NWS Hydrology Forecast Verification Team:

18th Meeting

09/22/2009 –12 pm EDT

Outline

- Final team report: consensus on

  – Recommended verification metrics and products

  – RFC verification case studies

  – Future team activities

- CHPS Verification Service

  – Status

  – Input from Verification Team

Final team report: consensus?

- Recommendations on
  - sets of verification metrics and products to be used at all RFCs
  - verification analyses
  - sensitivity analyses on impact of QPF horizon and impact of run-time mods made on the fly

- Future team activities
  - Second team charter: presented to HICs on 07/10/09 and reviewed in August 09

- Report to be finalized by 09/30/09

Final team report: recommendations

- Key verification metrics for 4 levels of information for single-valued and probabilistic forecasts

  1. Data information (scatter plots, box plots, time series plots)

  2. Summary information (e.g. skill scores)

  3. More detailed information (e.g. measures of reliability, resolution, discrimination, correlation)

  4. Sophisticated information (e.g. for specific events)

- Corresponding verification products

Final team report: recommended metrics

- 4 different levels of information

| Information level	Attributes	Single-valued forecasts	Probabilistic forecasts
1) Data information	Forecast and observed values	Scatter plots for each lead time Time series plots for set of forecasts	Scatter plots with box and whiskers for each lead time Time series plots with box and whiskers for set of forecasts
2) Summary information	Error	MAE	Mean CRPS
	Bias	Relative Bias	Relative Bias in ensemble means
	Skill	$MAE\text{-}SS_{ref}$	$CRPSS_{ref}$
3) More detailed information	Error	MSE	BS for set of events
	Skill	$MSE\text{-}SS_{ref}$	BSS_{ref} for set of events
	Reliability	$Reliability_{MSE}$	$Reliability_{CRPS}$
	Resolution	$Resolution_{MSE}$	$Resolution_{CRPS}$
	Discrimination	ROC Score for set of events	ROC Score for set of events
	Correlation	Correlation coefficient	Correlation coefficient for ensemble means
	Sample size	Number of forecast-observed pairs	Number of forecast-observed pairs
4) Sophisticated information	Reliability	FAR for set of events	Cumulative Talagrand Diagram, Reliability Diagram for set of events
	Discrimination	ROC curves for set of events	ROC curves for set of events
	Forecast value	Relative value	Relative value

Final team report: recommended products
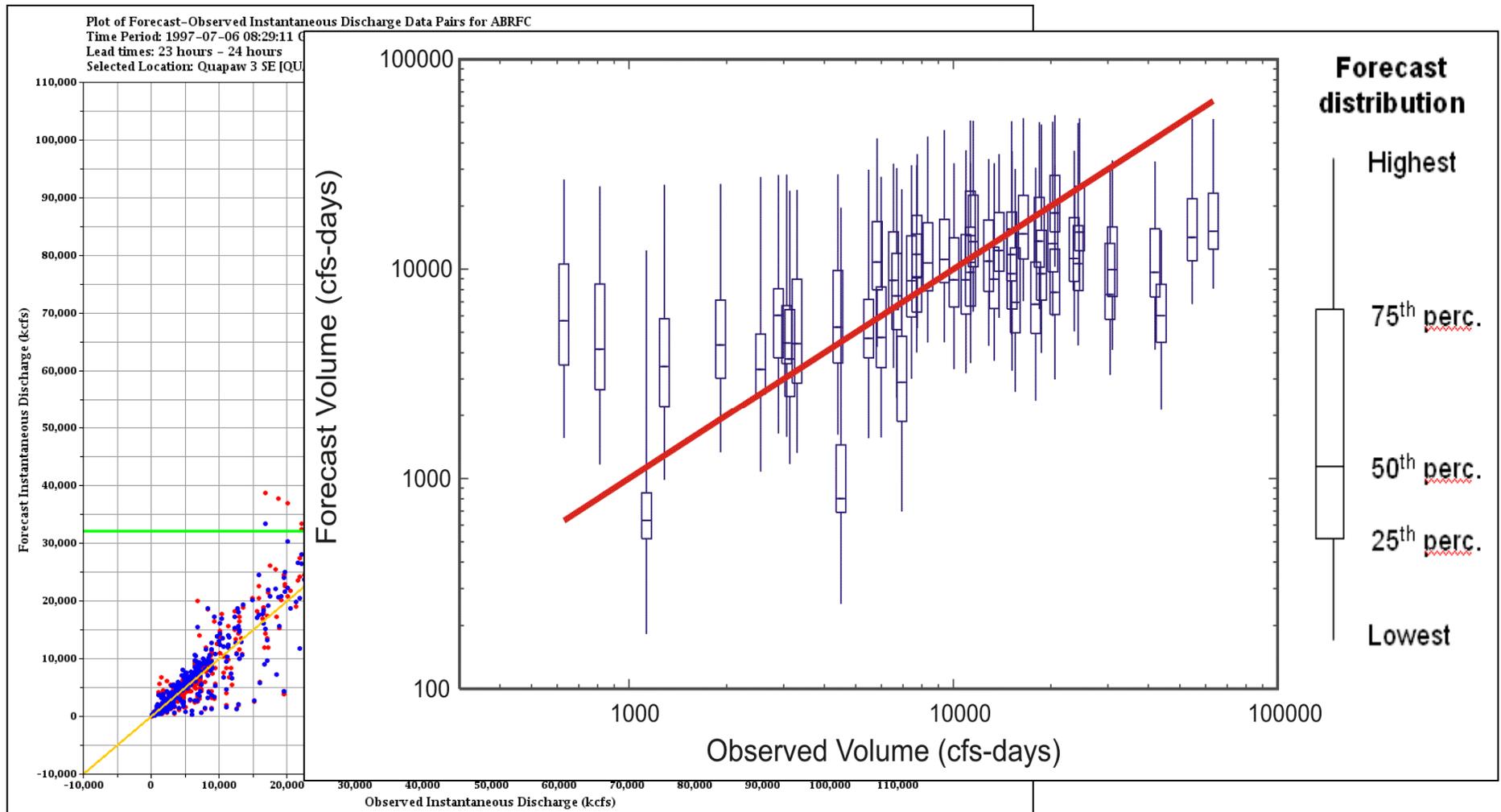
- Examples of verification products: level 1

Final team report: recommended products

- Examples of verification products: level 1
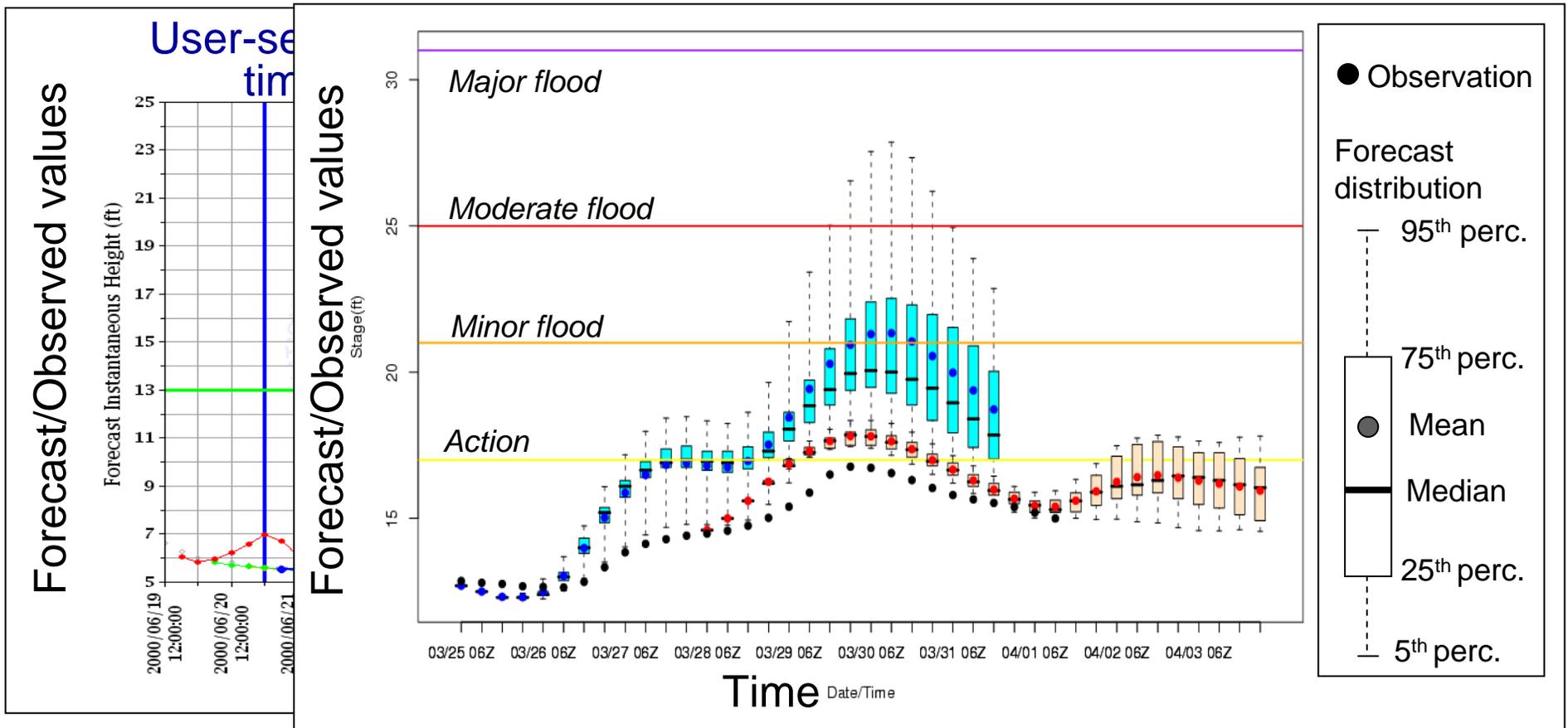
Final team report: recommended products

- Examples of verification products: level 2
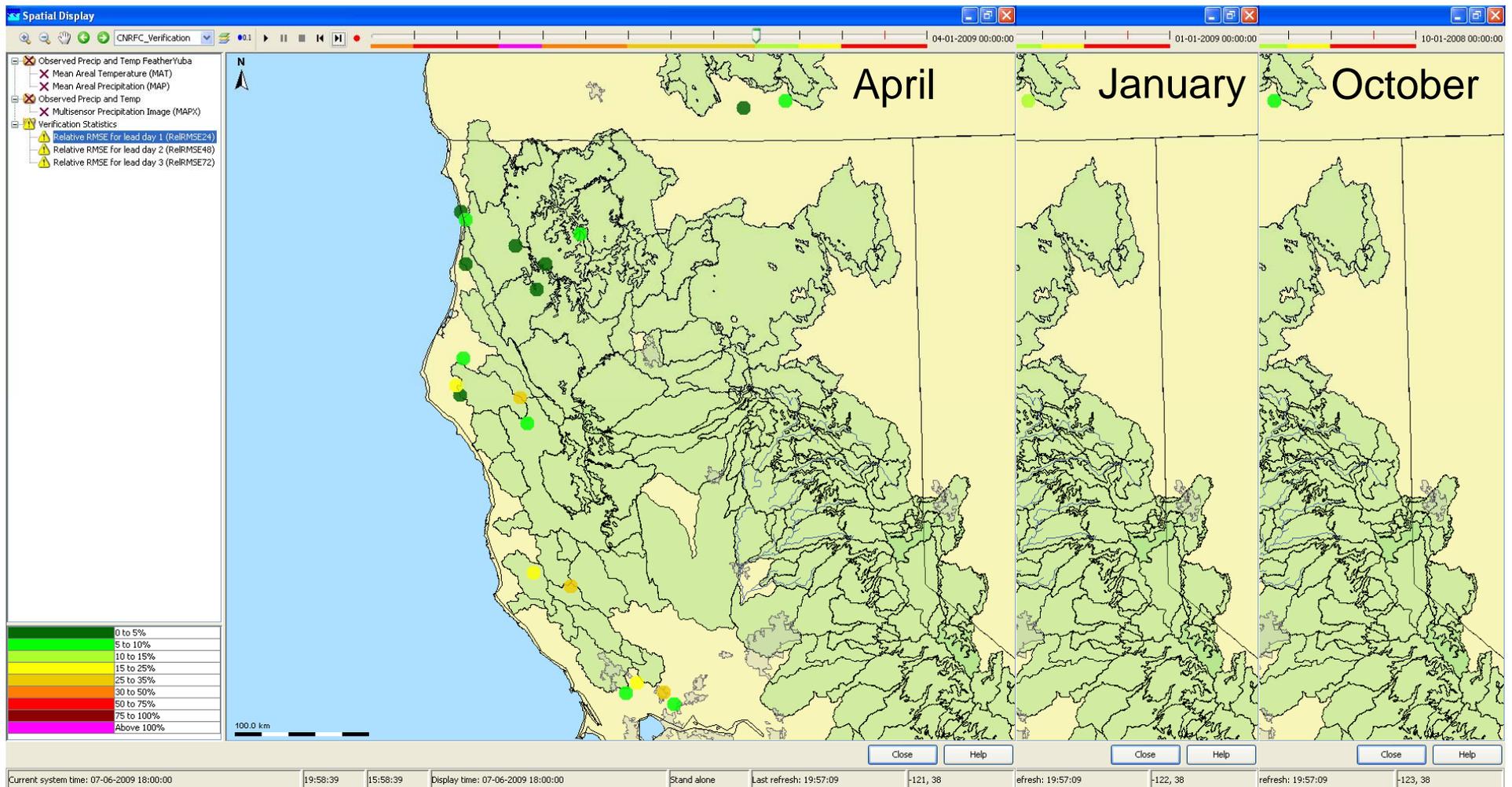
Final team report: recommended products

- Examples of verification products: level 3

Score

Score
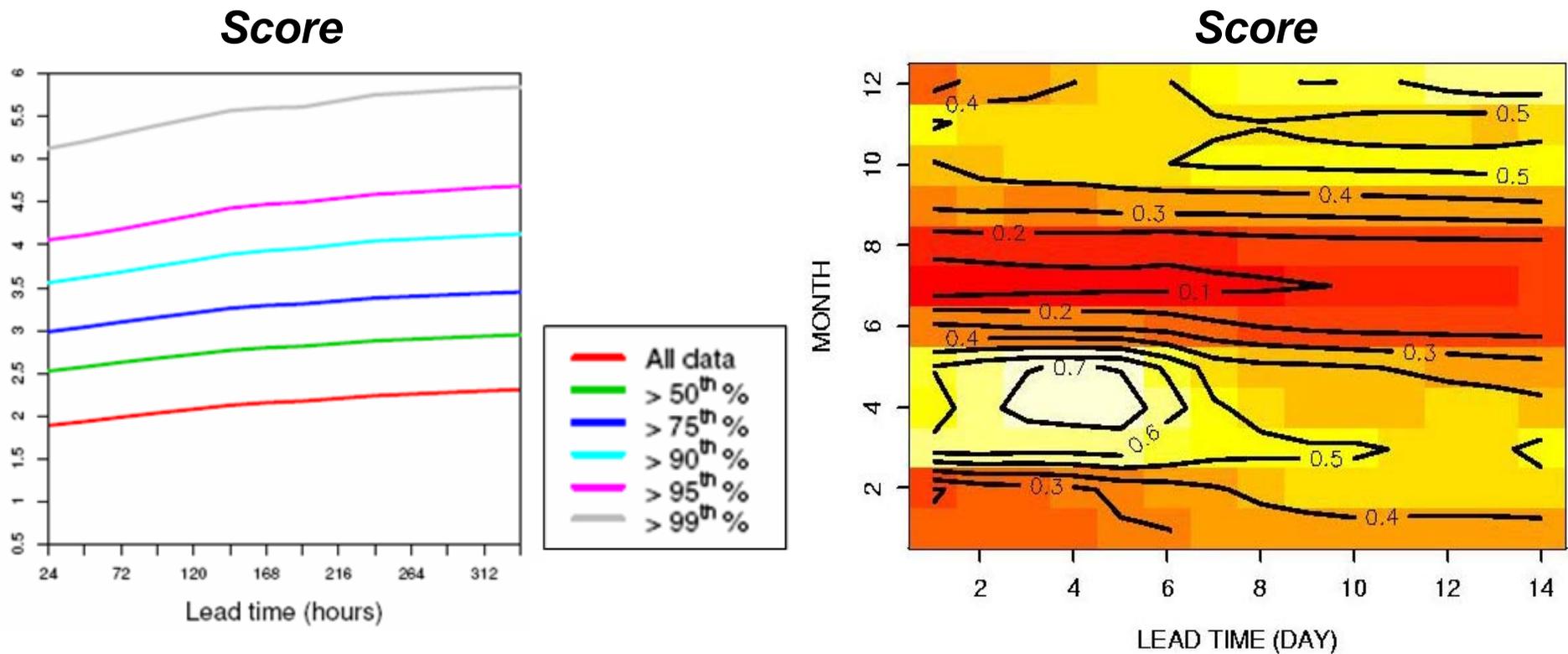
Final team report: recommended products

- ## Examples of verification products: level 4

Event: > 85th percentile from observed distribution

Reliability
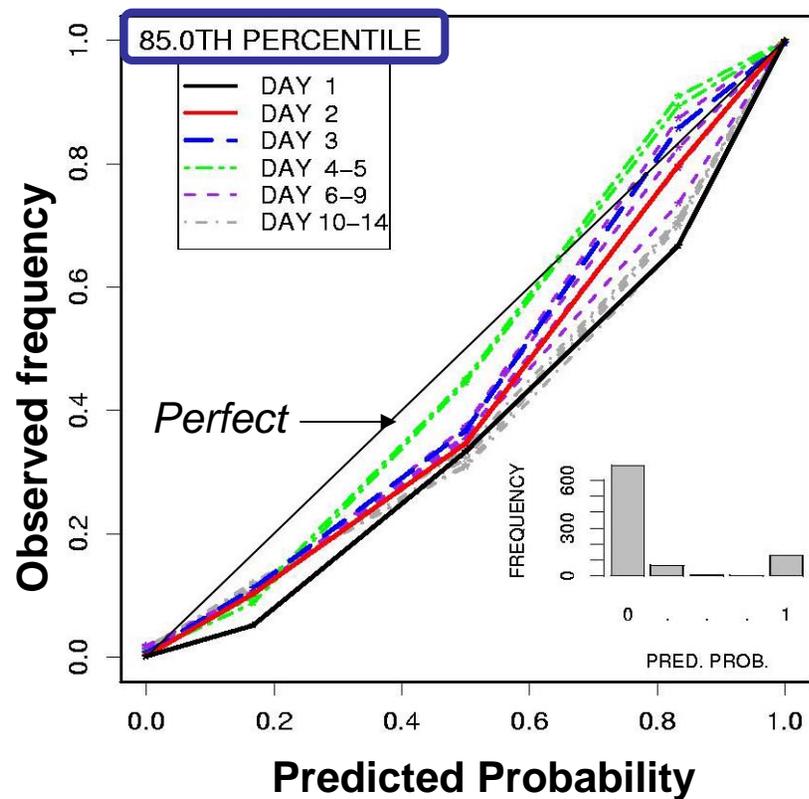
Discrimination
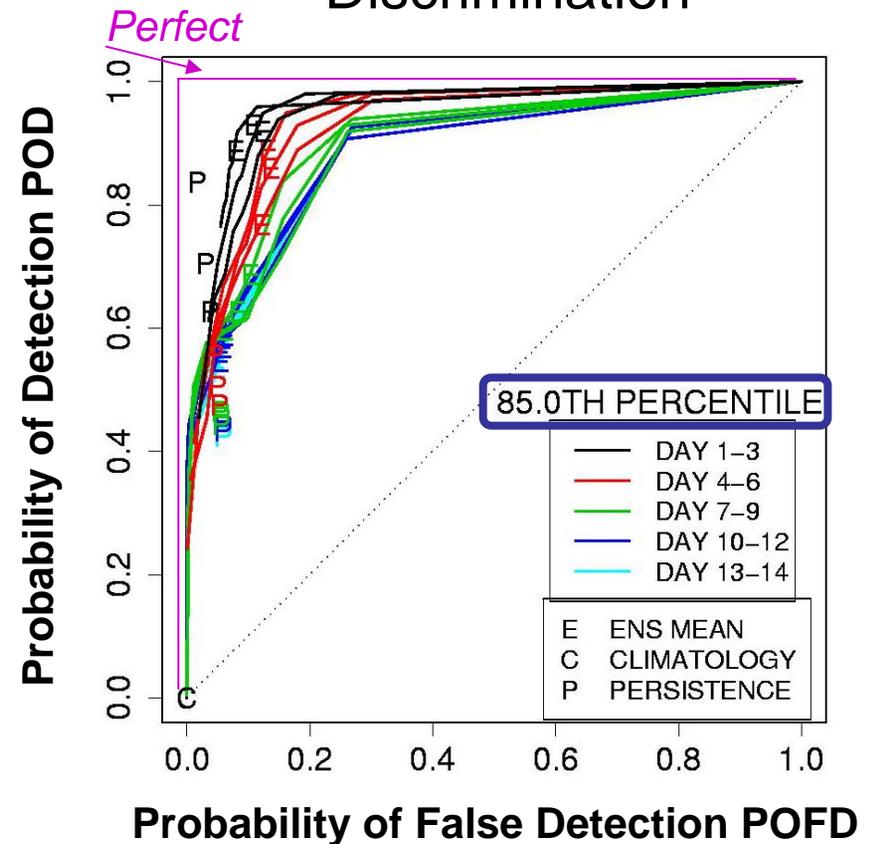
Final team report: recommended analyses

- Analyze any new forecast process with verification

- Use different temporal aggregations (e.g. weekly max. flow)

  – Analyze verification statistic as a function of lead time

  – If similar performance across lead times, data can be pooled

- Perform spatial aggregation carefully

  – Analyze results for each basin and results plotted on spatial maps

  – Use normalized metrics (e.g. skill scores)

  – Aggregate verification results across basins with similar hydrologic processes (e.g. by response time: fast, medium, and slow)

- Report verification scores with sample size

  – In the future, confidence intervals

Final team report: recommended analyses

- Evaluate forecast performance under different conditions
    - w/ time conditioning: by month, by season
    - w/ atmospheric/hydrologic conditioning:
        - low/high probability threshold
        - absolute thresholds (e.g., PoP, Flood Stage)
    - Check that sample size is not too small

- Analyze sources of uncertainty and error
    - Verify forcing input forecasts and output forecasts
    - For extreme events, verify both stage and flow
    - Sensitivity analysis to be set up at all RFCs:
        1) impact of QPF horizon
        2) impact of run-time mods made on the fly

Sensitivity analysis: QPF horizon

- Goal: what is the optimized QPF horizon for hydrologic forecasts?

- QPF horizon to test:
  - 0 (no QPF), 6-hr, 12-hr, 18-hr, 24-hr, 30-hr, 36-hr, 48-hr, 72-hr, 96-hr
  - Longer horizon: optional

- Model states to use:
  - Similar to operational mods except mods that impact future states
  - Metadata to store which mods were used in these runs

- What forecast to verify
  - 6-hr stage forecasts for 7-day window (longer window for slow response basins)

Sensitivity analysis: run-time MODs

- Goal: do run-time mods made on the fly improve forecasts?

- 4 scenarios
  - Operational forecasts (w/ all mods)
  - Forecasts w/ best available obs. and fcst. inputs wo/ on-the-fly mods
  - Forecasts w/ best available obs. inputs (no fcst) w/ all mods
  - Forecasts w/ best available obs. inputs (no fcst) wo/ on-the-fly mods

- What forecast to verify
  - 6-hr stage forecasts for same window as in operations

- Model states:
  - Carryover from 5 days ago (w/ past mods) + a priori mods (known before producing any forecast)

Sensitivity analyses: what to do at RFCs?

- Goal for FY10: run the different forecasting scenarios and store outputs to start building an archive

    - Within CHPS or outside CHPS

- Steps:

    - Define workflows for the different forecasting scenarios

    - Everyday: run scenarios + store outputs and metadata

    - Perform data quality control

- OHRFC experience with multiple forecasting scenarios

- Experience from other RFCs?

Final team report: future team activities

- Produce and evaluate recommended verification standards w/ new RFC verification case studies

  – Are recommended standard metrics/products meaningful to forecasters?

- Perform user analysis of verification products w/ RFC SCHs and OCWWS

  – What verification products should be delivered to users?

- Develop requirements for disseminating RFC verification information at NWS Performance Branch and at RFCs

  – What are the best methods to supply verification information?

Final team report: future team activities

- Support design and development of CHPS Verification Service (CHPS-VS):

  – Help develop user requirements

  – Review software design documents

  – Continue testing prototype functionality: EVS, CHPS displays

- Ultimate goal: provide useful verification information to

  – **modelers and forecasters** to guide improvements of forecasting system

  – **users** to maximize utility of forecasts in their decisions

CHPS Verification Service: development

- Development strategy:

  - Use existing capabilities in verification software (IVP, EVS, WR water supply website), Graphics Generator, FEWS displays to develop prototypes for verification products and services

  - Collaboration with RFCs and OCWWS to determine CHPS VS requirements and meaningful verification products

  - HSEB-HSMB-Deltares collaboration for design, development & implementation

CHPS Verification Service: current system
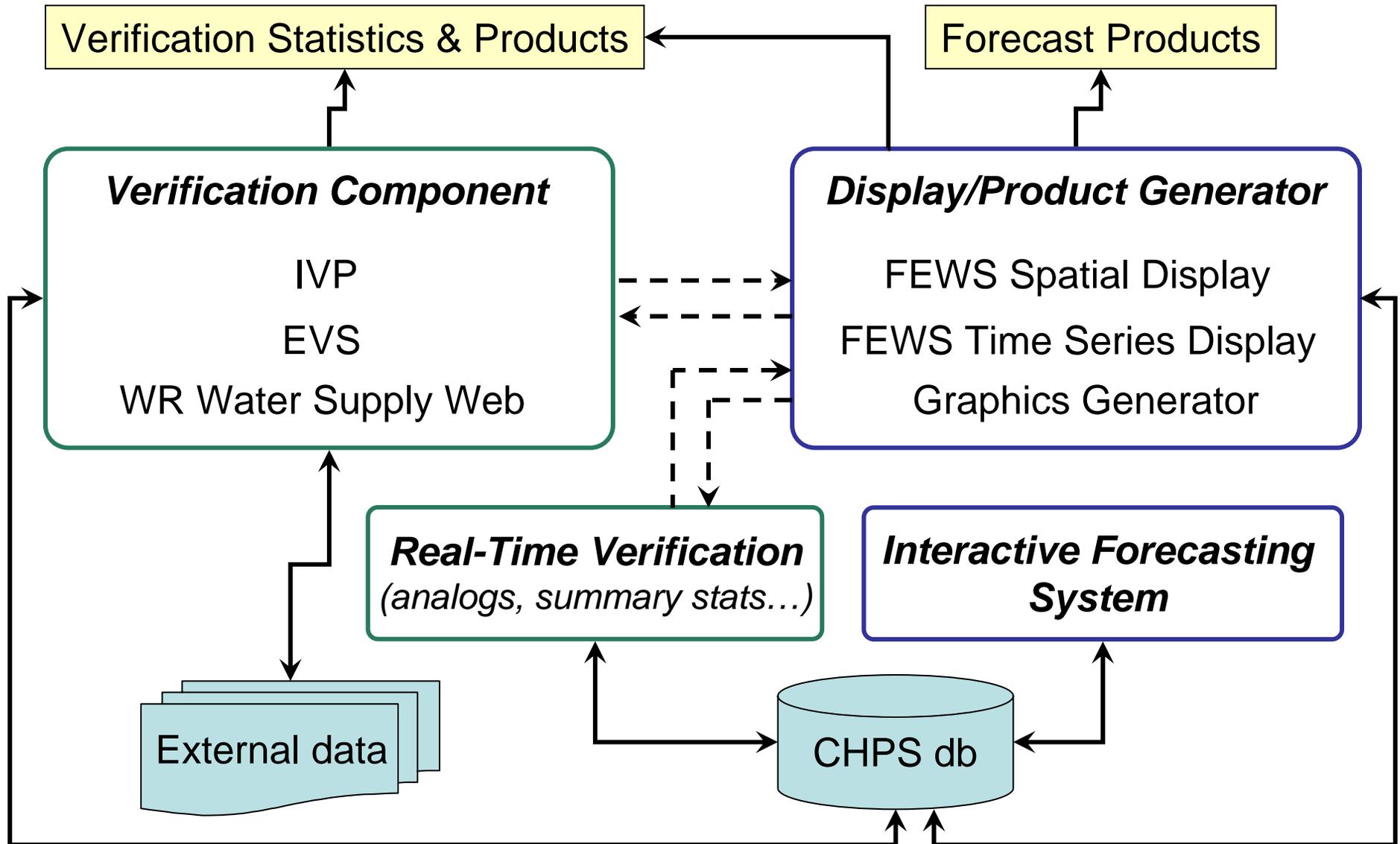
CHPS Verification Service: team input

- Verification prototypes

  ➢ EVS version 2.0 to be delivered in early October 09 (w/ CHPS adapter): feedback through verif-hydro list server

  ➢ CHPS prototype displays: demo available at RFCs for feedback

- Verification products

  ➢ Suite of products to be reviewed by RFCs (including SCHs) and OCWWS: initial examples from report; further examples in Nov.-Dec. 09

- Verification system requirements

  ➢ List of questions to be sent to RFCs: analog query and display, spatial verification displays

Next meeting

- 19th meeting: November-December 09
  - Present progress on
    - Setting up multiple forecasting scenarios at RFCs
    - Developing CHPS VS: prototype, products, requirements

Thank you!

Questions?