Summary of Results of Meteorological Forcings Comparisons from CNRFC, ABRFC, NERFC, NWRFC

Ziya Zhang, David Kitzmiller, Wanru Wu
Office of Hydrologic Development
April 2011

1. Introduction

Staff at all River Forecast Centers (RFCs) have agreed to initiate an operational transition to producing gridded meteorological forcings for river models (precipitation, temperature, freezing level, potential evapotranspiration). At some field offices, some or all of these forcings are routinely produced by legacy preprocessors that estimate basin-average values directly from input point station data. This legacy approach precludes the use of powerful grid editing tools that enable input quality control and manipulation of meteorological fields, both observed and forecasted. However, the transition is complicated by the possibility that basin averages values derived from gridded fields have nontrivial statistical differences from the same averages as calculated by the legacy preprocessor system. These include different long-term bias and spread characteristics, which could lead to degradation in the quality of river forecasts produced from models calibrated to the biases specific to the legacy preprocessor output.

Studies to assess these potential impacts were carried out at the four RFCs scheduled to be the first to implement the Community Hydrologic Prediction System (CHPS), namely the Northwest (NWRFC), Northeast (NERFC), Arkansas-Red (ABRFC), and California-Nevada (CNRFC) offices. Staff at these CHPS Acceleration or CAT RFCs have reported on results with new paradigms for generating hydrometeorological input to their river models.

This report contains a brief summary of the meteorological inputs and software tools considered by RFC staff (section 2), a synopsis of investigations (section 3), summaries of individual RFC reports (sections 4-7), and an overall summary (section 8).

2. Summary of inputs and software tools

At three of the RFCs (NWRFC, CNRFC, and NERFC) the primary precipitation driver is hourly and daily rain gauge reports. At ABRFC, a gauge-radar precipitation analysis known as P3 has been in use for many years.

All of the RFCs have relied mainly on daily point reports, usually daily maximum/minimum values, for temperature input. With the start of the CHPS effort, many of the RFCs are actively using externally-generated gridded input for

temperature and potential evapotranspiration (PE) as first guess fields for these variables. These include the Real-Time Mesoscale Analysis (RTMA; Horel and Colman 2005) and other regionally-produced products.

The legacy preprocessing system hereafter referred to as OFS featured tools for quality checking and editing of rain gauge precipitation and temperature input. Staff at some offices gained experience in the use of DQC for temperature and precipitation quality control during the course of the study period. The operational Multisensor Precipitation Estimator (MPE; Glaudemans et al. 2009) package features quality control tools for both gauge and radar input. The capabilities of the Gridded Forecast Editor (GFE; Wakefield 2008) have been exploited by a number of the field offices.

3. Summary of report topics

Material was generated and shared by the CAT RFCs during 2009-2010, a process which included a webinar exchange in April 2010.

Two of the offices' reports (NWRFC, ABRFC) concentrated on direct comparison of 6-hour basin average values derived from the same input data, but processed through both the legacy and grid-derivation systems. Correlations between the two sets of data for both mean-areal precipitation (MAP) and temperature (MAT) were generally high (> 0.7) but sometimes much lower in rugged terrain. Biases between the two were also the largest in mountain regions, as might be expected; there are generally much smaller spatial gradients in climatic values in flat terrain than in hilly areas, thus the particular assumptions about spatial interpolation are of less consequence.

The study by CNRFC staff documented an approach to adapting the gridded forcings toward the mean value of the basin averages from the legacy system. The aim of this approachis to ensure that there is no bias inconsistency between the real-time gridded forcing input and the historic data used in model calibration, by applying a correction factor to the gridded estimates. It appears to be successful, and has since been applied in real-time operations. The CNRFC staff also documented some improvement in river simulations at some gauging points through the use of gridded rather than legacy precipitation input.

The study carried out at NERFC focused on the impacts of different inputs from the legacy and from two gridding systems on river simulations. Based on studies carried out in early spring of 2010, gridded precipitation input from rain gauges, as processed with Daily Quality Control (DQC), was superior to the legacy forcings.

Subsequently, OHD staff have worked to provide new tools for quality control and analysis of historic meteorological point data, with the end goal of providing datasets for calibration of lumped and distributed hydrologic models based on

gridded inputs. These tools will enable RFC staff to re-calibrate their soil and channel models with meteorological data consistent with real-time inputs.

Summaries of reports from the CAT RFC staff are contained in the following four sections, followed by an overall summary.

4. CNRFC (Rob Hartman, Art Henkel)

The CNRFC studies were designed to address the following three basic points:

1. Validate Flood Early Warning System (FEWS) simulations vs. NWSRFS simulations relative to current RFS-derived basin forcings (MAP and MAT) to answer if FEWS get the same outputs as NWSRFS when using standard inputs;
2. Compute new basin forcings from grids, from the Daily QC (DQC) process, and compare them to forcings from the Operational Forecast System (OFS) preprocessors, to determine differences between basin inputs derived from grids and those derived from pre-determined station weights;
3. Assess sensitivity in simulated flows due to changes in forcings to answer if the input differences matter.

Approach:

1. Steps for MAP forcing analysis

    a. Determine target MAP 6-h climatology, based on PRISM 30-year monthly climatology datasets
    b. Validate calibration MAP climatology w/ target
    c. Validate current OFS MAP climatology w/ target
    d. Calculate DQC MAP climatology
    e. Rescale DQC MAP climatology to target
    f. Reanalyze water-year (WY) data for OFS & DQC with potentially new info
    g. Compare OFS and DQC MAP data for WY 2006 (Figs. 1-2).

    Results indicate that the re-scaling procedure results in measurable improvement in the bias and random differences between the new gridded MAP and the target climatology, relative to simply inputting DQC-derived precipitation (see Figs. 1-2).

2. Steps for MAT forcing analysis

    a. Determine target MAT TX/TN climatology based on height/lapse rate dependent interpolation of available station data
    b. Validate calibration and current OFS MAT climatology w/ target
    c. Calculate DQC MAT climatology
    d. Rescale DQC MATs to target climatology & RFS techniques
    e. Compare scaled DQC MAT's with OFS MAT's for WY 2006.

    Again, results indicate that the rescaling and elevation compensation procedures lead to closer agreement with the target climatology (labeled CREC1 in Figs. 3-4)

than did simply applying the DQC technique alone. Results were verified for both points (Fig. 3) and basin averages (Fig. 4).

3. MAZ or rain-snow (freezing) levels

    a. MAZ calculations in MM-DQC
    b. Use of RUC NWP model analysis field as observed data (*RUC-Rapid Update Cycle, NWP-Numerical Weather Model*)
    c. Supplemented with rain-snow profiler observations
    d. OFS RSNWELEV operation in at the CNRFC currently initialized with surface MAT in observed domain
    e. Switch to explicit observed freezing level input should be an improvement

Results/Conclusions:

1. MAP results to date with comparisons between OFS and MM-DQC are promising
2. MAT validation has been very encouraging, particularly with the use of synthetic stations
3. MAZ (rain-snow) analysis & validation is being scoped. May represent biggest change
4. Hydrologic validation to date has been positive but only episodic for WY06
5. Analysis over 10 year window is ongoing; will include hydrologic validation and sensitivity
6. Migration of existing methodologies has necessitated significant analysis, cleanup and improvement

Hydrologic simulations for some basins such as Smith River - Crescent City (Fig. 5) showed overall improvement relative to the legacy forcings.

Issues: Only a single year sample was available.

Next Steps:

1. for MAP

    a. Continue detailed spatial analysis of differences in MAP forcings
    b. Expand analysis window to full operational data period of record (2000-forward)
    c. Iterate

2. for MAT

    a. Steps are similar to those for MAP
    b. Comparisons are somewhat simplified with use of pseudo-stations in OFS MAT

3. for MAZ

    a. Complete freezing level (MAZ) analyses

b. Validation and sensitivity tests
c. Complete hydrologic validation and sensitivity test

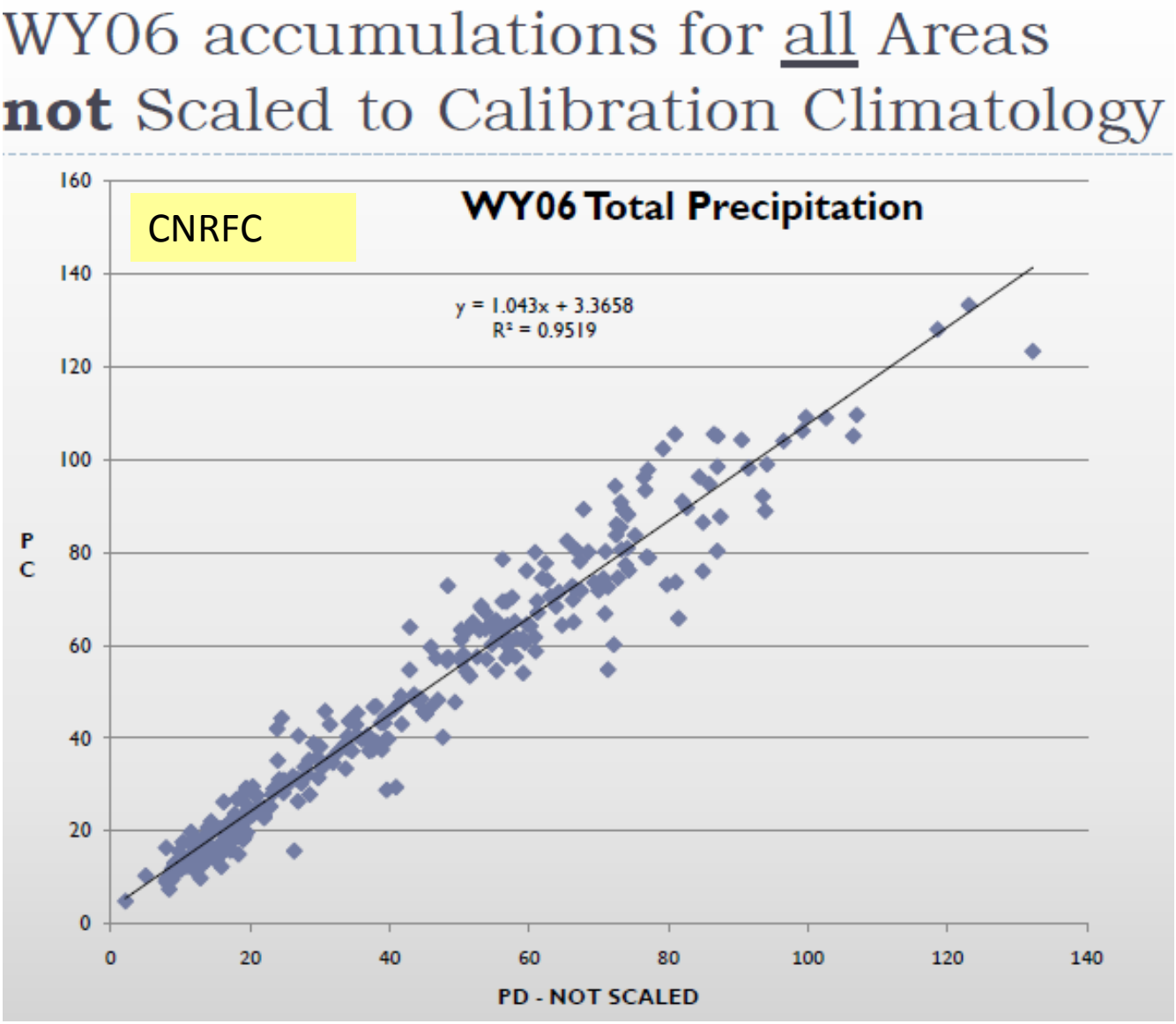

Figure 1. Comparison of total precipitation based on NWSRFS and MM-DailyQC, for multiple basins.

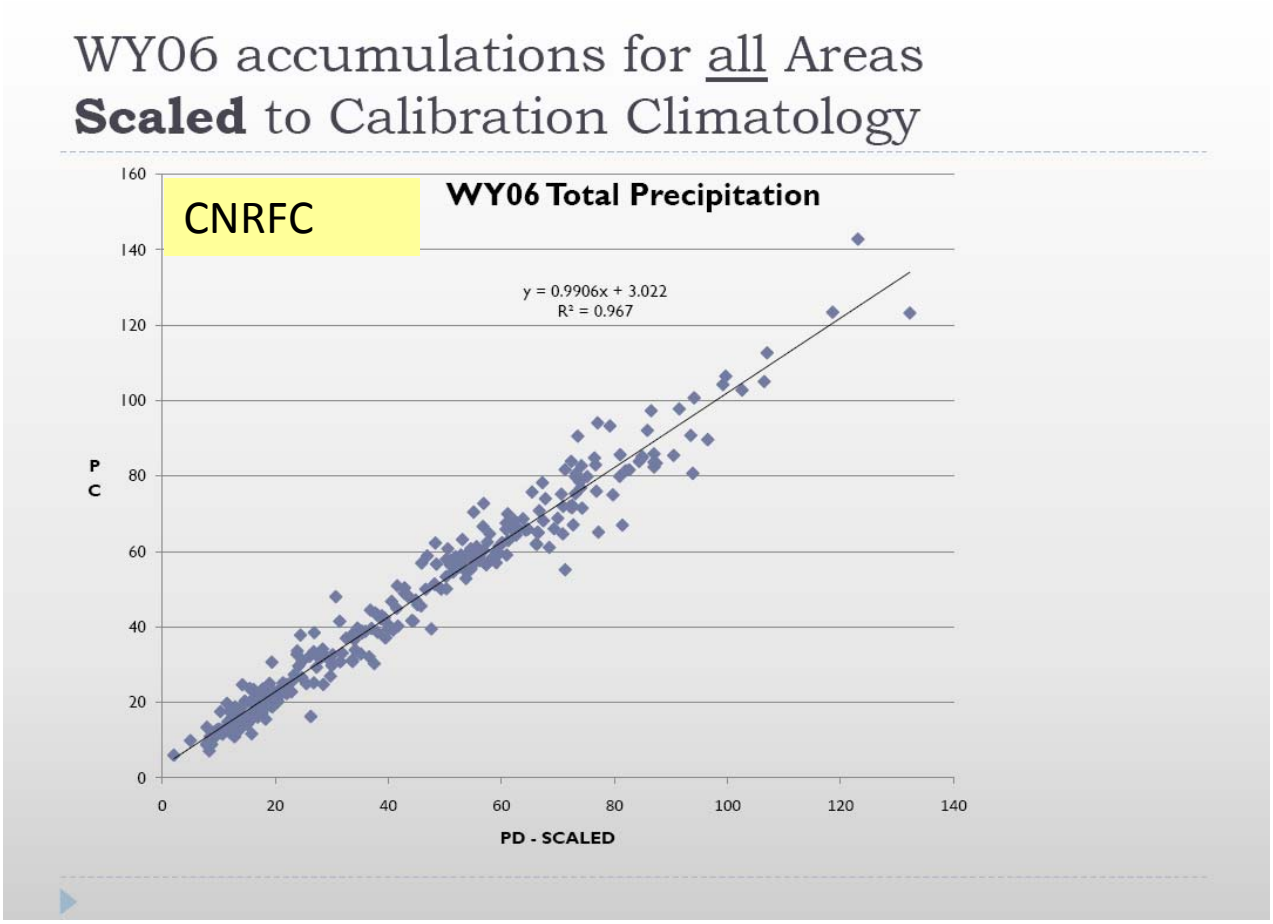

Figure 2. As in Fig. 1 except that a scaling factor based on long-term biases between OFS and MM-DQC precipitation has been applied, slightly improving bias and errors.

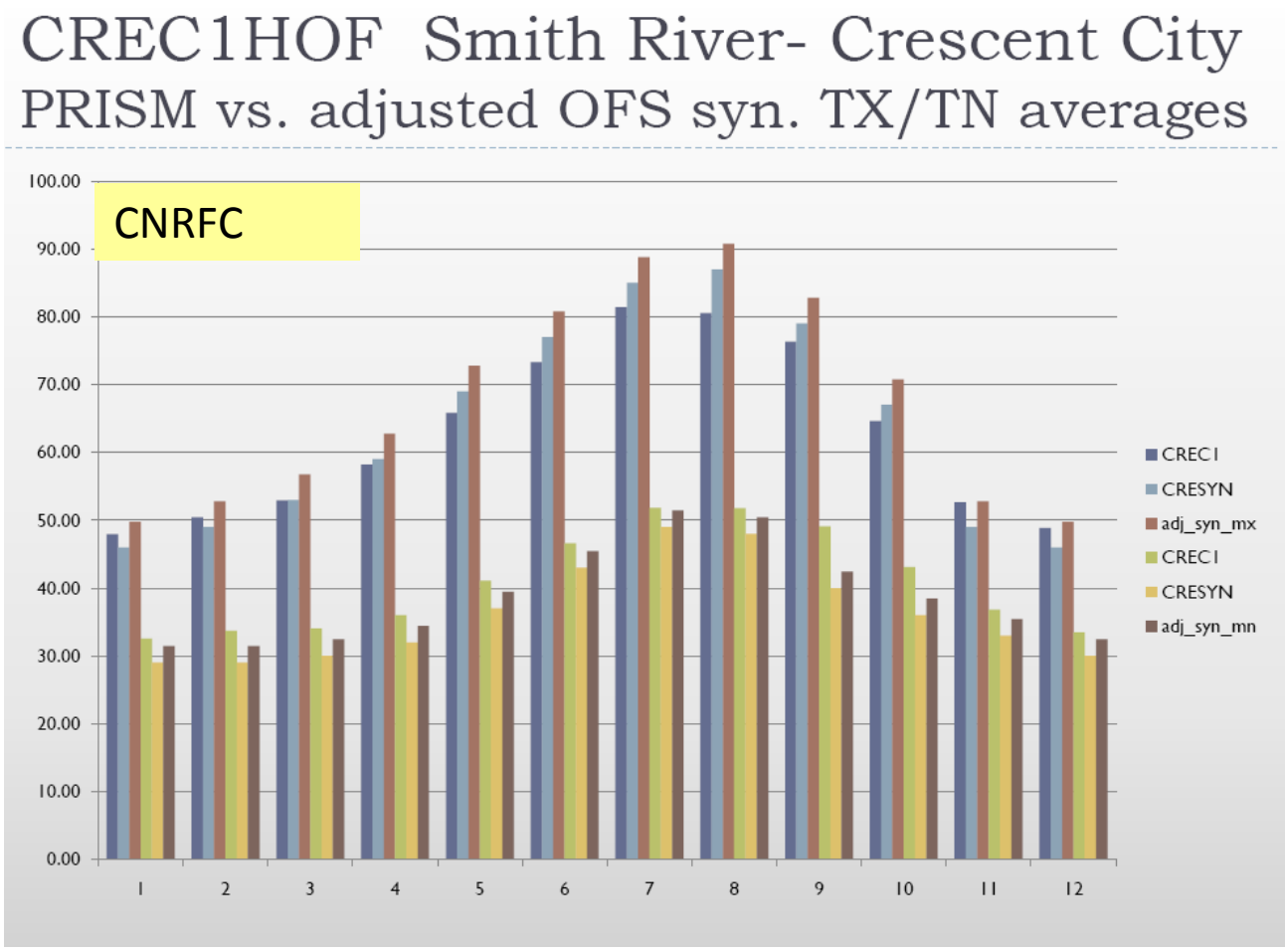

Figure 3. Maximum/minimum temperature scaling procedure for Smith River, Crescent City, for 12 month in WY2006. Climatology scaling and correction for different assumptions about mean basin elevation (adj_syn values) lead to improvements in estimating target values (labeled CREC1).

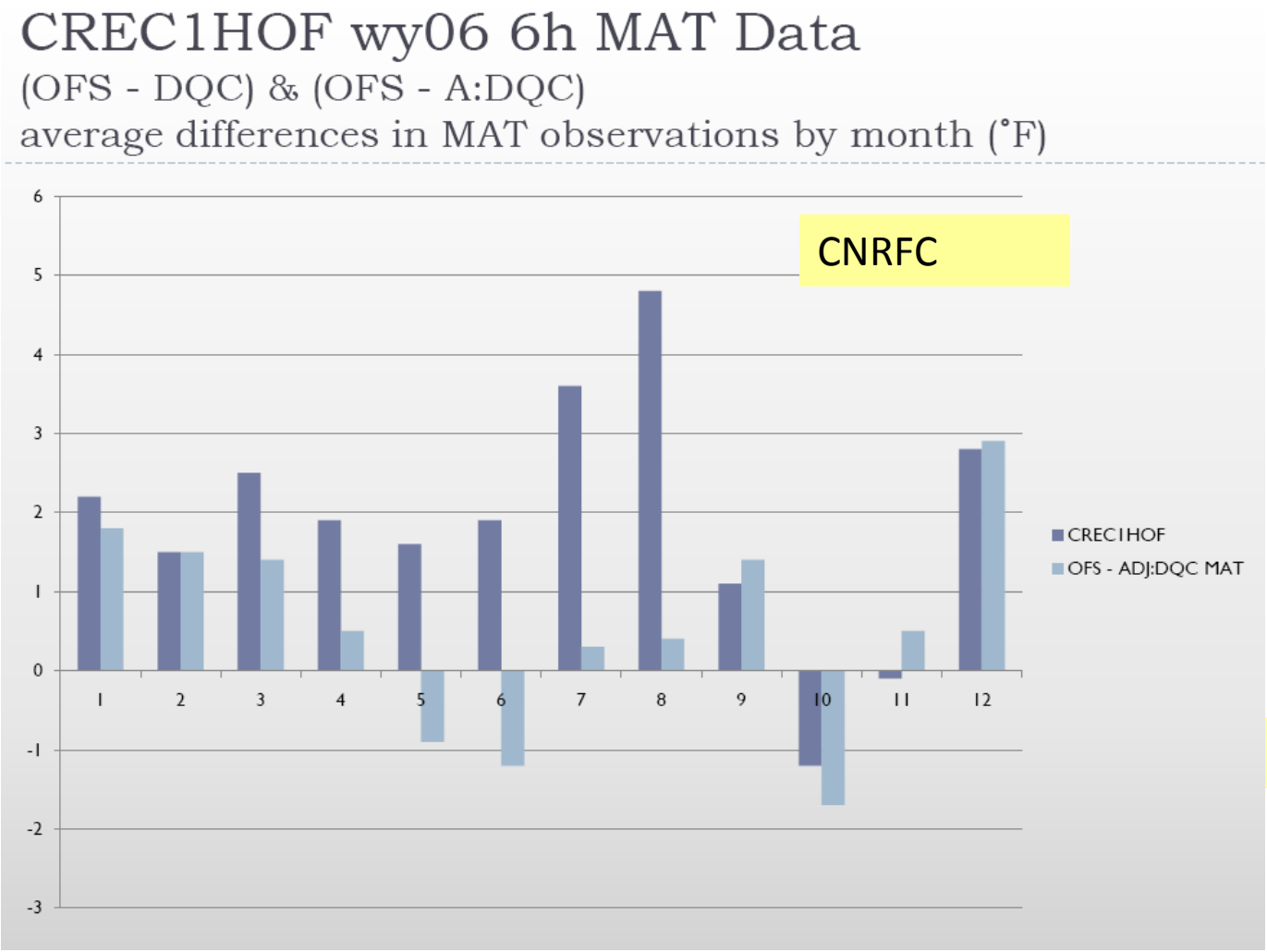

Figure 4. Temperature scaling procedure for Smith River, Crescent City. Scaling procedure as applied to 6-h mean temperatures from DQC leads to smaller errors (darker vs. lighter toned bars).

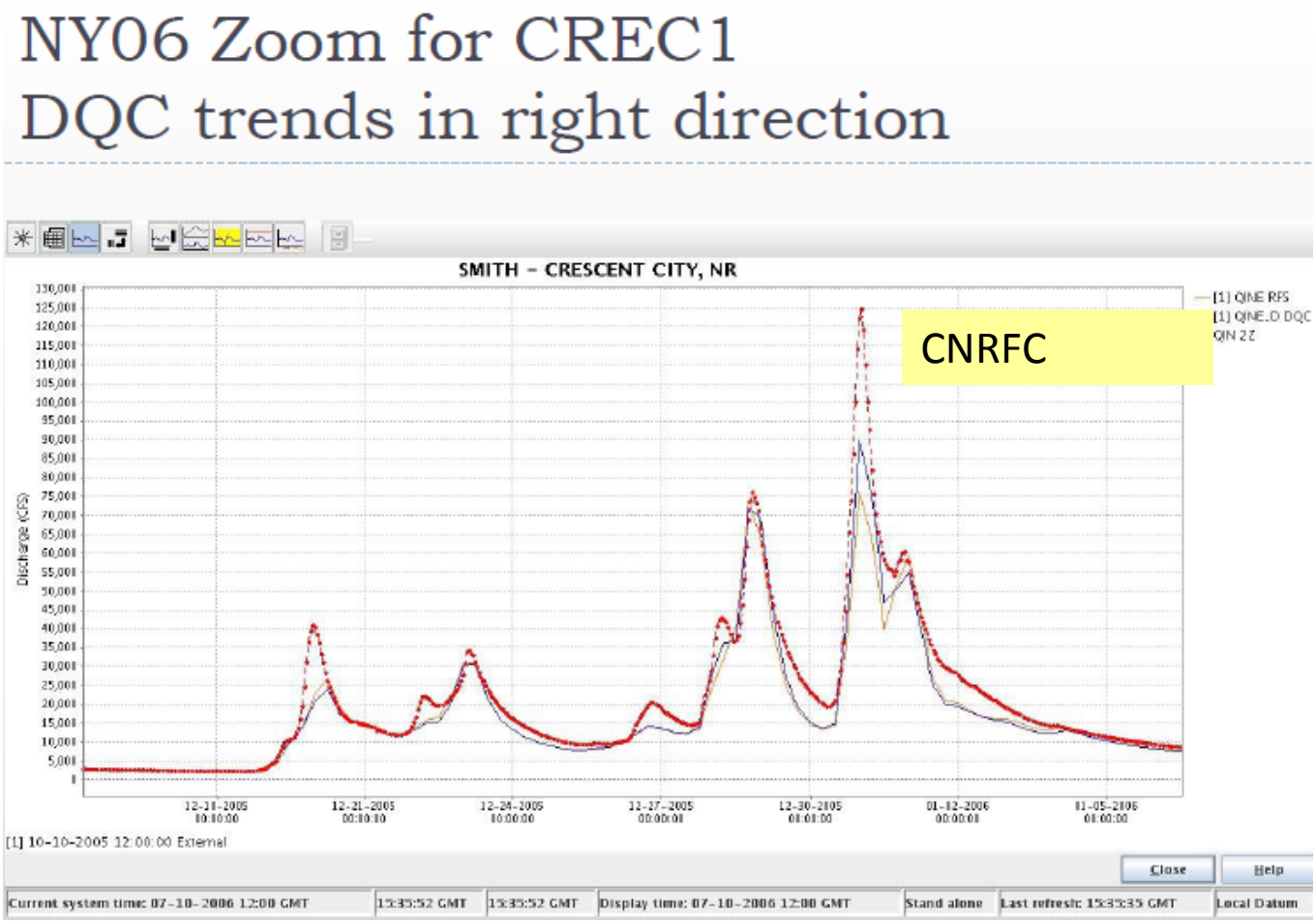

Figure 5. Comparison of flow simulations from RFS and DailyQC forcings. Over most of the year the DQC-based simulations were closer to observations than were the OFS-based.

5. ABRFC (Eric Jones; Briona Chester)

ABRFC staff executed comparison of NWSRFS vs CHPS forcings

They compared three variables: MAP, MAT, and potential evapotranspiration (PE) for all basins within the ABRFC domain

Approach:

1. MAP: Presented correlation by basin, mean absolute error by basin, mean error by basin for 'PRECIP WINTER - 10/1/2009 - 1/25/2010 -- a little over a year of records
2. MAT: Presented correlation by basin, bias by basin, and RMSE by basin for WINTER (10/1/09 - 1/13/10); Presented correlation by basin, bias by basin, and mean error by basin for temperatures near the freezing point (labeled FREEZE) (6/11/09 - 1/13/10);
3. Potential Evapotranspiration (PE): CHPS used PE computed from NCEP Real-Time Mesoscale Analysis (RTMA) and then gridded in the Gridded Forecast Editor (GFE) AWIPS application - Mean areal PE computed in CHPS; NWSRFS basins SAC-SMA basins tied directly to PE computed from METAR stations. A table of average values and differences from NWSRFS and GFE for 19 cases were presented

Findings on data comparisons:

1. MAP: Gridded gauge-radar forcings from the P3 process have been used operational for some years; differences between older and newer values are mainly due to computation changes along basin boundaries between NWSRFS and CHPS definitions
2. MAT:
   a. Generally high correlations and small biases between OFS legacy and GFE values over the eastern and central portion of the area, where terrain has a gradual slope (Figs. 6-7).
   b. Larger differences noted in mountainous areas
   c. Probably due to way RTMA computes temperature (lapse rate)
   d. May have to recalibrate SNOW 17
   e. Might subdivide basins based on elevation, which cannot be done in NWSRFS (discontinuous)
   f. Some lower correlations and larger biases were observed for cold temperatures near freezing (Figs. 8-9).
3. PE: May require some recalibration, especially in west/east basins. Results summarized in Table 1.

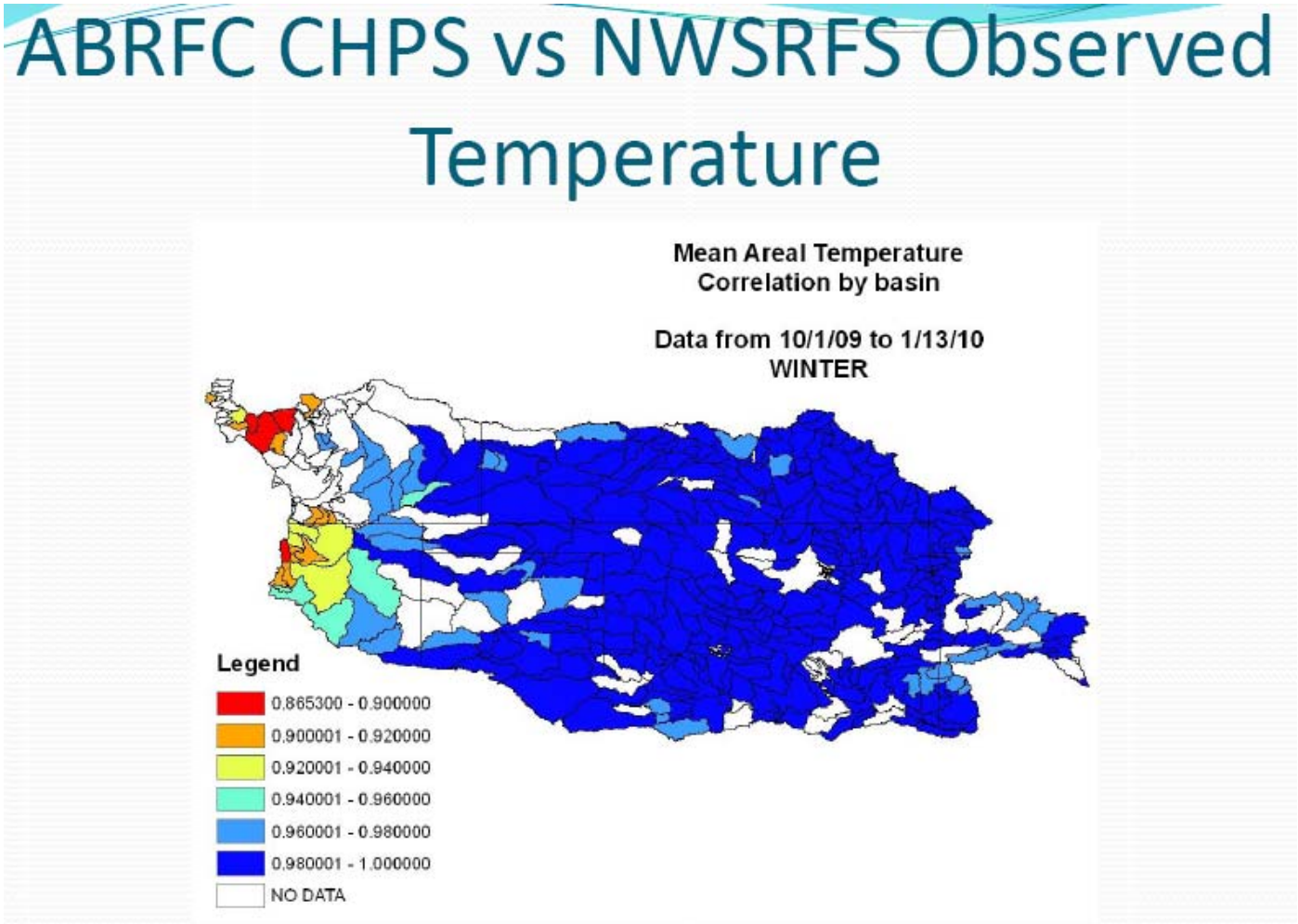

Figure 6. Map of MAT linear correlation by basin

ABRFC CHPS vs NWSRFS Observed Temperature

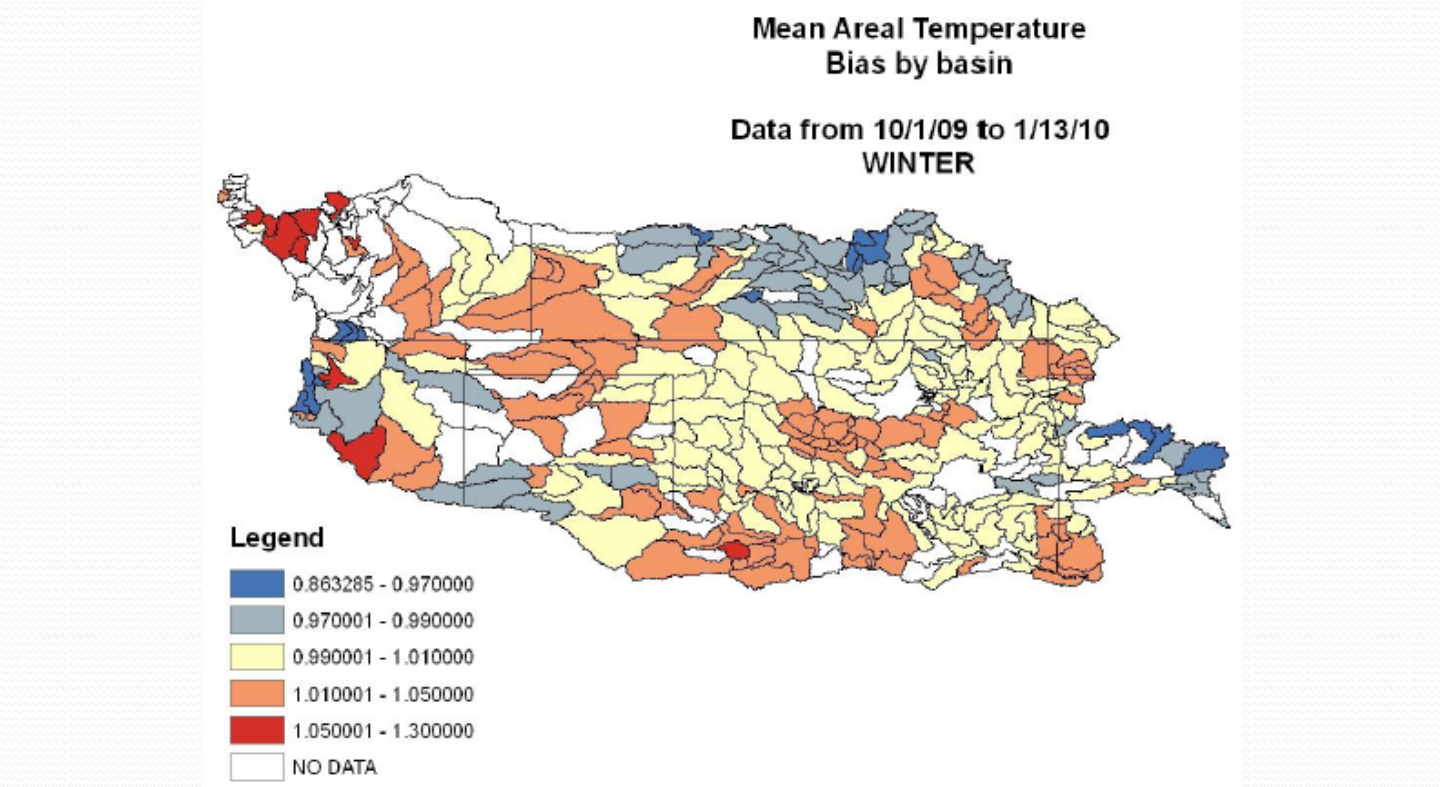

Figure 7. Map of MAT bias (deg F) by basin

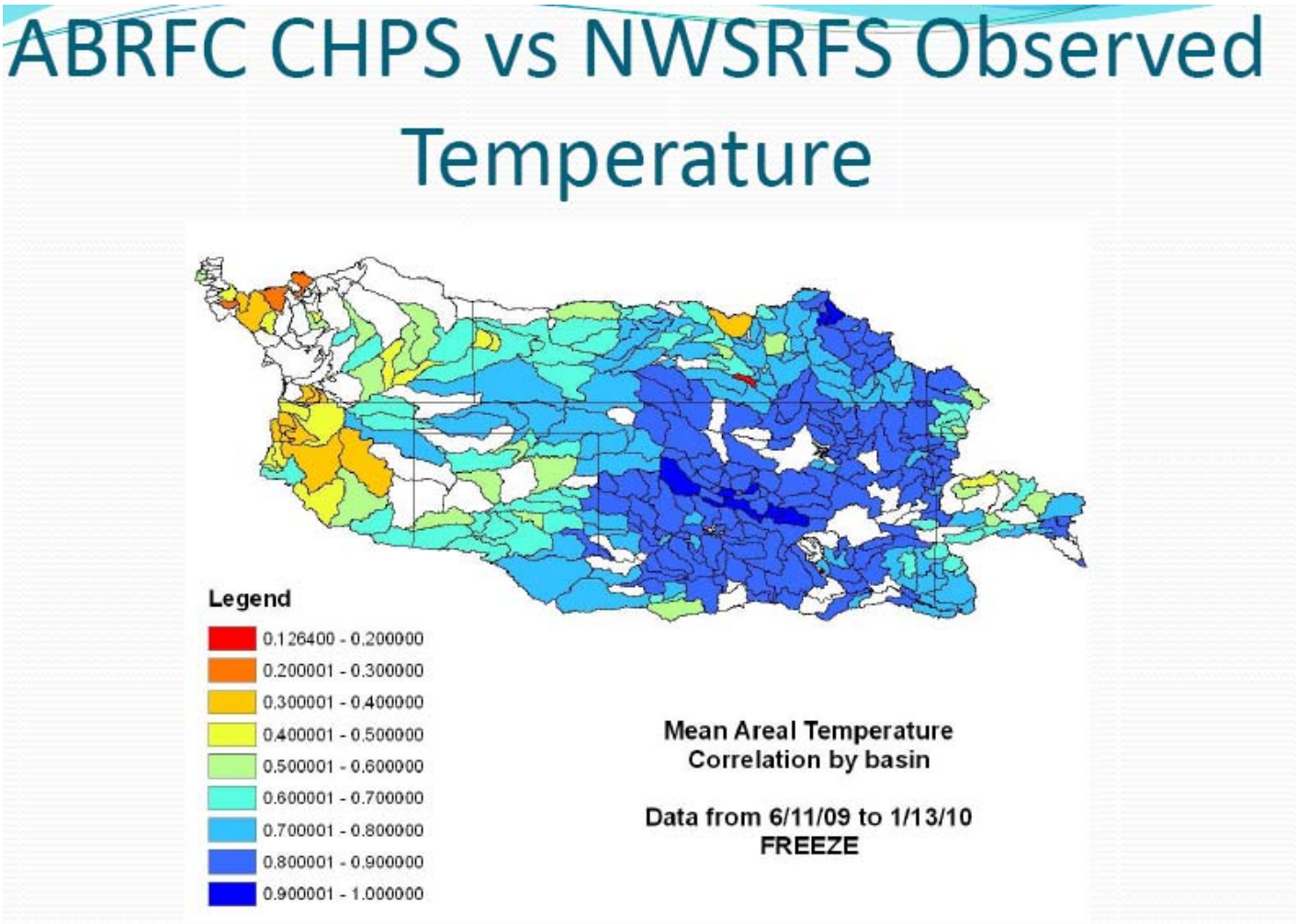

Figure 8. As in Fig. 6, but for cases with temperatures near freezing.

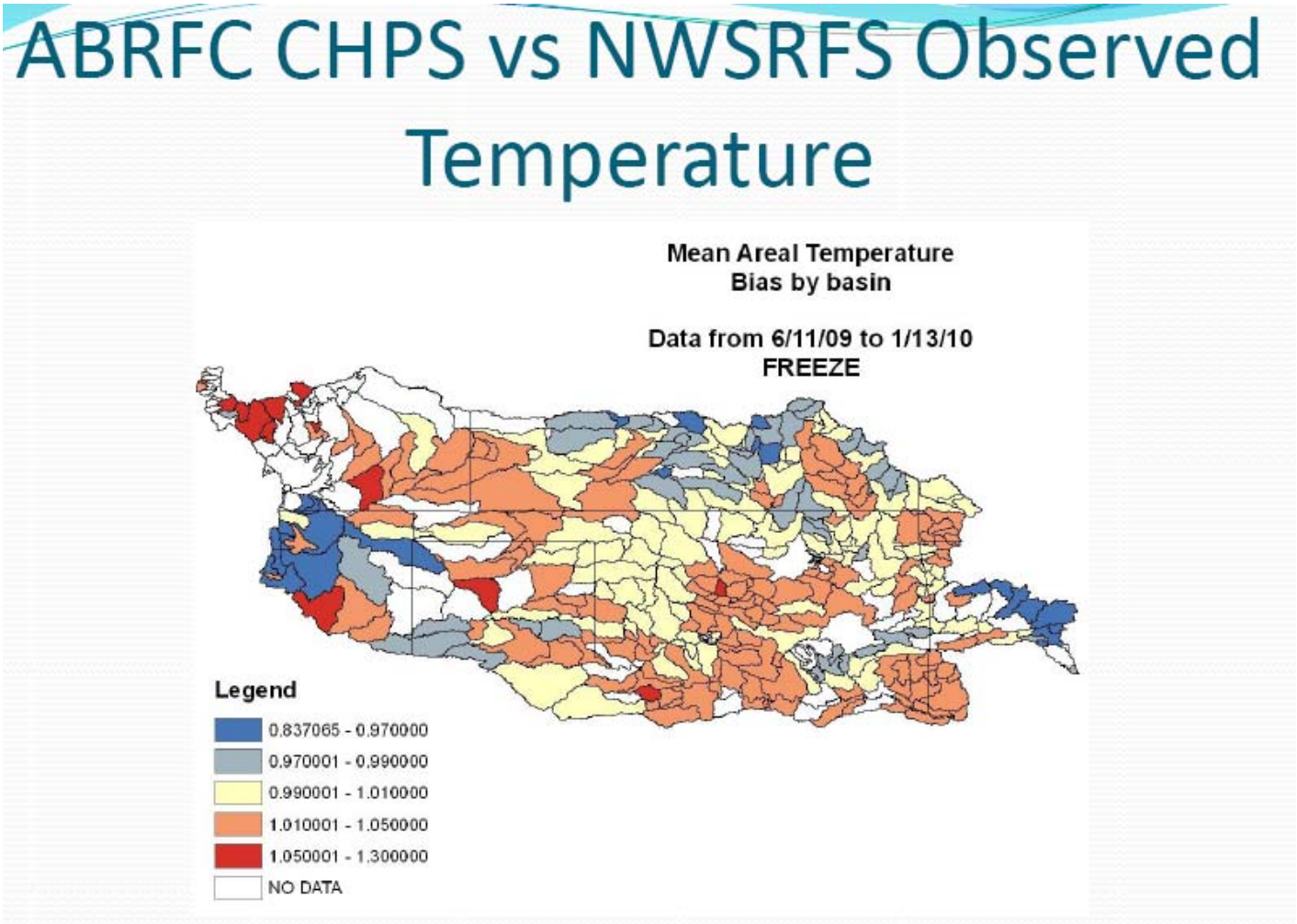

Figure 9. As in Fig. 7, but for cases with temperatures near freezing.

ABRFC PE Results

GFE to NWSRFS Biases				average NWSRFS (in)	average GFE (in)	Difference (in)
	ratio					
AAO	-0.00318			0.174225	0.173775	0.000
AMA	0.150627			0.189788	0.218376	0.029
CAO	0.031504			0.217128	0.223968	0.007
DDC	-0.00991			0.189415	0.187537	-0.002
EMP	-0.10864			0.169255	0.150867	-0.018
FSM	-0.03722			0.169202	0.162904	-0.006
GAG	0.014835			0.208387	0.211478	0.003
LAA	-0.01217			0.197581	0.195387	-0.002
LBB	0.09933			0.206882	0.227602	0.021
LIT	0.184013			0.134839	0.159855	0.025
LVS	-0.0487			0.203118	0.193226	-0.010
LXV	-0.01445			0.132204	0.130446	-0.002
MLC	-0.05416			0.174032	0.164608	-0.009
OKC	0.21413			0.166935	0.202758	0.036
PUB	-0.03298			0.204140	0.197435	-0.007
RVS	-0.03119			0.178387	0.172871	-0.006
SGF	-0.0781			0.162688	0.150011	-0.013
SPS	0.122699			0.200914	0.225565	0.025
TXK	0.030547			0.164032	0.169043	0.005
Average Bias =		0.021946			average inches =	0.004
					max =	0.036
max bias=	0.21413				min =	-0.018
min bias =	-0.10864				absolute inches =	0.012

Table 1. MAPE comparison between NWSRFS and GFE

6. NWRFC (Don Laurine)

Staff at NWRFC compared forcings OFS MAP and DQC MAPX only. Results were obtained for eight basins (Fig. 10). No hydrologic modeling runs were undertaken, but are planned.

Goals:

1. Compare the current operational process with the calibration MAP
2. Create a historical grid or Local Analysis of Record (LAOR) for use in the evaluation process and bias assessments
3. Develop a method to produce an unbiased grid MAP, representing the volumes used in the calibration phase
4. Consider methods that can be employed in future calibration activities that use historical grids in lien of point data

Approaches:

1. Generate quality and consistent hourly point precipitation from the point data network used in the calibration process using OHD-provided MAP preprocessor program
2. Create 6-h and 24-h station precipitation totals from hourly point data from above
3. Use DailyQC to generate historical 6 and 24 hour grids for the calibration period of 1949 to 2003
4. Compare calibration OFS MAP and new generated historical DQC by employing an instance of a stand alone system
5. Use CHPS and Excel to generate statistics for analysis

Issues:

1. MAP fails to accurately disaggregated the 6-h data to 1-h values
2. Need to address multiple time zones
3. Adjustments must be made in CHPS so MAP and MAPX represent the same time period
4. Need to split 55 years of grids into multiple groups due to CHPS limitation to 20 years of data for historical calculation

Results/Summaries:

1. The majority of sites (total 16 basins) analyzed exhibited R^2>0.80 over the 55 years (example in Fig. 11 for Snake River).
2. Considerable scatter for all sites and also need to look at categories
3. Bitterroot basins have the weakest MAP/MAPX relationships, possible because of especially rugged terrain (Fig. 12).
4. Bias between sites is not consistent
5. Bias for a site is not consistent from month to month
6. Analysis using data from1971-2000 only shows a minor improvement in the relationships
7. Process is very simple and easy to update

Future work:

1. Develop QC points for Western Oregon. Needed for final LAOR grid
2. Develop method to disaggregate values from MAP that were not time distributed
3. Analysis runs on 6-h data
4. Analysis using simulations and runoff to determine impacts of this solution
5. Automate the bias correction process

a

Site-Zone	R^2	Bias	MAE
DWRI1U	.7123	.915	.08
DWRI1L	.815	.952	.05
HHWM8U	.754	1.029	.06
HHWM8L	.764	.992	.05
PLDI1U	.832	.708	.04
PLDI1L	.841	.778	.03
SAKW1G	.894	.662	.16
SAKW1U	.893	1.002	.08
SAKW1L	.896	1.138	.08
WFBM8U	.413	.998	.07
WFBM8L	.439	1.159	.06
TOPI1	.865	.95	.05
DARM8U	.540	.883	.06
DARM8L	.598	1.080	.04
BONM8U	.701	.848	.06
BONM8L	.773	.832	.03

b

Figure 10. Results of multi-year comparison between OFS MAP and DQC-MAPX, for basins and subbasins in (a). Reduction of variance, multiplicative bias, and MAE (inches) are shown in table in (b).

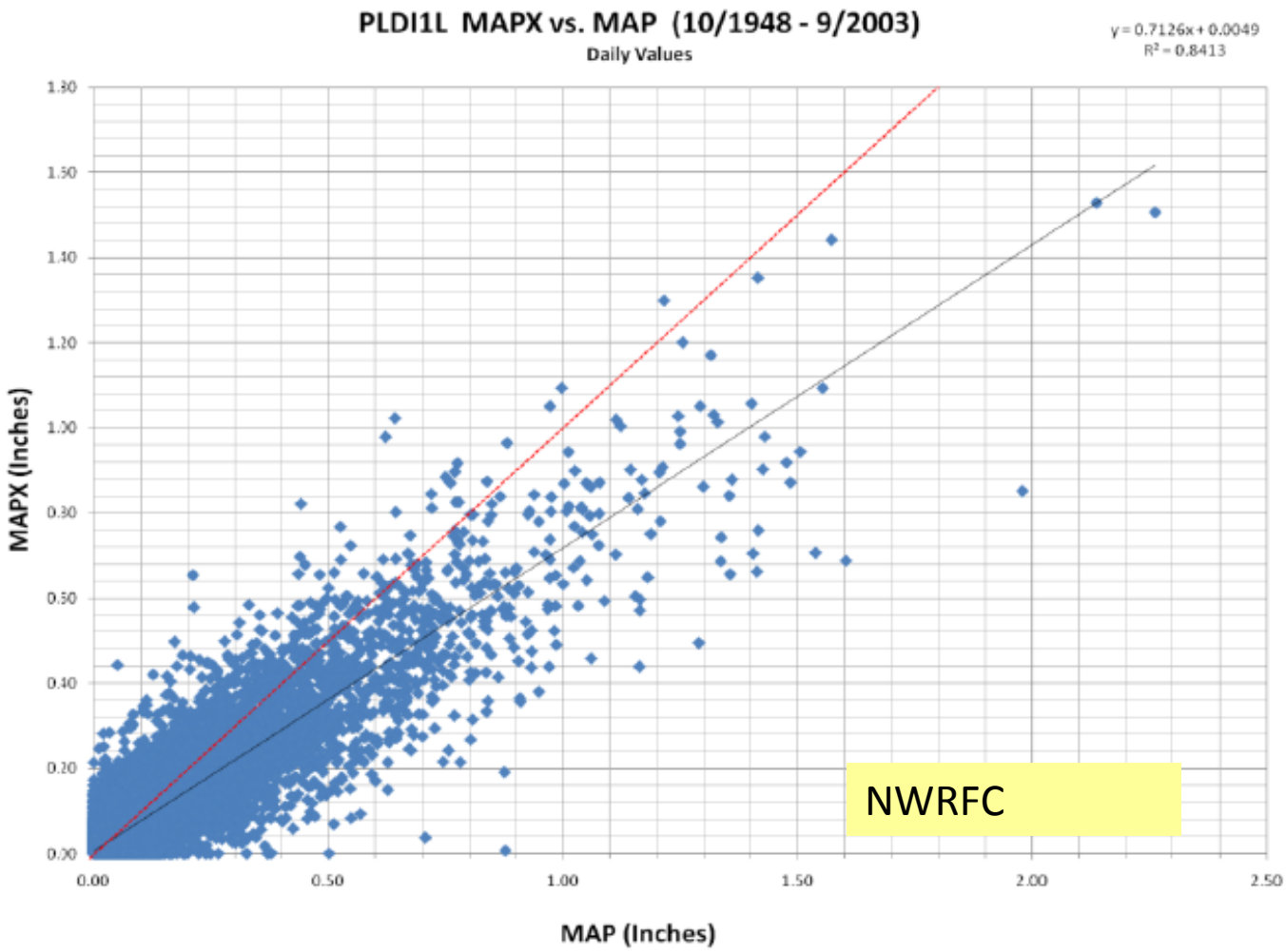

Figure 11. Comparison of mean areal precipitation based on OFS and DQC procedures; reduction of variance is 0.84 with some bias.

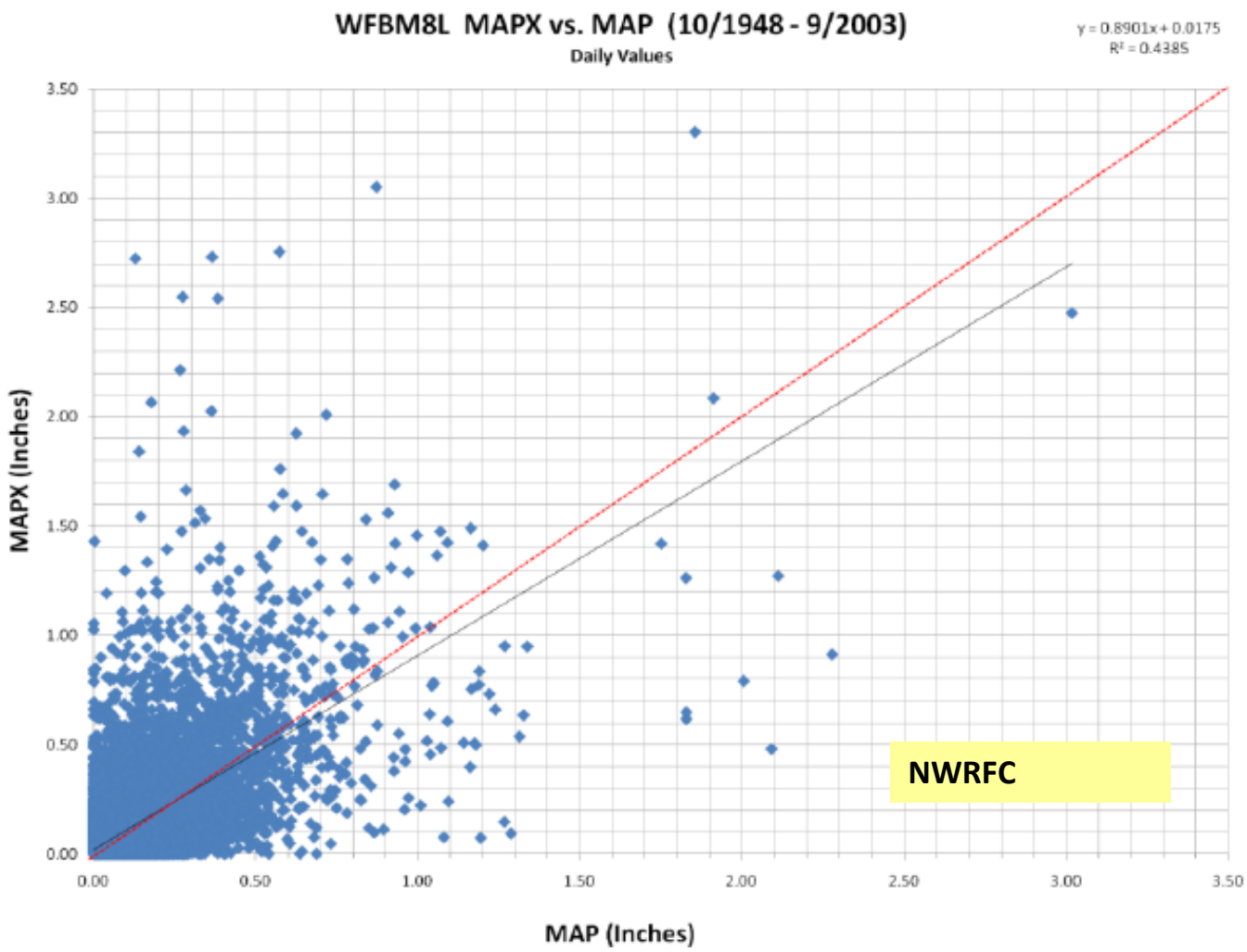

Figure 12. As in Fig. 11 but for a basin in the Bitterroot Mountains area. Reduction of variance is 0.43.

7. NERFC (Robb Shedd)

Compared three MAP forcings from OFS, DailyQC, and MPE (hourly radar based grids) and their hydrologic impacts

Compared two MAT forcings from observed (RUC13 or RTMA) grids, and forecasted (WFO Intersite Forecast Coordination and MOS outside of WFO region).

During this period use of radar data in MPE appeared to give few advantages over gauge-only input (Fig. 13).

Approaches:

Did hindcast runs for the period of 1/21/2010 - 4/5/2010 on four basins

Results:

Both DailyQC and MPE perform better than OFS, as shown in Table 2.

Summaries:

1. Initial assessment looks very promising. The extra input QC required by DailyQC resulted in improved forecasts

Additional work:

1. More detailed comparisons of different forcings
2. Longer hindcast analysis at more points
3. Continue to develop DailyQC station lists

March 2010 Precipitation Comparison - NERFC

DailyQC (Gauge input) **Hourly MPE (Gauge-radar input)**

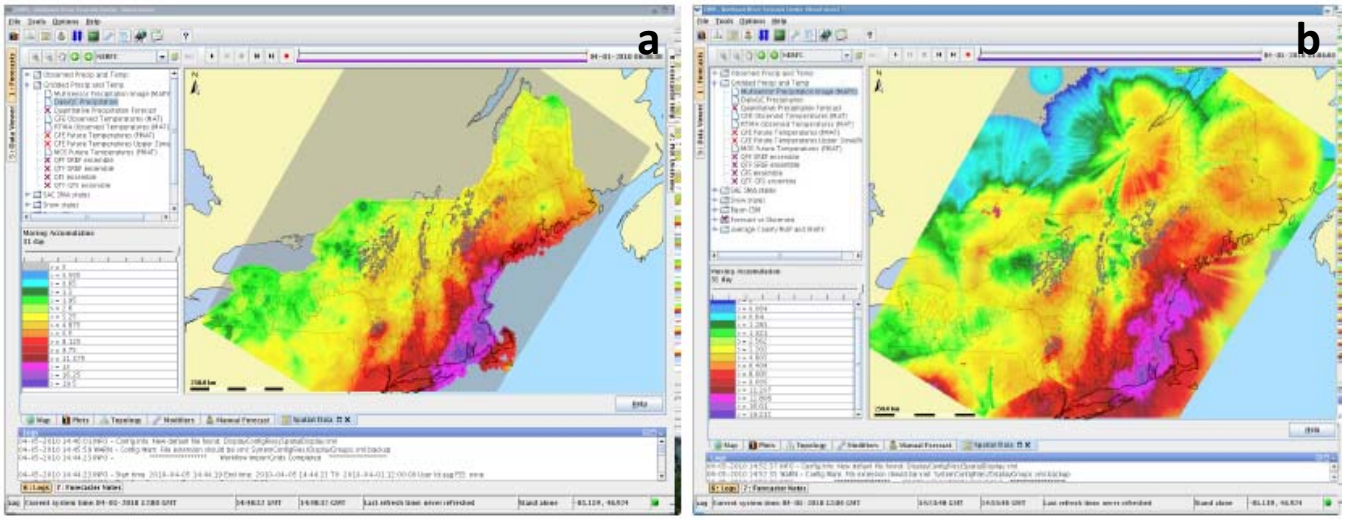

Figure 13. Monthly total precipitation from (a) gauge-only DQC and (b) gauge-radar MPE algorithms.

Correlations

	Cranston	Woonsocket	Yantic	Maynard
OFS	.948	.985	.958	.966
DailyQC	.959	.990	.952	.980
MPE	.959	.987	.984	.977

Lowest Highest

NERFC

Table 2. Comparison of correlations between simulated and observed flows using different forcings

Maynard, MA - MPE

NERFC

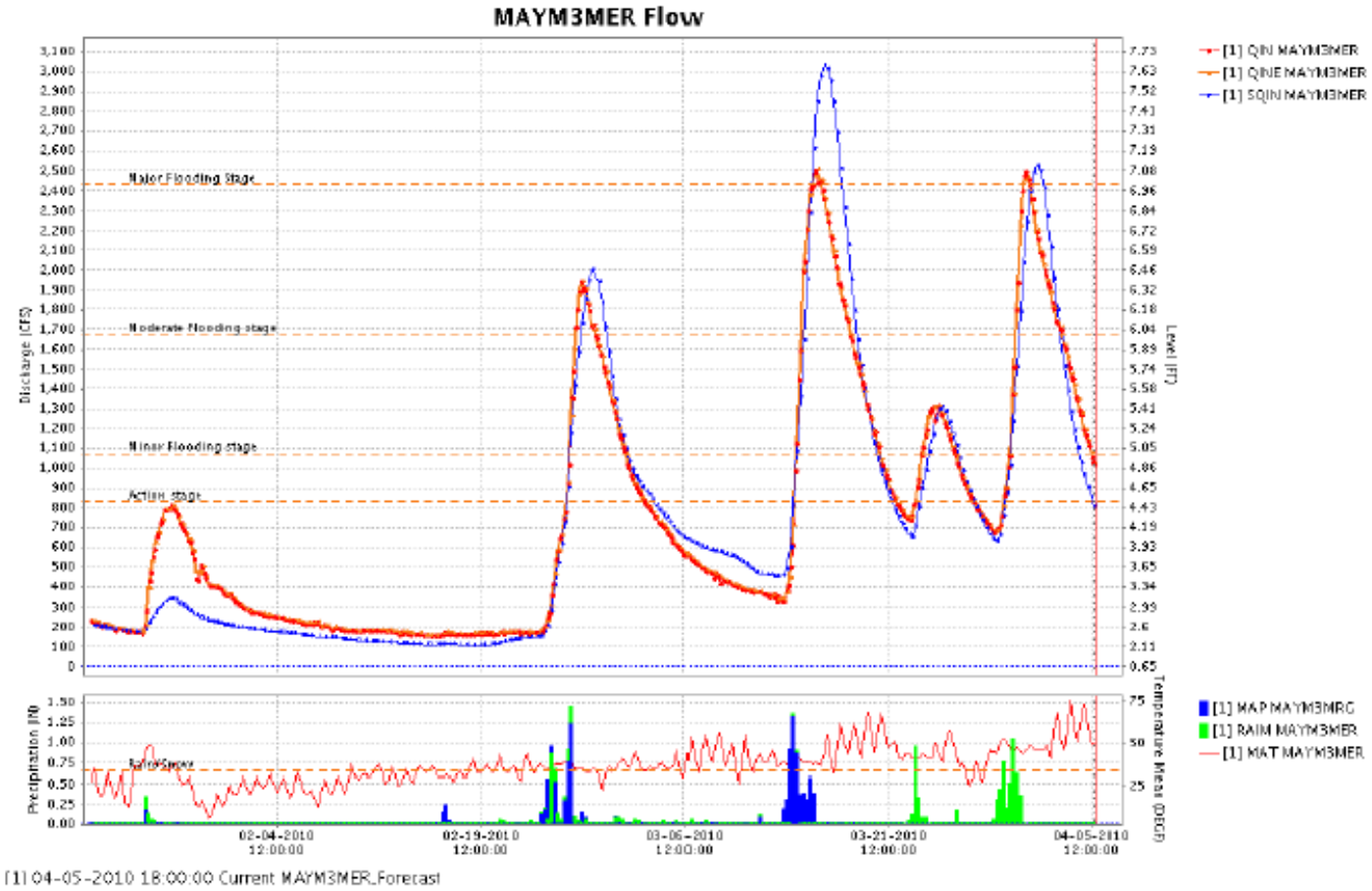

Figure 14. Observed (QIN and QINE) and CHPS simulation (SQIN) of flow for major precipitation and melt events at Maynard. Precipitation from MPE (radar-gauge). Temperature, rain, and melt traces are at bottom.

Maynard, MA -DailyQC

NERFC

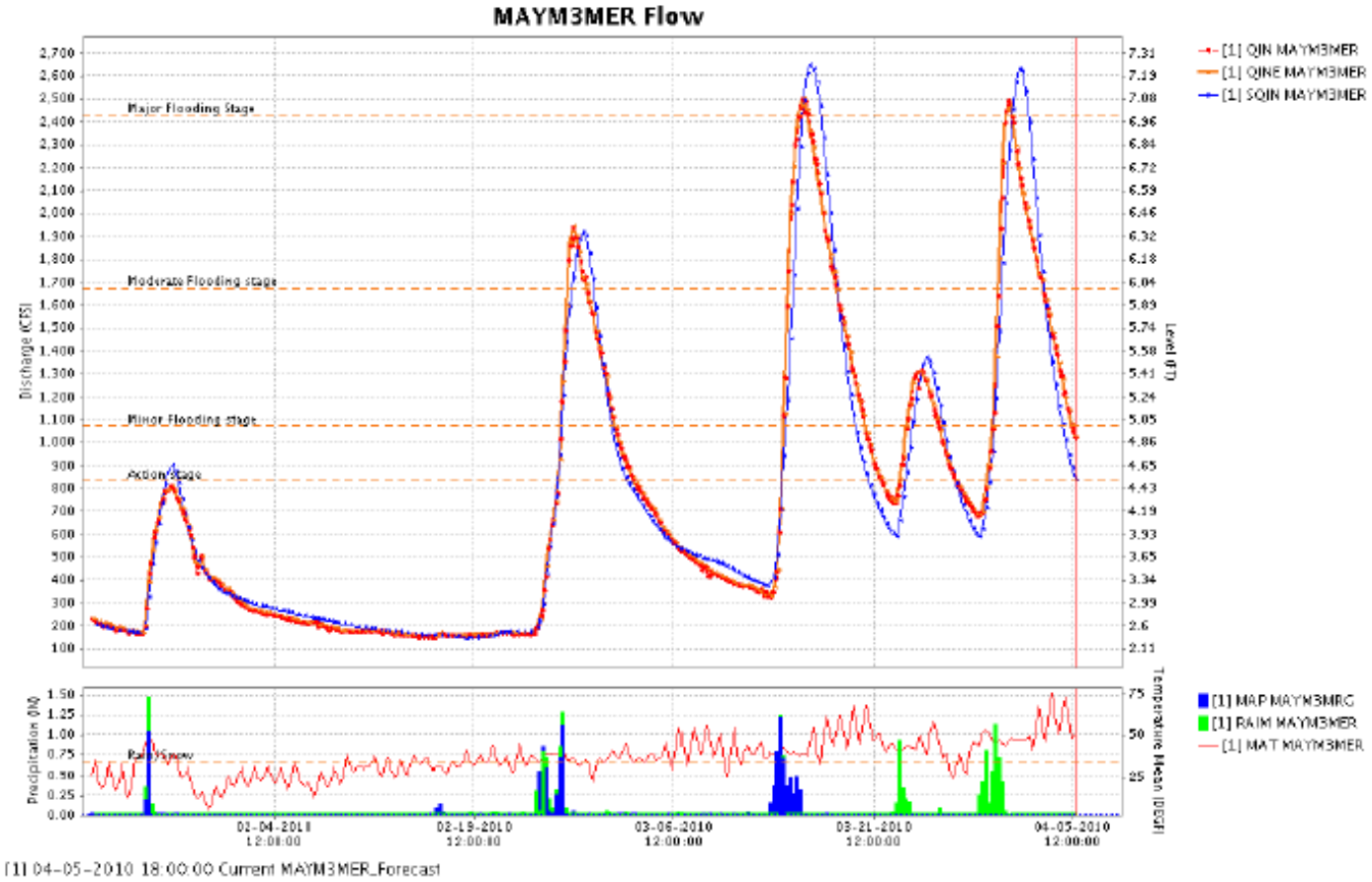

Figure 15. As in Fig. 14, except precipitation is from DQC (rain gauge).

8. Summary

A goal was set to make a transition from deriving meteorological forcings from point data to deriving them from gridded input. The quantities involved were basin-average precipitation, temperature, and potential evapotranspiration.

The studies were carried out to obtain documentation on differences between currently operational forcings as they were produced by legacy preprocessors and the same forcings when produced from gridded data.

Results indicated that the two analysis methods (point data to basin average vs. grid to basin average) can yield substantially different results, depending on the nature of the terrain and the density of the observing network. Given flat terrain and a dense gauge network, as in the ABRFC area, we can expect basin-average precipitation to be insensitive to the areal-averaging method in the cool season and basin-average temperature is generally insensitive to the derivation method.

The hydrologic impacts from forcings processed with DQC, rather than legacy OFS, were generally positive, as shown in Fig. 5 from CNRFC and Table 2 from NERFC.

One method for adjusting grid-based forcings toward better agreement with legacy point-based values was tested successfully by CNRFC staff, for temperatures.

It appears that a multi-decade high-resolution gridded forcings archive is necessary to recalibrate existing lumped models and to supply input for future efforts to calibrate distributed models. Any meaningful hydrologic forecast verification statistics will also depend on a longer simulation period.

Ultimately, a unified approach and statistics might be better to evaluate the differences between different methods, different basins, and different RFCs. Within OHD, a plan is being developed to undertake such studies. This plan is being developed with the goal of serving both calibration needs and those of partners in the Integrated Water Resources Science and Services agreement.

Acknowledgements

We wish to thank the study authors listed above and other RFC staff who executed and reported on experimental results, and provided guidance on interpretation.

References

Wakefield, S., 2008: Information Generation Section.
[Available from http://gfesuite.noaa.gov/EFTHome.html]

Horel, J. and B. Colman, 2005: Real-time and retrospective mesoscale objective analysis. *Bull. Amer. Meteor. Soc.*, **86**, 1477-1480.

Glaudemans, M., P. Tilles, and B. Lawrence, 2009: Interactive quality control and operational product generation of hourly multi-sensor precipitation estimates in the NWS. *Preprints, 25th Conference on International Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Amer. Meteor. Soc., **5B.3**.