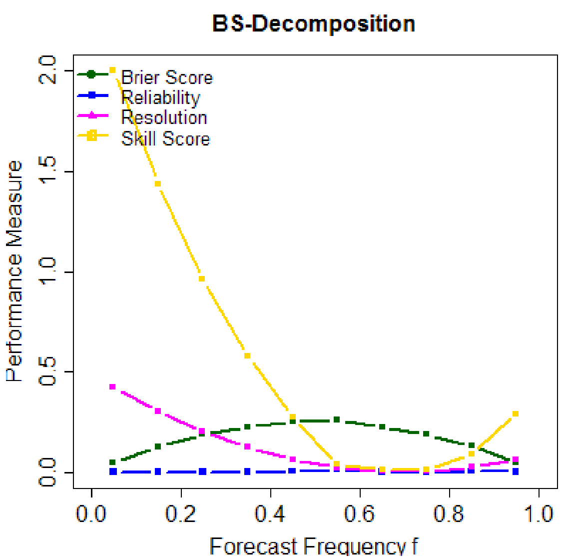


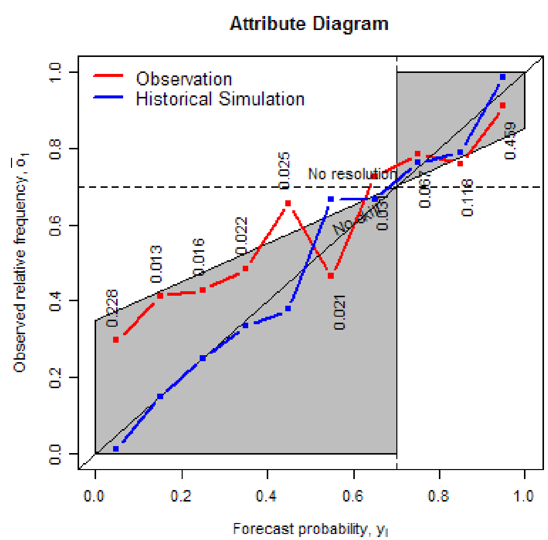
Combining Information

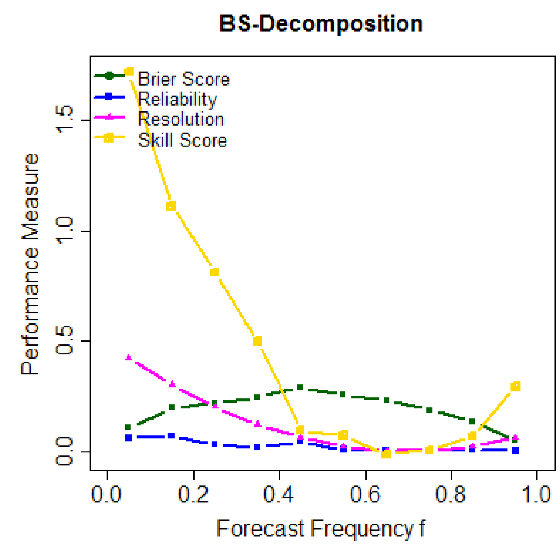


Non-exceedence Threhold:0.7		
Overall Perfromance: Observation		
Brier Score (BS)	=	0.1657
Brier Score - Baseline	=	0.2102
Skill Score	=	0.2113
Reliability	=	0.0187
Resolution	=	0.0631
Uncertaintity	=	0.2102
Overall Perfromance: Historical Simulation		
Brier Score (BS)	=	0.06388
Brier Score - Baseline	=	0.2102
Skill Score	=	0.696
Reliability	=	0.001809
Resolution	=	0.1481
Uncertaintity	=	0.2102

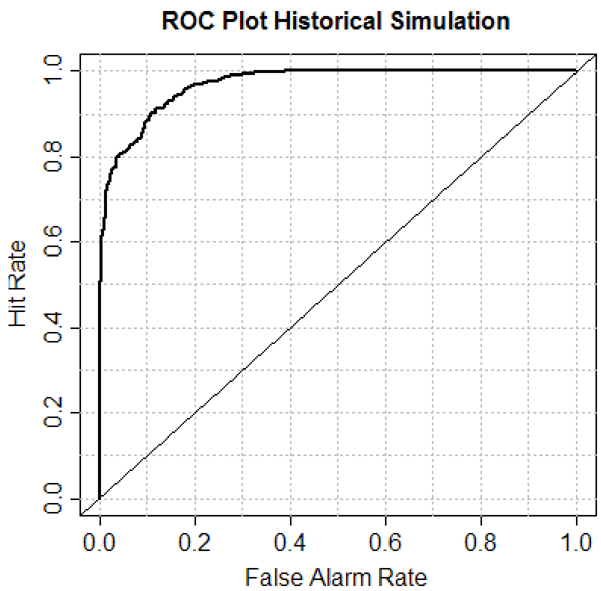




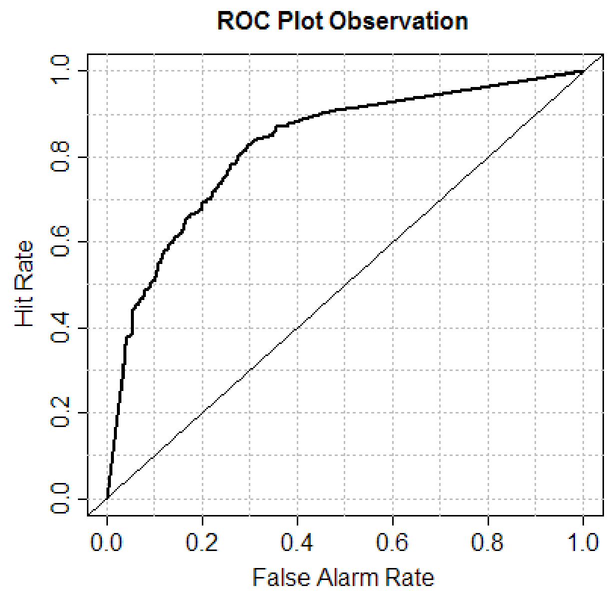


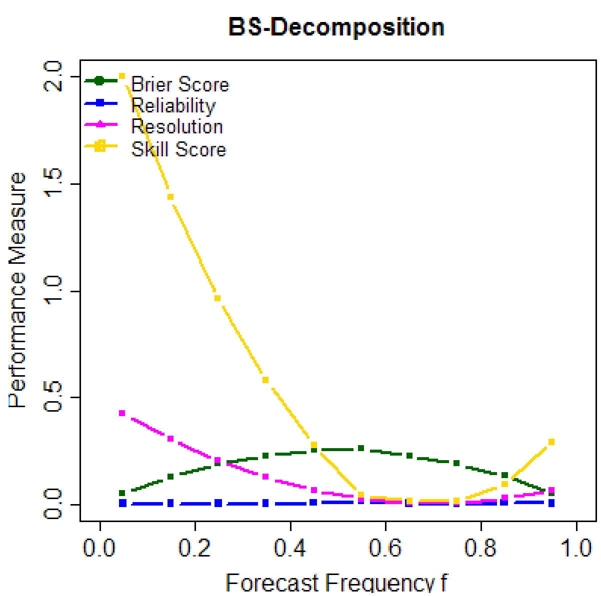


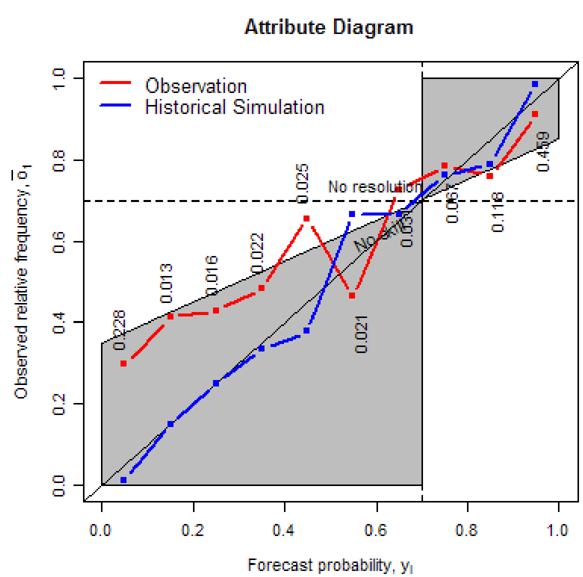
Screen Shot (15 days)

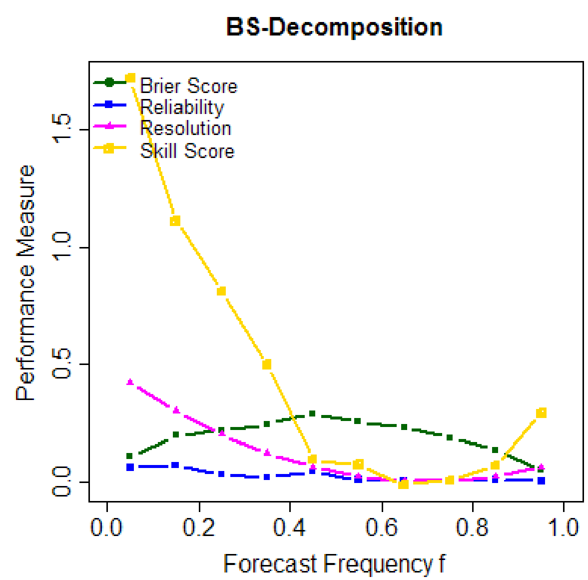



```
Non-exceedence Threshold:0.7
Overall Perfromance: Observation
Brier Score (BS)             =  0.1657
Brier Score - Baseline       =  0.2102
Skill Score                  =  0.2113
Reliability                  =  0.0187
Resolution                   =  0.0631
Uncertaintity                =  0.2102
Overall Perfromance: Historical Simulation
Brier Score (BS)             =  0.06388
Brier Score - Baseline       =  0.2102
Skill Score                  =  0.696
Reliability                  =  0.001809
Resolution                   =  0.1481
Uncertaintity                =  0.2102
```

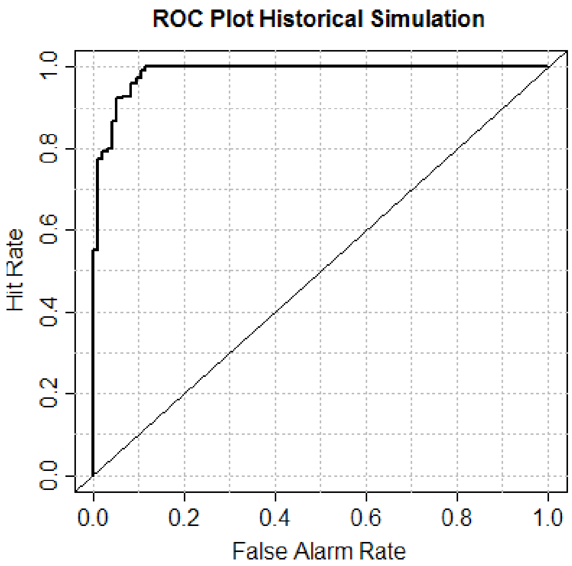




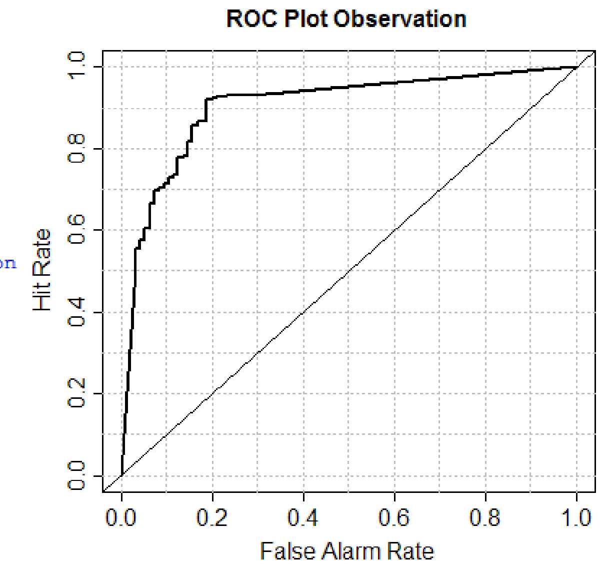


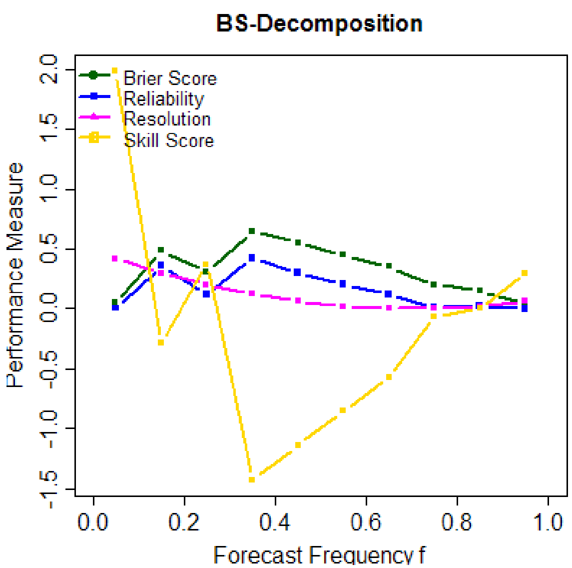


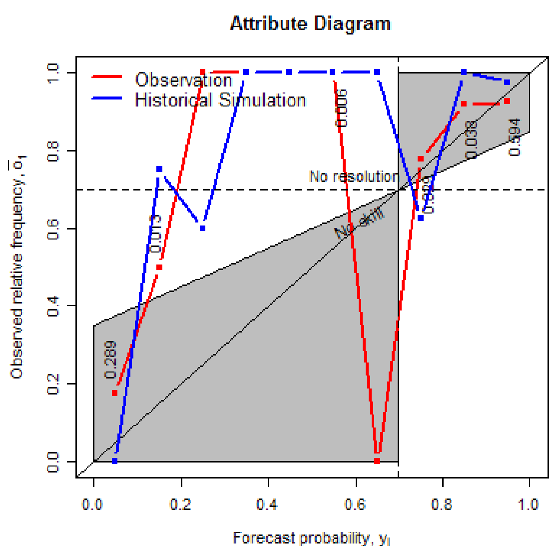
For smaller Data Set

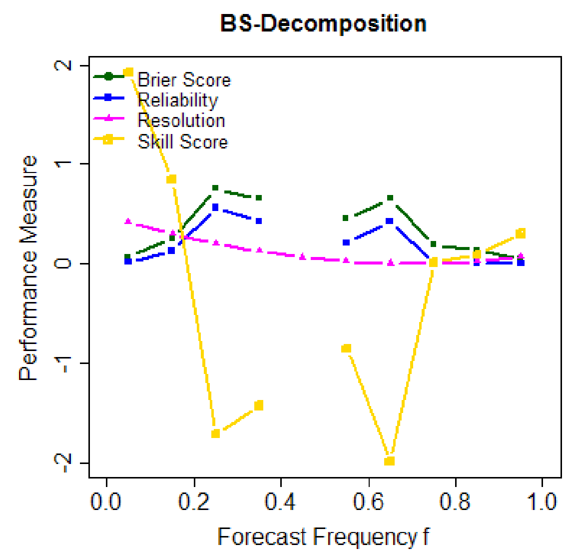


Non-exceedence Threshold:0.7
Overall Perfromance: Observation
Brier Score (BS) = 0.1189
Brier Score - Baseline = 0.2106
Skill Score = 0.4353
Reliability = 0.02494
Resolution = 0.1166
Uncertaintity = 0.2106
Overall Perfromance: Historical Simulation
Brier Score (BS) = 0.04961
Brier Score - Baseline = 0.2106
Skill Score = 0.7645
Reliability = 0.01962
Resolution = 0.1806
Uncertaintity = 0.2106





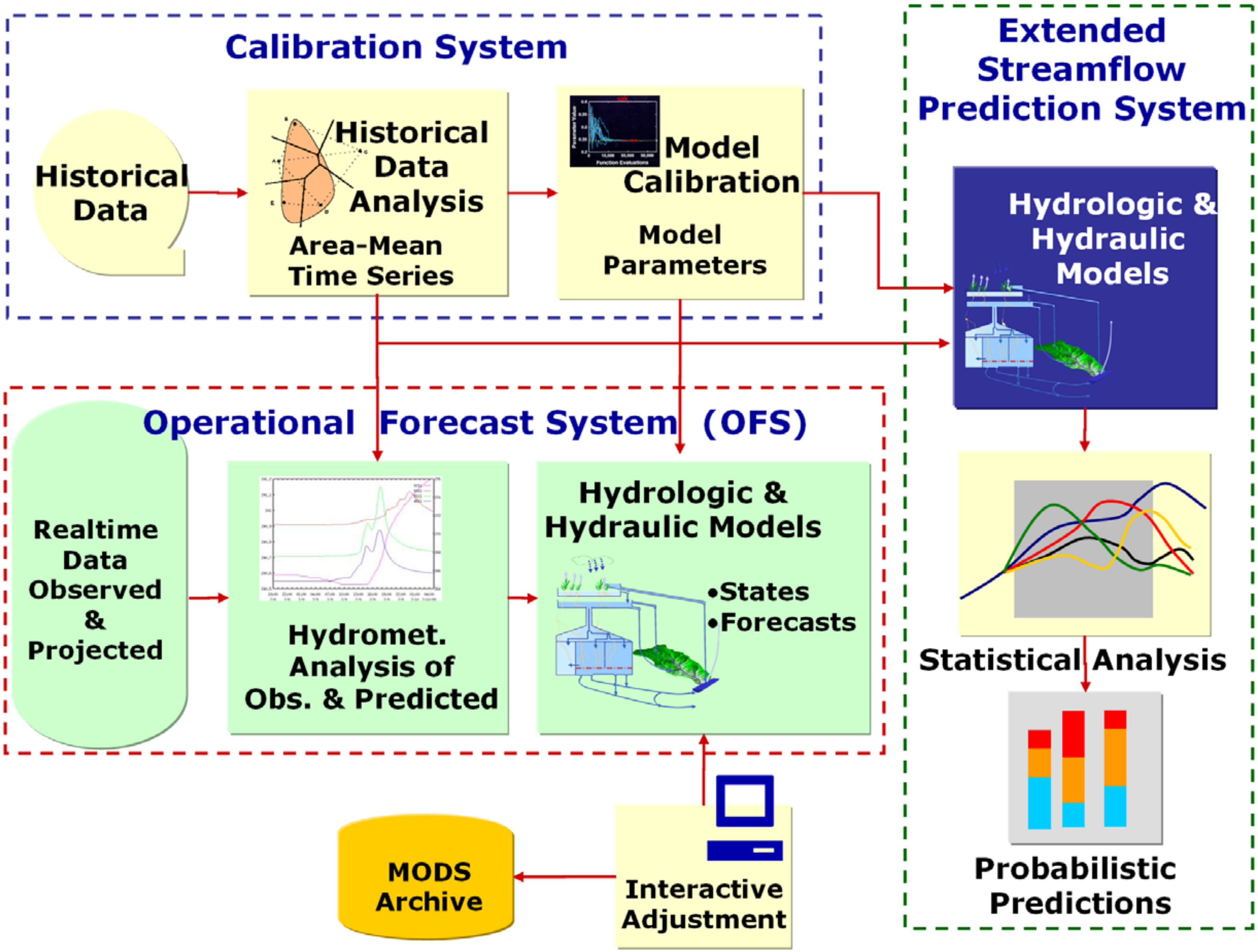




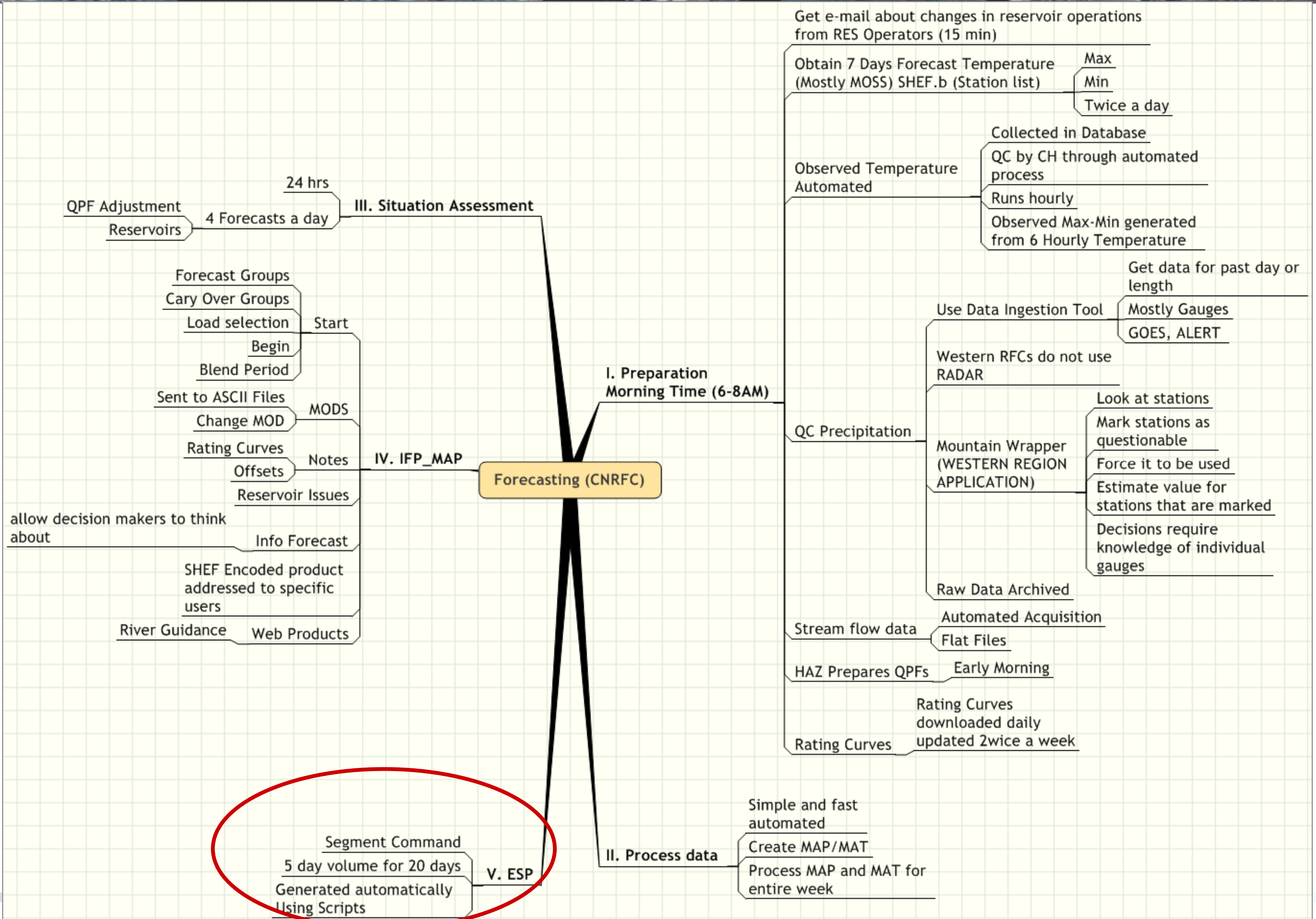
Other Measures (Bradley 2004, Hashino, 2003)

Measured Quality	Base dist.	Measure	Definition (Murphy 1996 and 1997)	Formulae (Based on Bradley (2003)
Bias Unconditional	p(f) t(o)	ME	Mean error. Difference between the mean forecast probability and the climatological probability of the specific threshold.	$ME = \mu_f - \mu_o$
Accuracy	h(f,o)	MSE	Mean square error. Overall degree to which forecast corresponds with observations.	$MSE = (\sigma_f^2 + \mu_f^2) + \mu_o(1 - 2\mu_{f\|o=1})$
Skill or Relative Accuracy	h(f,o)	SSMSE	A normalized measure of accuracy. Also known as the Brier Skill Score.	$SS_{MSE} = 1 - \frac{(\sigma_f^2 + \mu_f^2) + \mu_o(1 = 2\mu_{f\|o=1})}{\sigma_o^2}$
Association	h(f,o)	$\rho_{f,o}$	A measure of association between forecasts and observations. Potential skill of the perfectly reliable forecast (See definition of reliability)	$\rho_{f,o} = \frac{\mathrm{cov}(f,o)}{\sigma_f\ \sigma_f} = \sqrt{\frac{\mu_o}{1-\mu_o}}\left(\frac{\mu_{f\|o=1} - \mu_f}{\sigma_f}\right)$
Reliability/ Type 1 Conditional Bias)	y(o|f) p(f) CR	REL	Degree of correspondence between the observations associated with a given forecast and the forecastAlso known as Type 1 conditional bias. In meteorology it is also known as Calibration	$REL = E_f(\mu_{o\|f} - f)^2$
Resolution	y(o|f) p(f)	RES	Degree of spread of observations around the conditional mean of observation for a given forecast.	$RES = E_f(\mu_{o\|f} - \mu_o)^2$
Sharpness	p(f)	σ^2_f	Degree to which probability forecasts approach 0 and 1.	σ_f^2
Concurrence	y(o|f) p(f)	MSECR SSCR	Concurrence between reliability and resolution	$MSE_{CR} = \sigma_o^2 - REL - RES$ $SS_{CR} = 1 - \frac{MSE_{CR}}{\sigma_o^2} = \frac{RES}{\sigma_o^2} - \frac{REL}{\sigma_o^2}$
Discrimination	r(f|o) t(o) p(f)	DIS	Degree of deviation between the conditional mean for forecast from the mean of forecast	$DIS = E_o(\mu_{f\|o} - \mu_f)^2$ $DIS = (1-\mu_o)(\mu_{f\|o} - \mu_f)^2$ $+ \mu_o(\mu_{f\|o=1} - \mu_f)^2$
Type 2 conditional bias	r(f|o) t(o)	B2	Degree of correspondence between the mean of forecast conditioned by observation and the mean of observations.	$B2 = E_o(\mu_{f\|o} - o)^2$
Relative Type 2	r(f|o) t(o)	MSECR SSCR	Concurrence between discrimination and B2	$MSE_{LBR} = \sigma_f^2 + B2 - DIS$ $SS_{LBR} = 1 - \frac{\sigma_f^2}{\sigma_o^2} + \frac{MSE_{LBR}}{\sigma_o^2} - \frac{DIS}{\sigma_o^2}$

NWSRFS


Calibration System
Historical Data
Historical Data Analysis
Area-Mean Time Series
Model Calibration
Model Parameters
Extended Streamflow Prediction System
Hydrologic & Hydraulic Models
Operational Forecast System (OFS)
Realtime Data Observed & Projected
Hydromet. Analysis of Obs. & Predicted
Hydrologic & Hydraulic Models
•States
•Forecasts
Statistical Analysis
Probabilistic Predictions
MODS Archive
Interactive Adjustment

Forecasting is a Demanding Task



Operational Verification

Wells, 2005. The objective of administrative verification of deterministic river stage forecasts is to determine:

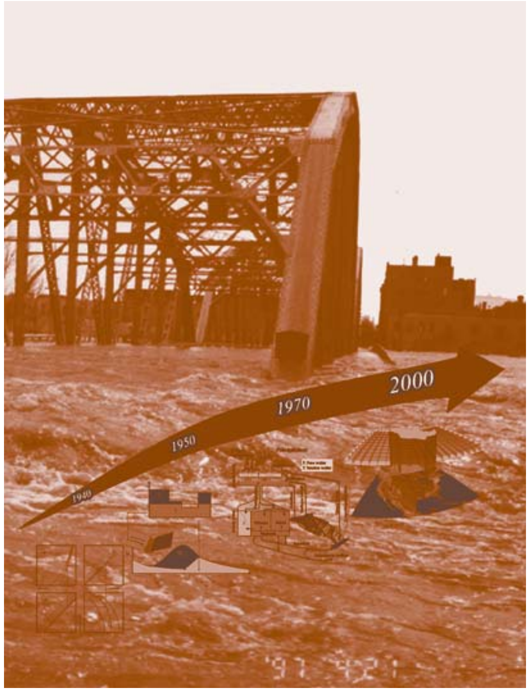
1. How does the performance of the actual forecasts compare to the performance of persistence forecasts?
2. How does the forecast performance change with lead time
3. How does the forecast skill change with time

For ESP, No archive of forecasts exist because of:

(1)Recent implementation and continuing evolution of ESP procedures at RFCs
(2)Lack of archival procedures of actual ESP forecasts in their numeric (ensemble) format.

Towards a New AHPA (NRC, 2006)

While the Inclusion of a verification subcomponent in AHPS NWSRFS is recommended, there is a pressing need for a long-term strategy and maintenance of forecast archive for future verification and NWSRFS evaluation



Possible Structure