Tara Jensen

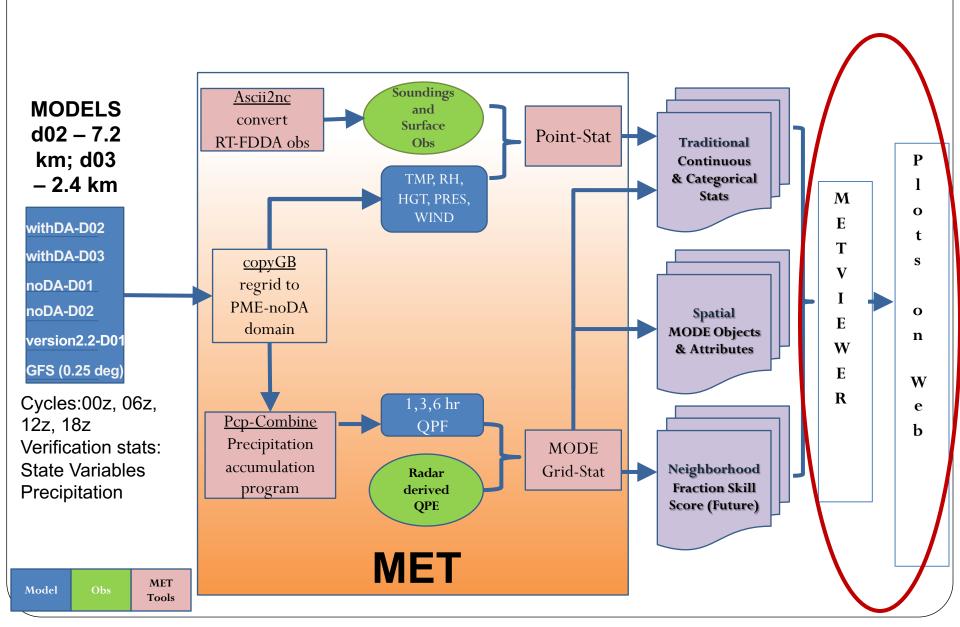
METplus Tutorial

Jul 31 – August 2, 2019

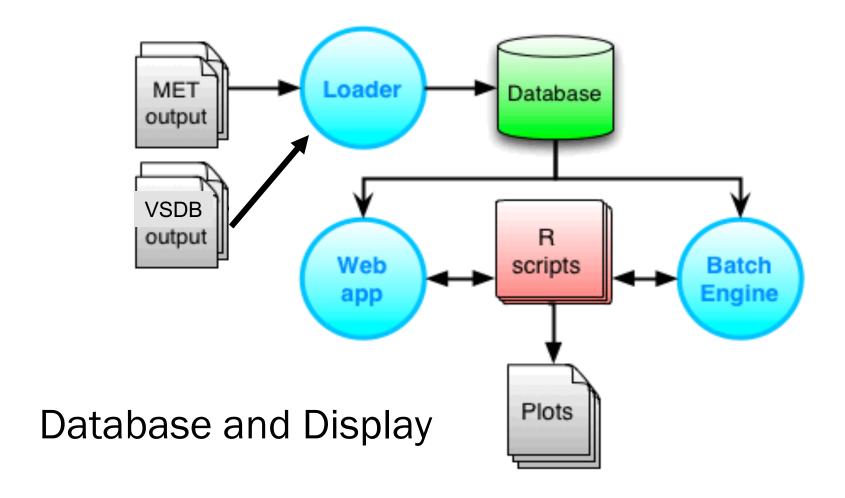
Monterey, CA



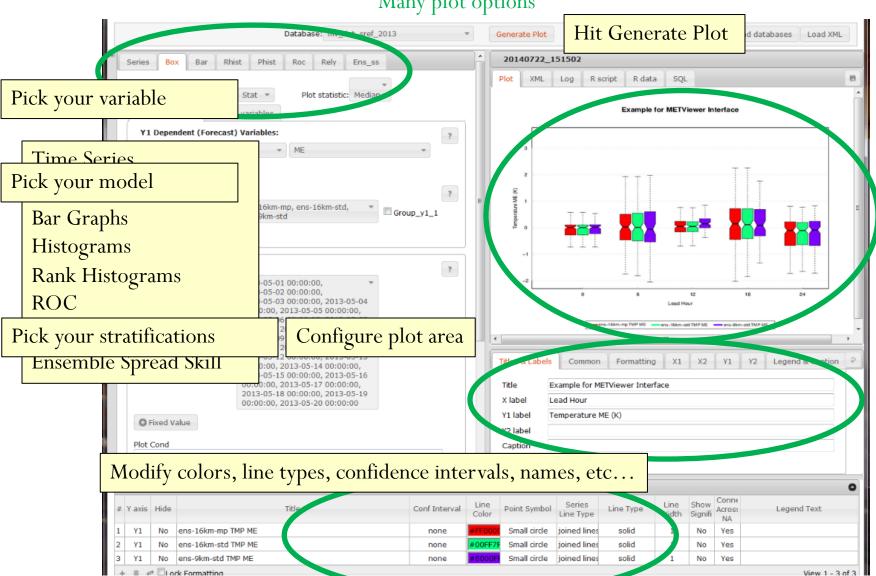
Example Design of a Verification System



METviewer Components

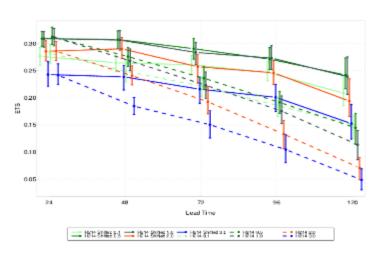


http://www.dtcenter.org/met/ metviewer/metviewer1.jsp

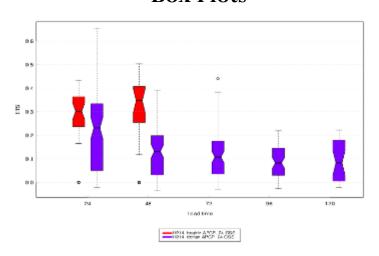


Examples of plots you can make

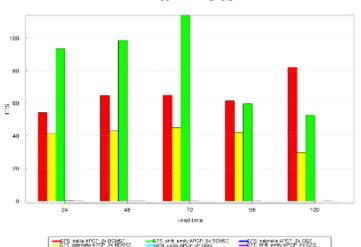
Time Series



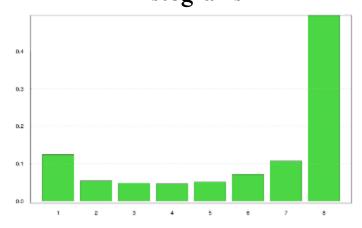
Box Plots



Bar Plots



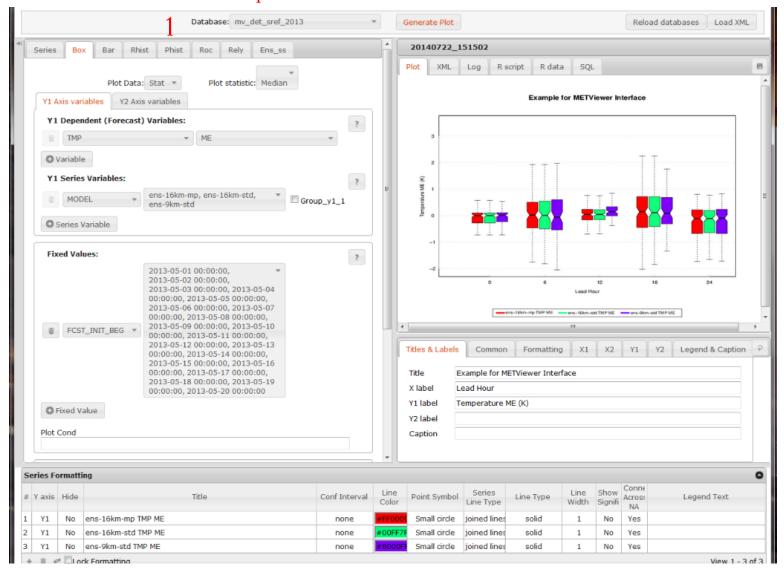
Histograms

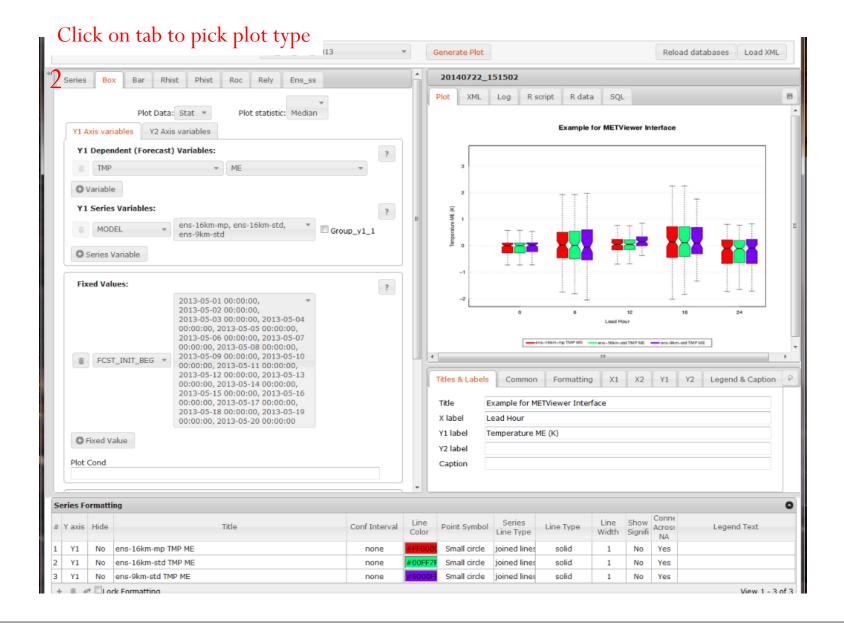


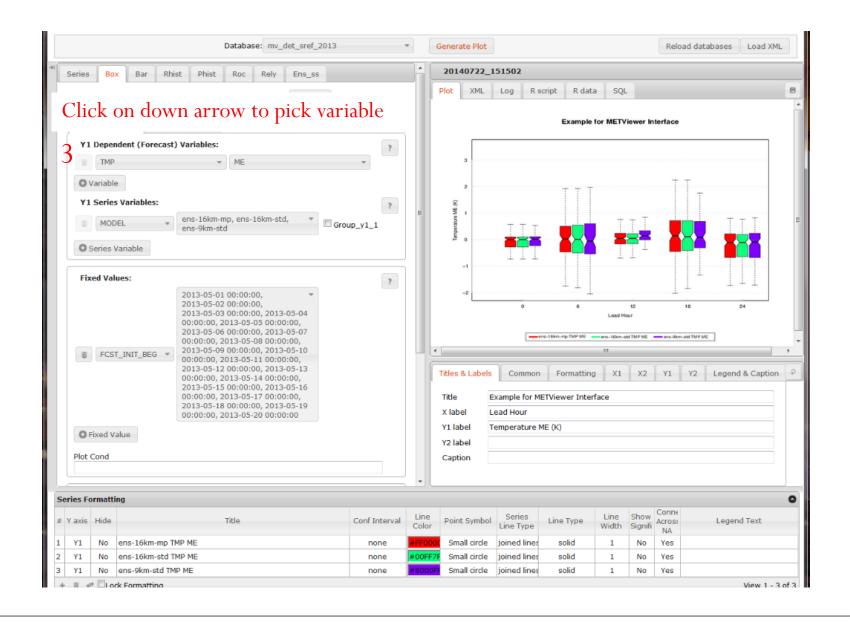
DTC's METviewer Instance

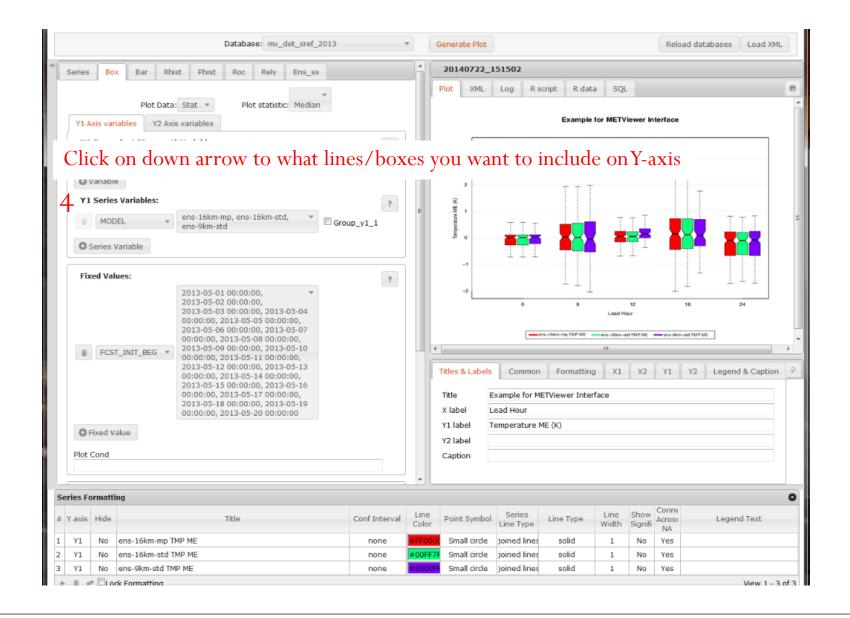
- http://www.dtcenter.org/met/metviewer.jsp
- Always Select :
 - 1. Database
 - 2. Plot Type
 - 3. Variable (Y1 Dependent variable)
 - 4. What the lines are (Y1 Series variable)
 - 5. Any stratifications like forecast levels and verification masking regions (FCST_LV, VX_MASK)
 - 6. X-axis (Independent variable)
 - 7. Aggregation Statistics: CTC (FHO) or SL1L2
 - 8. Whether you want to Mean or Median plotted

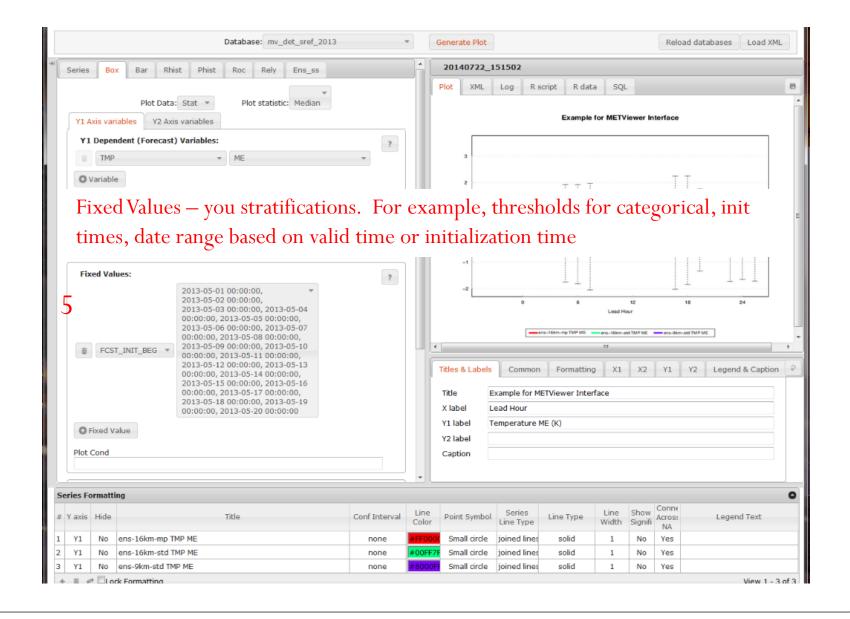
Click on down arrow to pick database



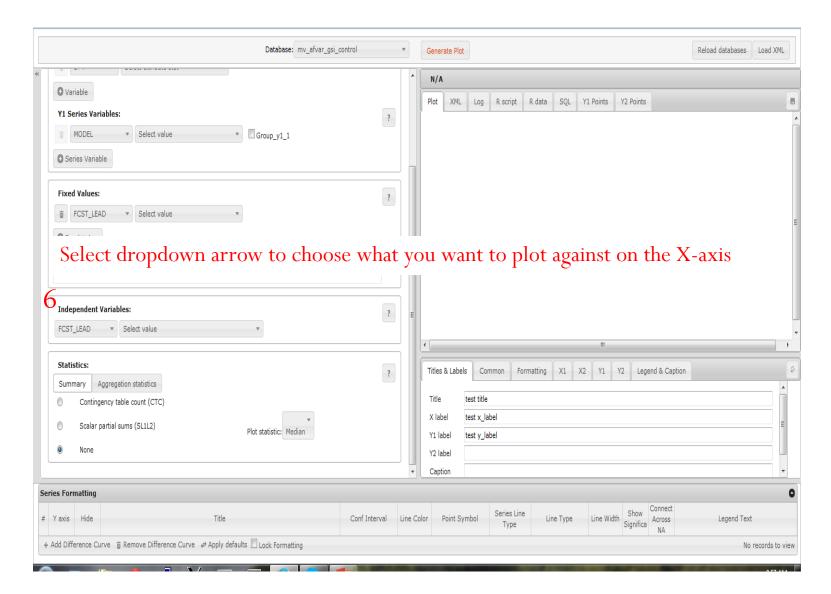




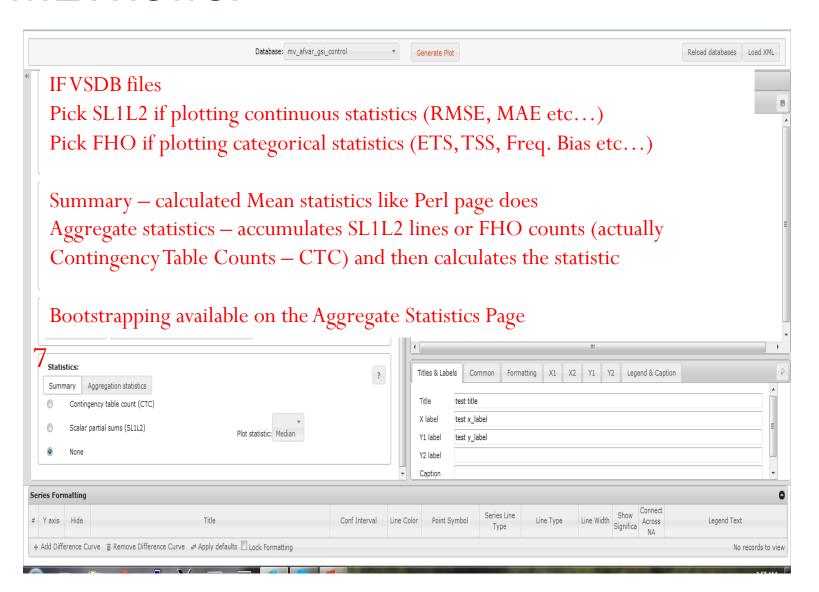




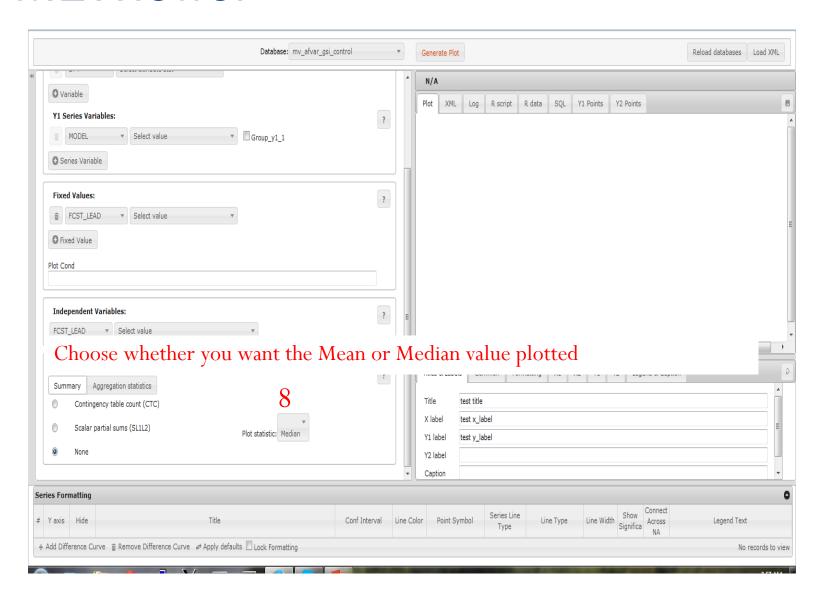
Scrolled down to see remainder of page

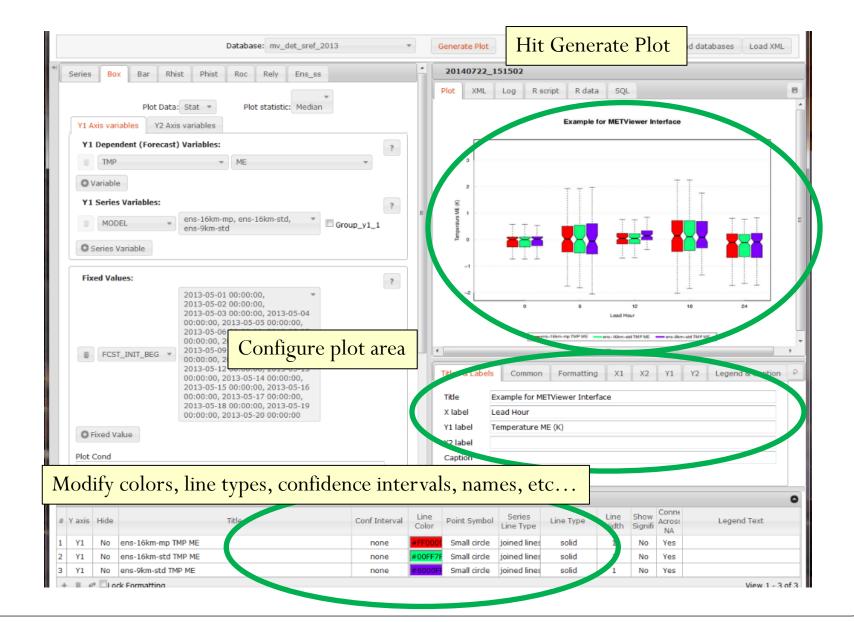


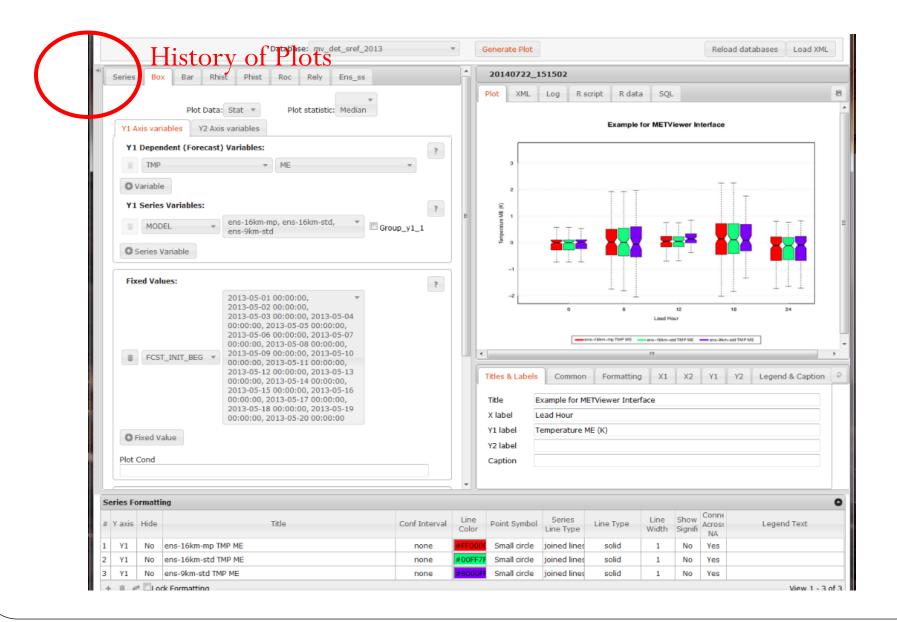
Scrolled down to see remainder of page



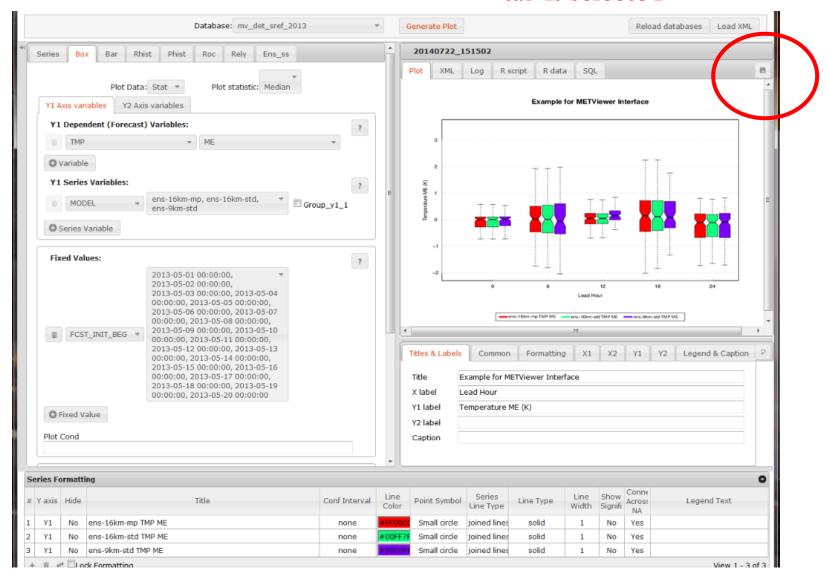
Scrolled down to see remainder of page



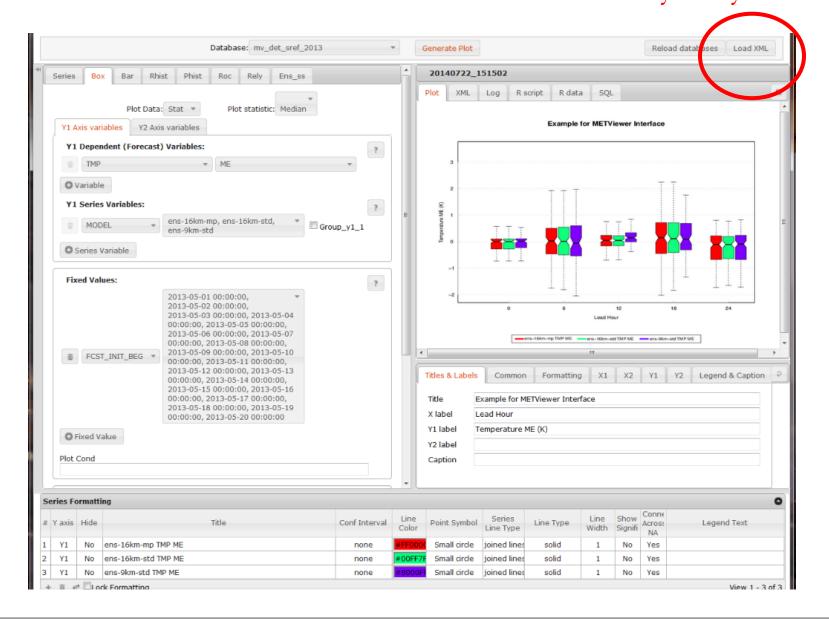


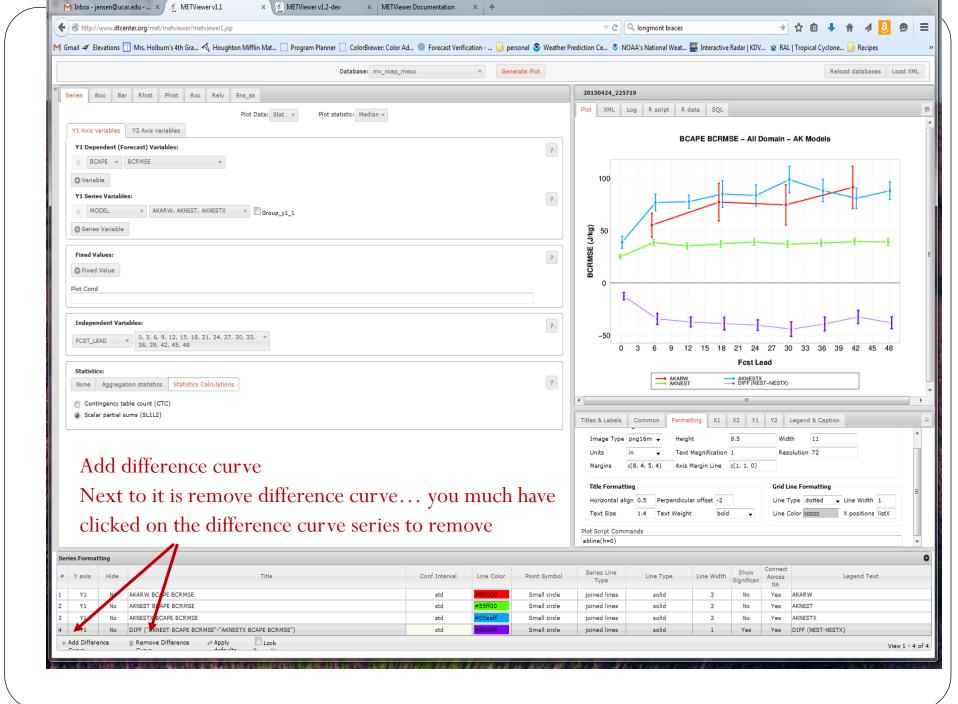


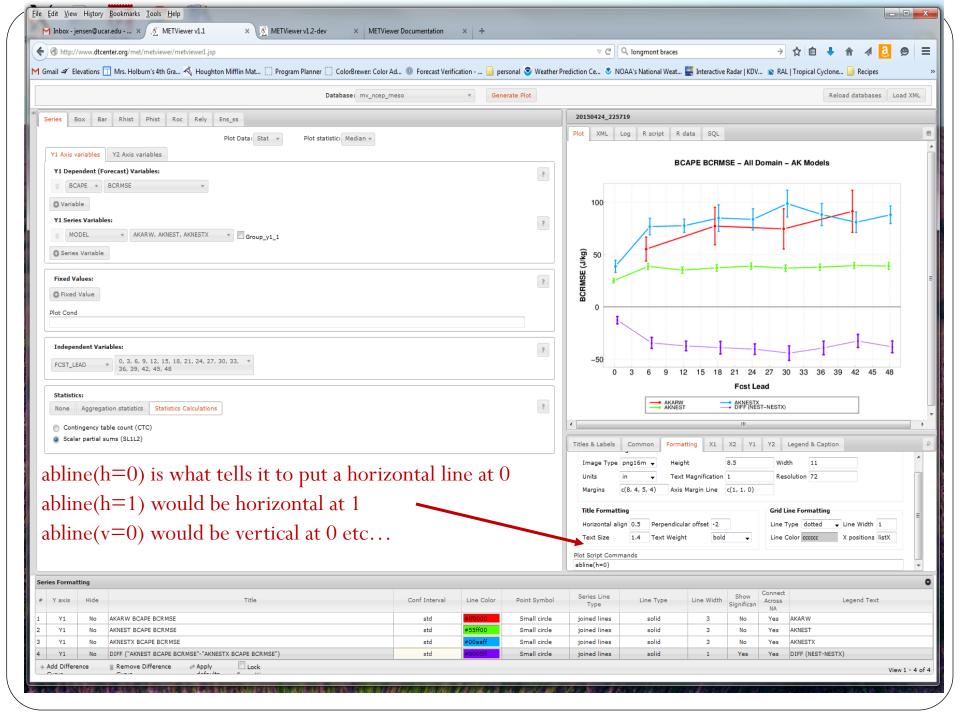
Save Plots, Save XML, Save Data based on which tab is selected

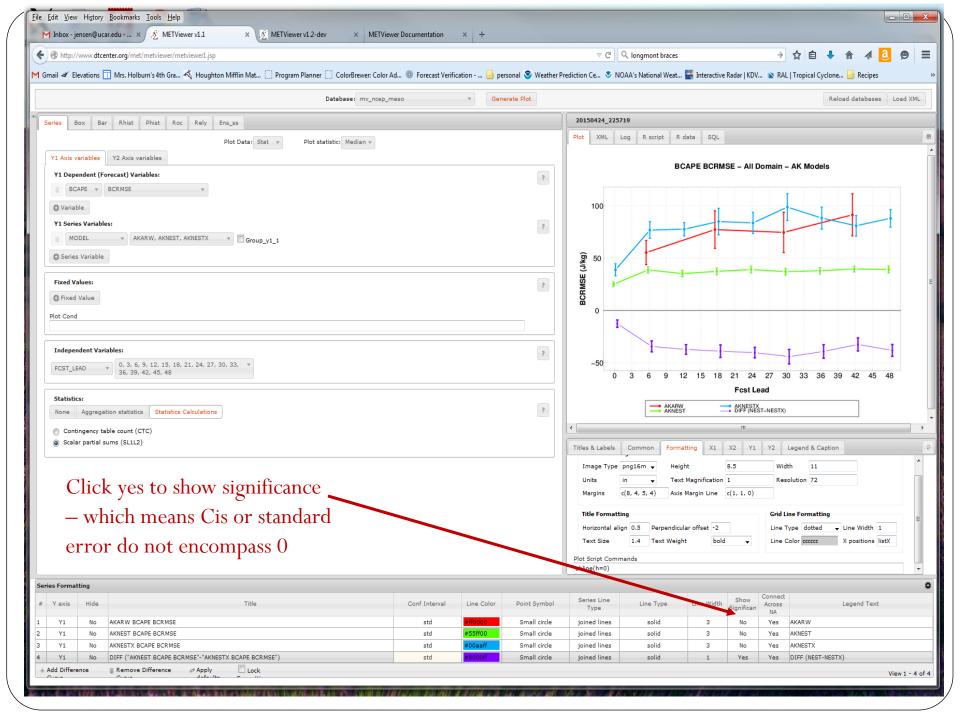


Upload XML scripts from your system





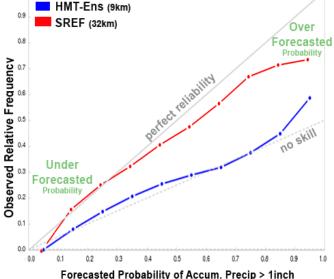


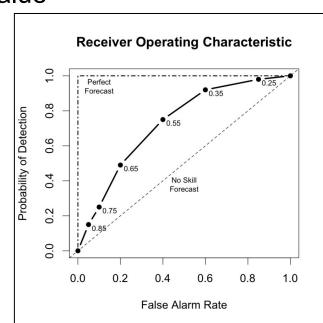


MET Probability Output (line-types)

- Output written to MET .stat file and, if desired, to individual text files:
 - PCT Probability Contingency Table Counts
 - PSTD Probability Contingency Table Scores
 - · Brier Score, Reliability, Resolution, Uncertainty, Area Under ROC
 - PJC Joint/Continuous Statistics of Probabilistic Variables
 - Calibration, Refinement, Likelihood, Base Rate, Reliability points
 - PRC ROC Curve Points for Probabilistic Variables
 - ECLV Economic Cost Loss Value

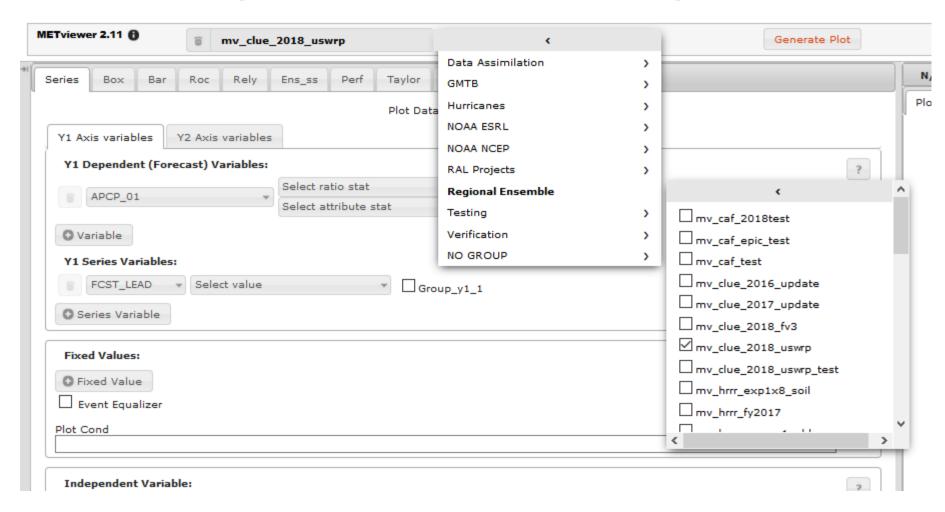
PJC SREF (32km)



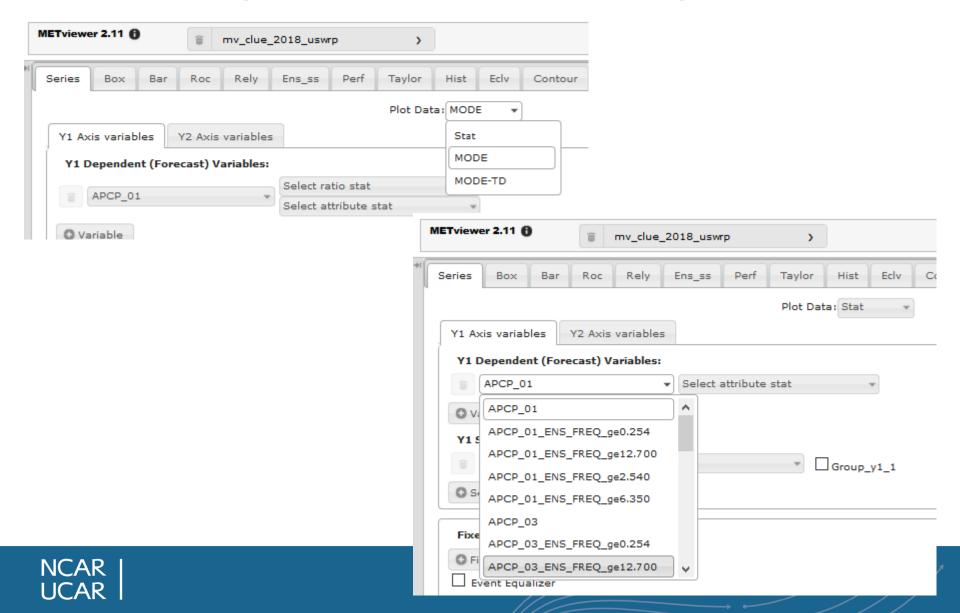


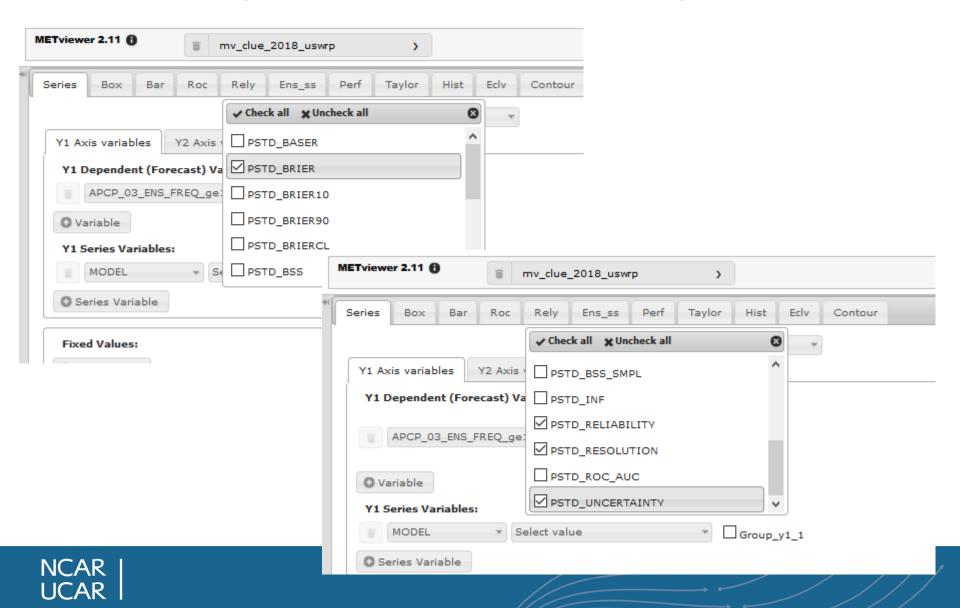
We will use these data for the Verificatio n Exercise

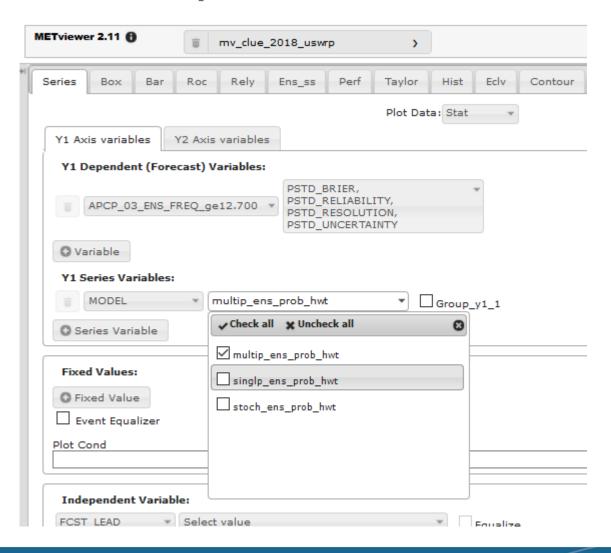
PRC



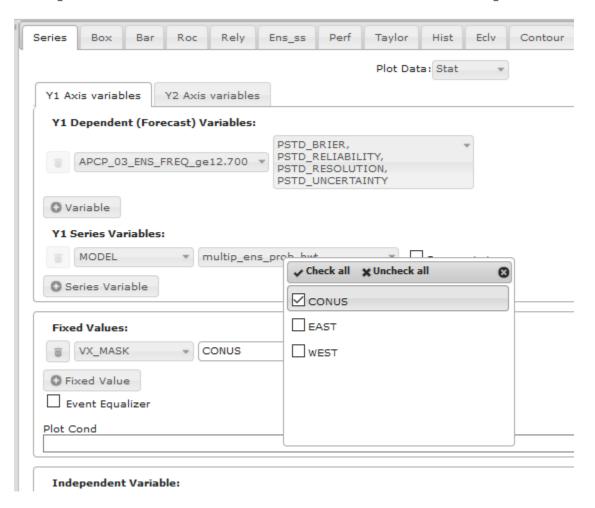




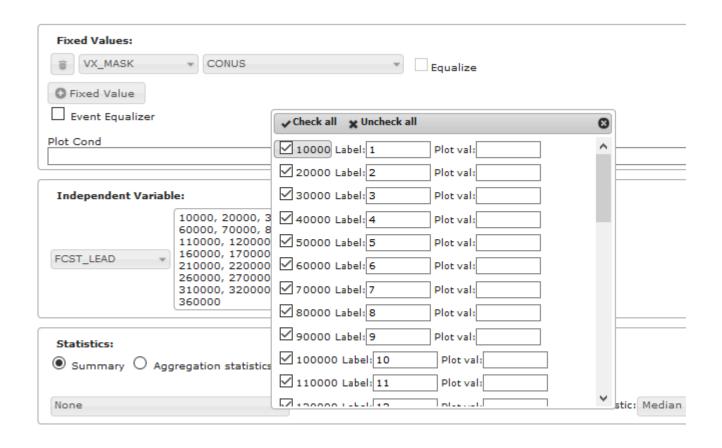




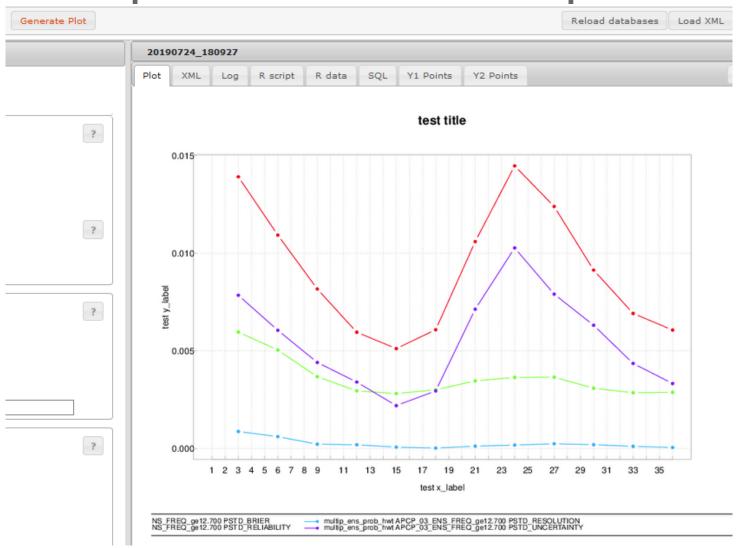




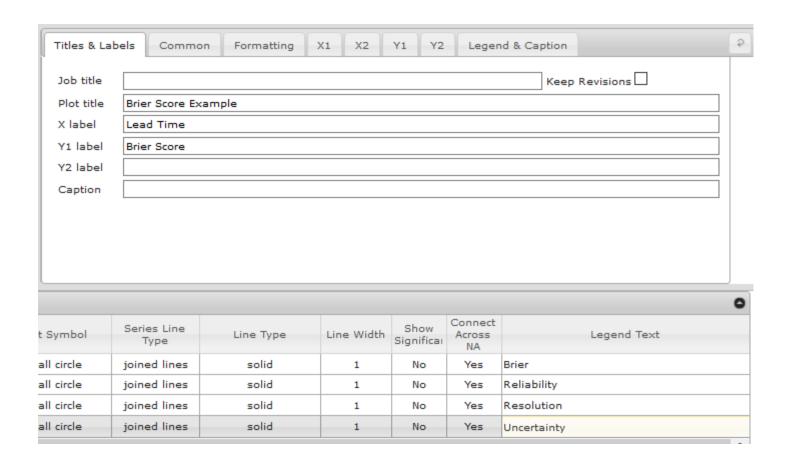




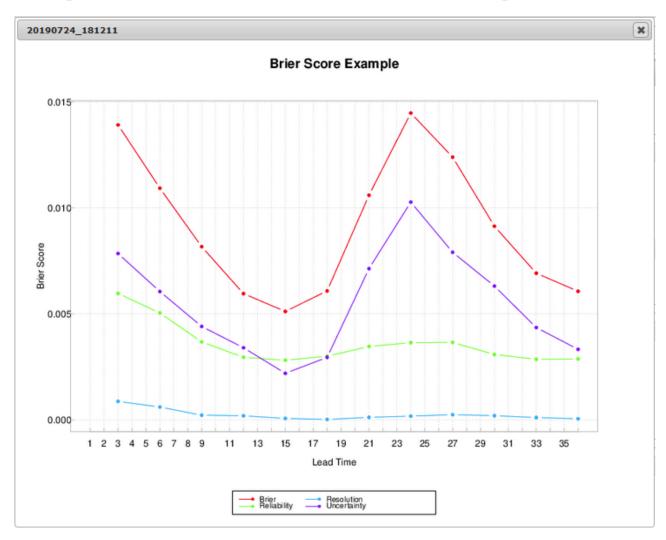








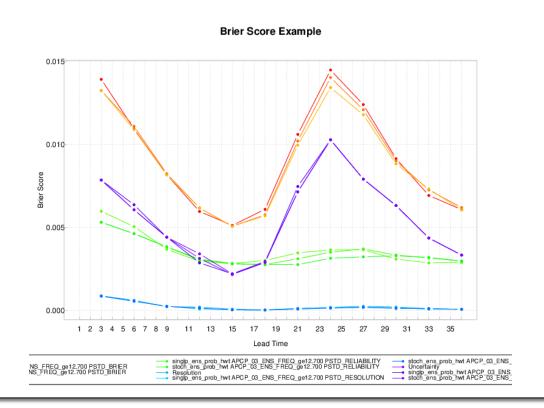






us? Can we camplate selveral models?

Add the two other models



East vs. West?

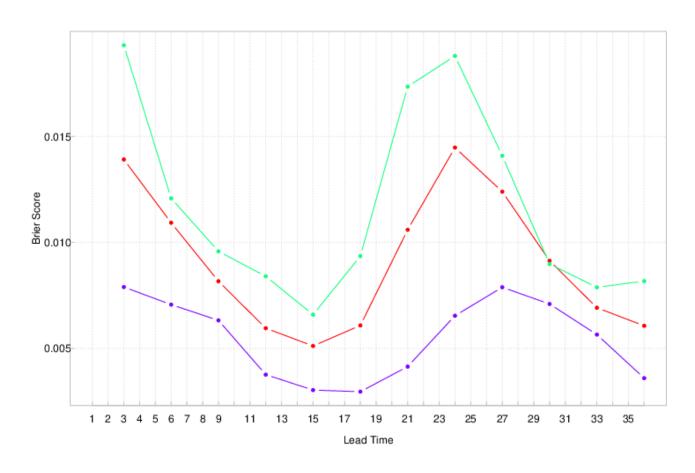
Select only Brier Score Remove two models

Add Y1 Series Variable and select

VX Mask -> CON

Remove Fixed Value entry

Brier Score Example

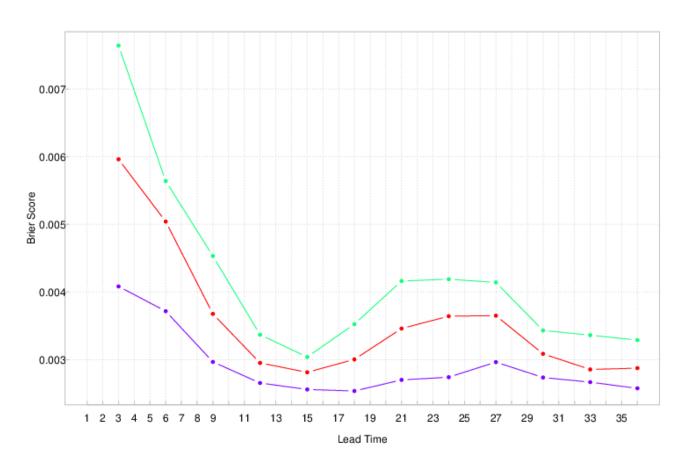


Explore a little

Do scores change if stratified East vs. West?

Select only Reliabil

Brier Score Example

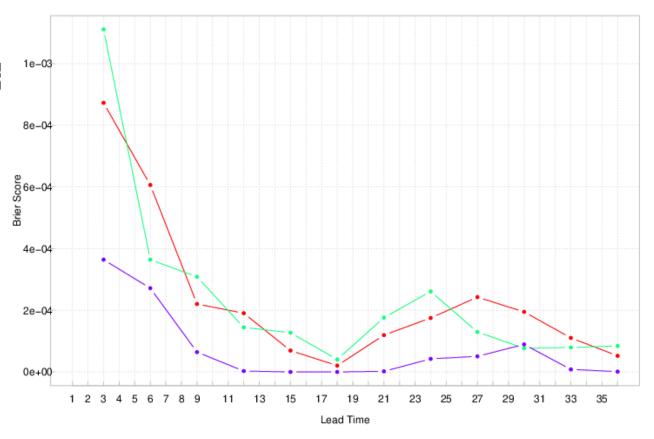


Explore a little

Brier Score Example

Do score East vs.

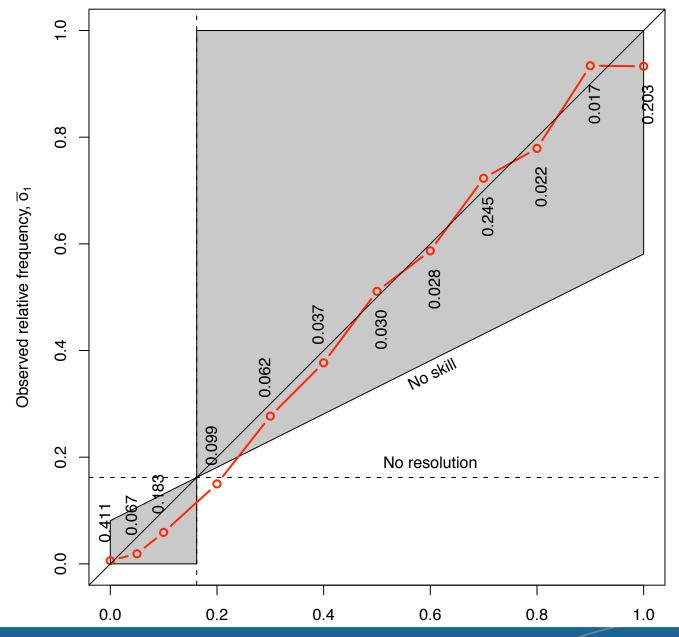
Select only Resoluti



Reliability (Attribute) diagram

- Analogous to the scatter plot- same intuition holds.
- Data must be binned!
- Hides how much data is represented by each
- Expresses conditional probabilities.
- Confidence intervals can illustrate the problems with small sample sizes.

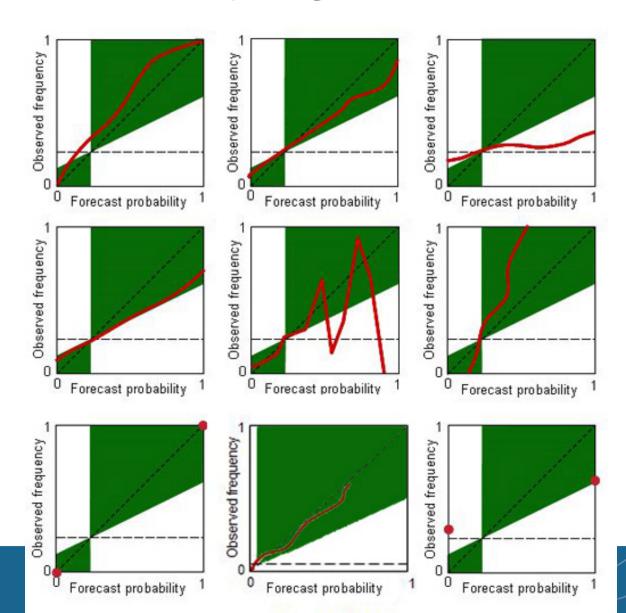




Reliability
diagram plots
observed
frequency of
event vs
probability
forecasted for
event;

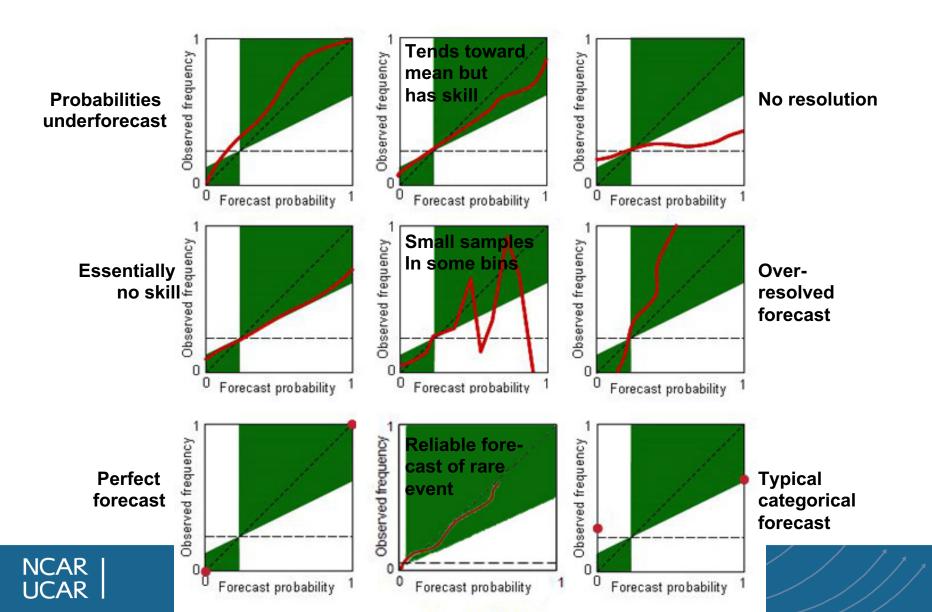
Attribute
diagram adds
lines to show
how connected
line (reliability)
relates to
Resolution and
skill

Reliability Diagram Exercise

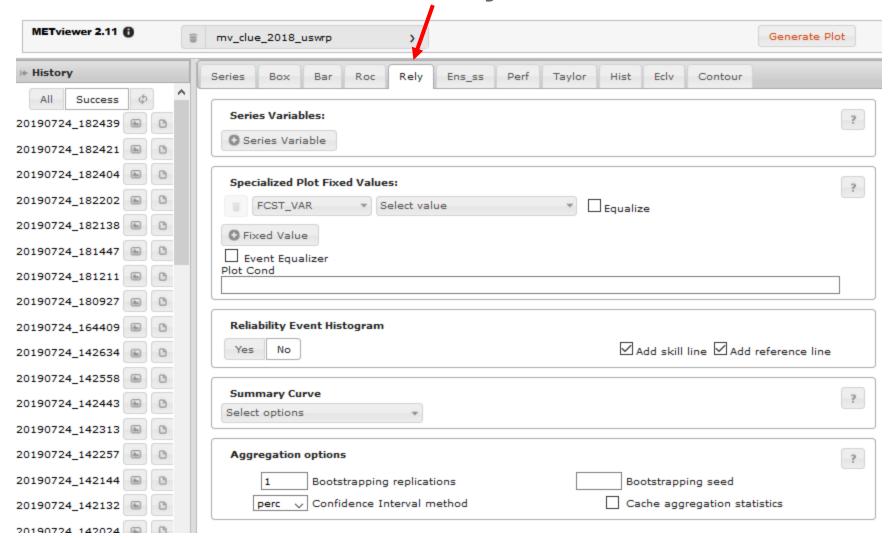




Reliability Diagram Exercise

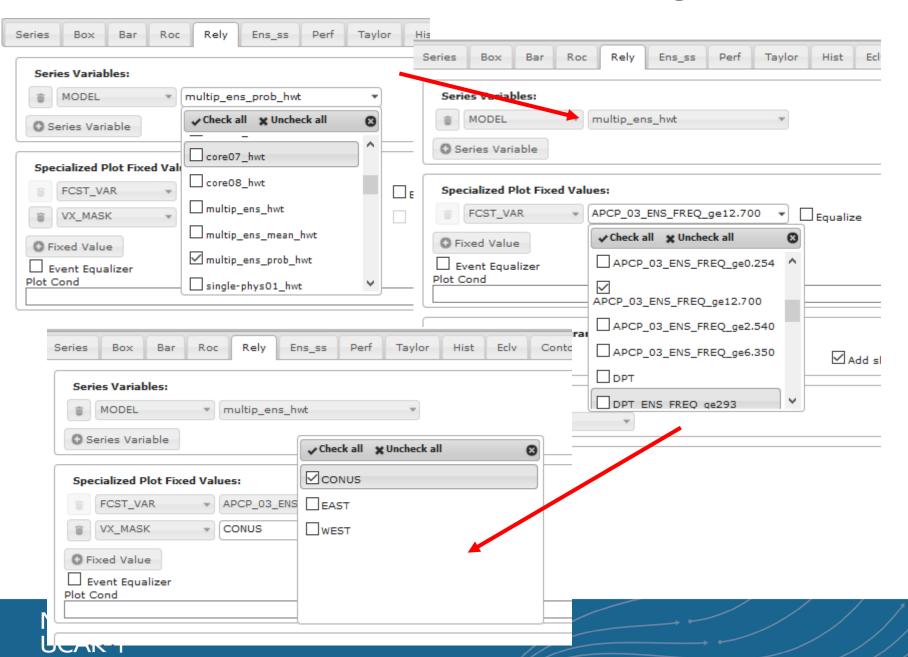


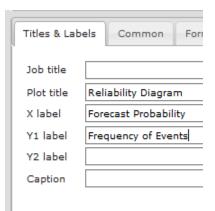
Select Rely Tab





Select Model, Fcst Var and Region

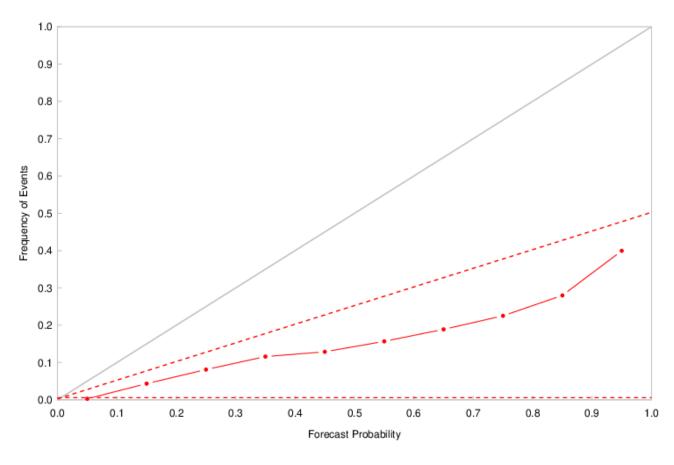




Generate Plot

multip_ens_prob_hwt Reliability Curve

Reliability Diagram





Does the reliability change with different ensemble compositions (multip, singlep, stocachastic)?

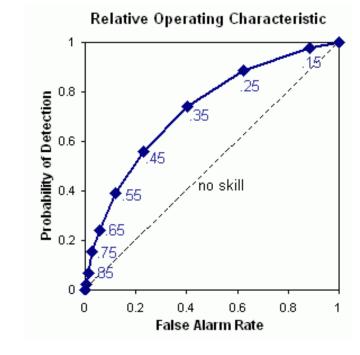
Add the other two modes_ens_prob_hwt to the MODEL list Does the reliability change with different preparations.

Change to a different thresholds to the FCST_VAR drop-down with different regions?

Move VX_MASTK to Series Variables and Add East, West

Interpretation of ROC

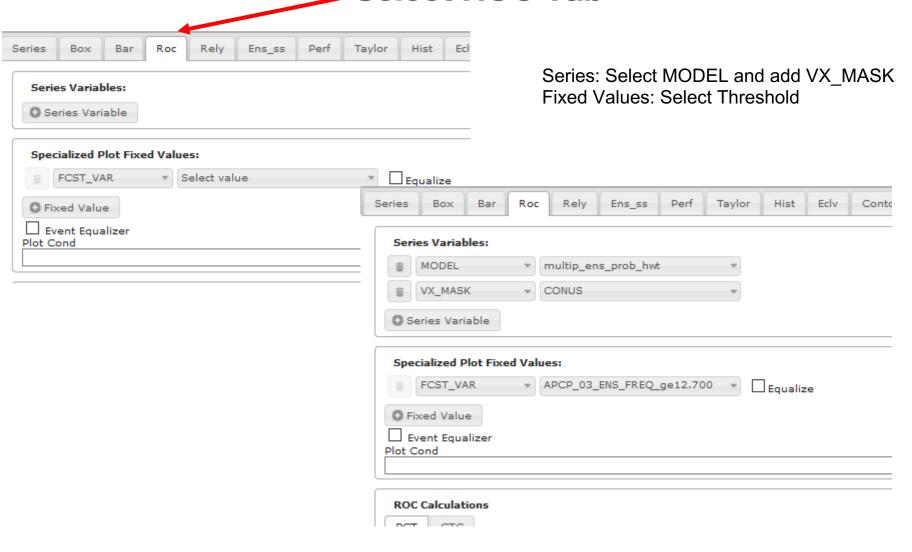
- Close to upper left corner good resolution
- Close to diagonal little skill
- Area under curve ("ROC area") is a useful summary measure of forecast skill
- Perfect: ROC area = 1
- No skill: ROC area = 0.5
- ROC skill score ROCS = 2(ROCarea-0.5)
- Not sensitive to bias.



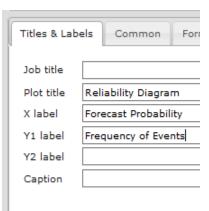
- ROC is conditioned on the observations (i.e., given that Y occurred, what was the corresponding forecast?)
- Reliability and ROC diagrams are good companions



Select ROC Tab

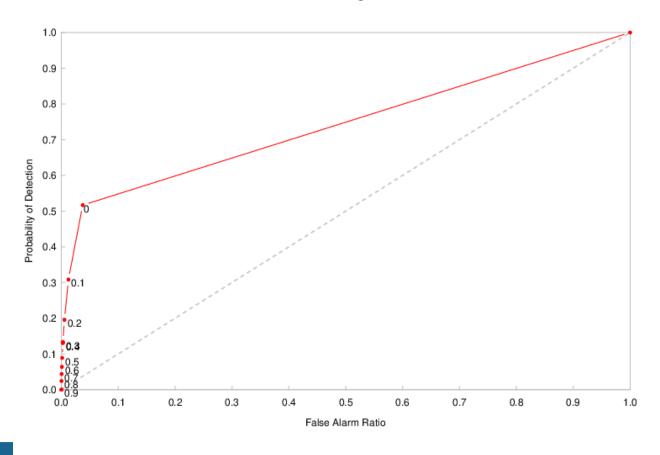






Generate Plot

ROC Diagram





Explore a little

Does the ROC change with different ensemble compositions (multip, singlep, stocachastic)?

Add the other two modes_ens_prob_hwt to the MODEL list

Does the ROC change with

different thresholds?

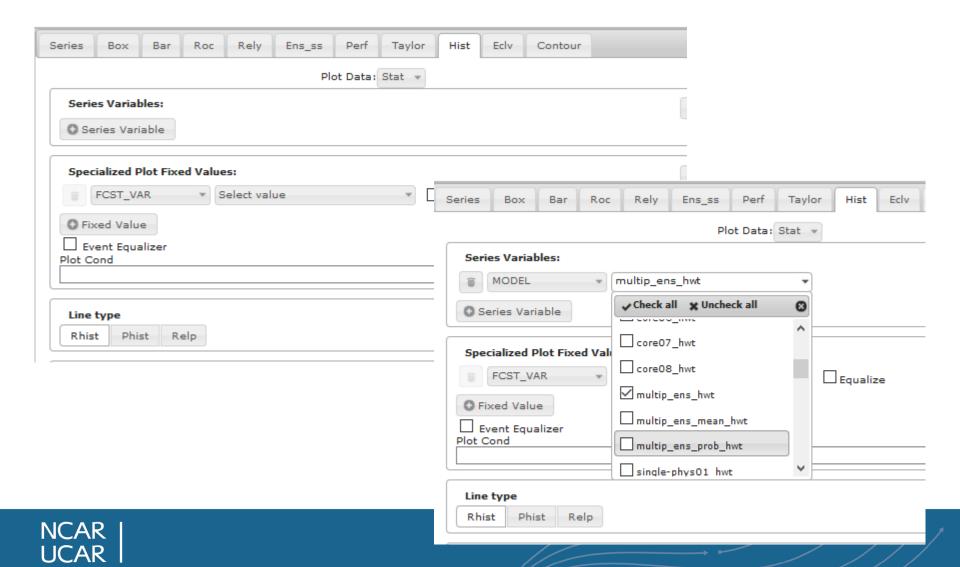
Change to a different thresholds to the FCST_VAR drop-down

Does ROC change with different regions?

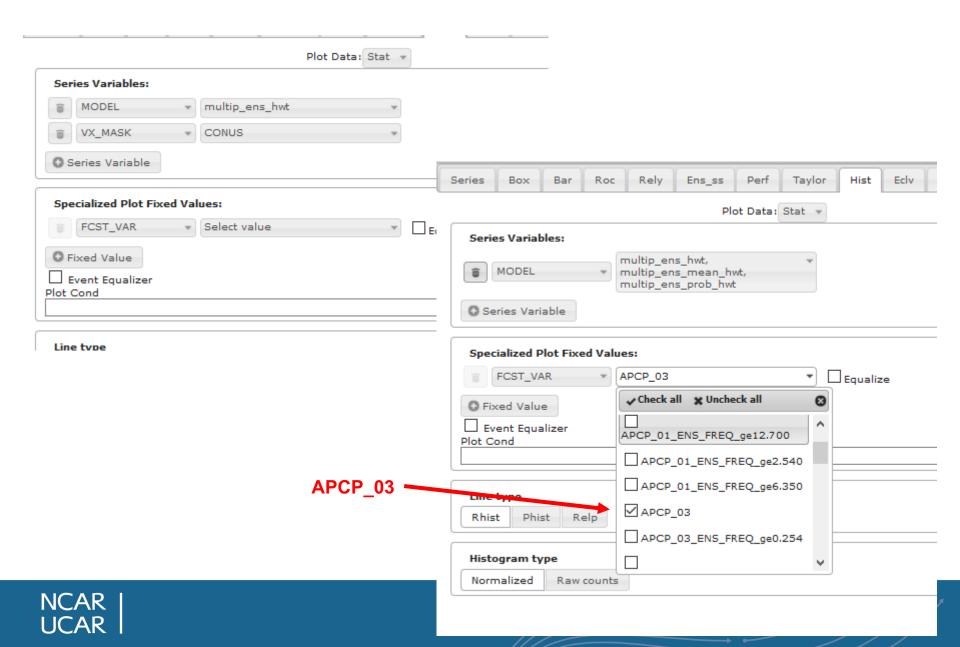
Move VX_MASTK to Series Variables and Add East, West

What else?

Rank Histograms

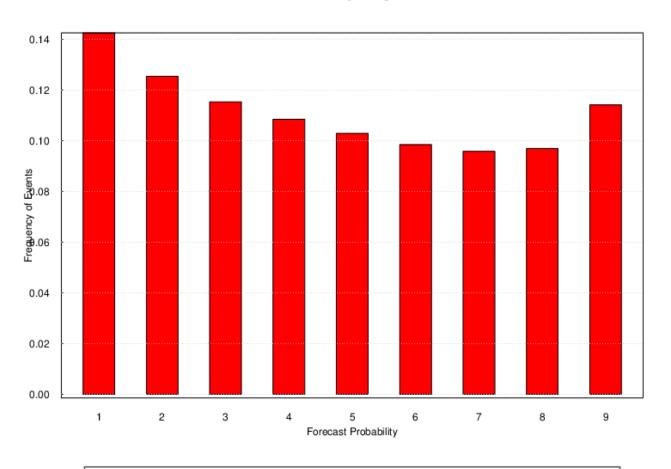


Add VX_MASK to Series; Select FCST_VAR



Plot Rank Histogram

Reliability Diagram



multip_ens_hwt Rank Histogram



Explore a little

Does the rank histogram change with different ensemble compositions (multip, singlep, stocachastic)?

Add the other two modes_ens_prob_hwt to the MODEL list Does the rank histogram change with different thresholds?

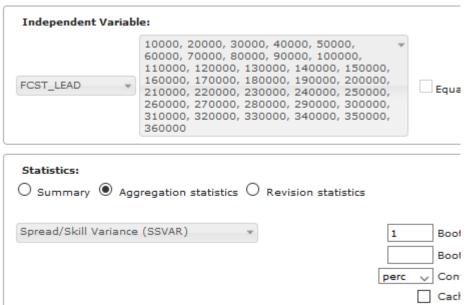
Change to a different thresholds to the FCST_VAR drop-down
Does rank histogram change with different regions?

Move VX MASK to Series Variables and Add East, West

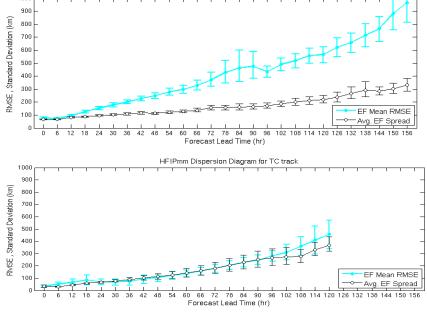
What else?

Evaluating ensembles

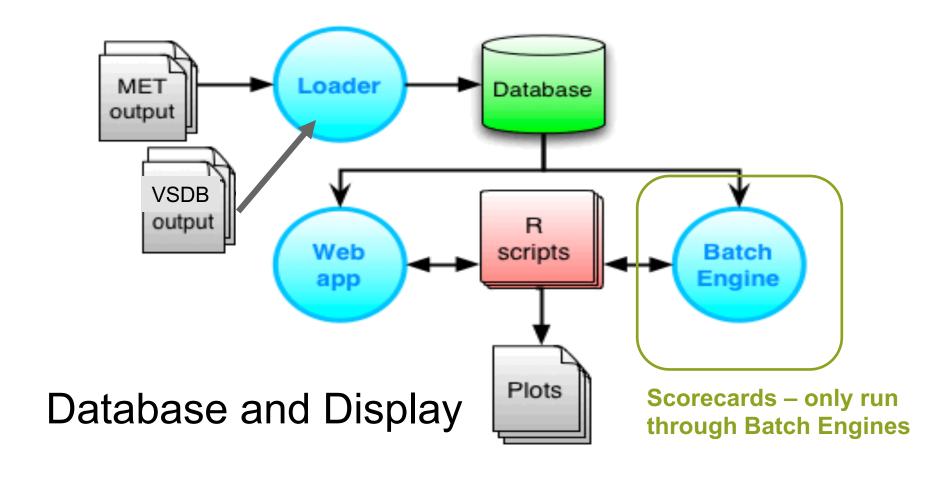
- Select Series Tab
- Y1 Dependent is APCP_03 -> SSVAR_RMSE, SSVAR Spread
- Add Y1 Dependent VX_MASK -> CONUS
- Y1 Series Var MODEL -> multip ens hwt
- FCST LEAD -> Select all leads
- Statistics
- Aggregation (see below)



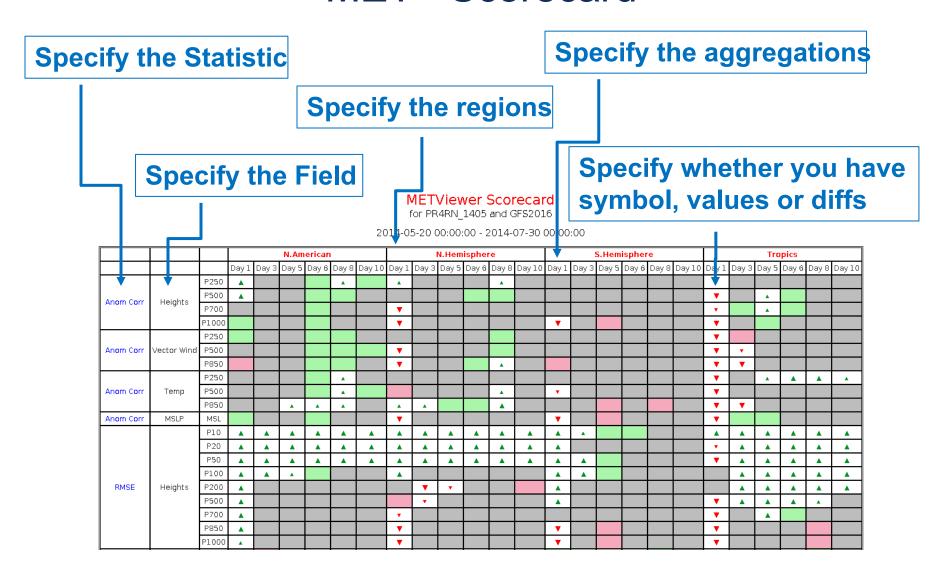
Spread-skill



Scorecards



MET+ Scorecard





METViewer CAM Scorecard for GFDLFV3 and NSSLFV3

2018-04-30 00:00:00 - 2018-05-22 00:00:00

METViewer CAM Scorecard for NSSLFV3 and HRRR 2018-04-30 00:00:00 - 2018-05-22 00:00:00

Daily & CONUS 2018-04-30 00:00:00 - 2018-05-22 00:00:00

METViewer CAM Scorecard for HREFv2 and HRRRE

Daily Domain CONUS Daily

001100
Domains
Domains

		Daily Domain	CONUS					Daily Domain	CONUS	
		Daily	′					Daily		1
	>=0.02						>=0.02		•	
	>=0.05			9			>=0.05		▼	
	>=0.10		•				>=0.10	*	▼	
NBR 50	>=0.15		•) ^ (NBR 50	>=0.15	*	▼	
	>=0.30			Se			>=0.30		•]
	>=0.45			Φ)			>=0.45	*	•]
	>=0.60			atı			>=0.60			
	>=0.02		•) G			>=0.02		•	
	>=0.05		•	rrc			>=0.05		•	
	>=0.10		•	Surrogate Severe		NBR 75	>=0.10		•	
NBR 75	>=0.15						>=0.15	*	▼	
	>=0.30						>=0.30	*	▼	
	>=0.45						>=0.45	+	•	
	>=0.60				CSI		>=0.60			
	>=0.02		•		3		>=0.02		▼	
	>=0.05		•	Prob			>=0.05		▼	
	>=0.10		•				>=0.10		▼	
NBR 100	>=0.15		•			NBR 100	>=0.15		▼	
	>=0.30		*				>=0.30	+	▼	
	>=0.45			UH -			>=0.45		*	
	>=0.60						>=0.60			
	>=0.02		*	exceeding			>=0.02		▼	
	>=0.05						>=0.05		▼	
	>=0.10						>=0.10		▼	
NBR 125	>=0.15					NBR 125	>=0.15		▼	
	>=0.30		*				>=0.30			
	>=0.45						>=0.45			
	>=0.60						>=0.60	<u> </u>		

		>=0.02		
	NBR 50	>=0.05		
		>=0.10		
		>=0.15		
		>=0.30		
		>=0.45		
		>=0.60		
		>=0.02		
		>=0.05		
		>=0.10		
	NBR 75	>=0.15		
		>=0.30		
		>=0.45		
CSI		>=0.60	*	4
031		>=0.02		
	NBR 100	>=0.05		
		>=0.10		
		>=0.15		4
		>=0.30		*
		>=0.45		
		>=0.60		
	NBR 125	>=0.02		*
		>=0.05		
		>=0.10		
		>=0.15		
		>=0.30		
		>=0.45		
		>=0.60	A	*

- ▲ GFDLFV3 is better than NSSLFV3 at the 99.9% significance level
- GFDLFV3 is better than NSSLFV3 at the 99% significance level

GFDLFV3 is better than NSSLFV3 at the 95% significance level

No statistically significant difference between GFDLFV3 and NSSLFV3

- GFDLFV3 is worse than NSSLFV3 at the 95% significance level GFDLFV3 is worse than NSSLFV3 at the 99% significance level
- GFDLFV3 is worse than NSSLFV3 at the 99.9% significance level

Not statistically relevant

- NSSLFV3 is better than HRRR at the 99.9% significance level
- NSSLFV3 is better than HRRR at the 99% significance level

NSSLFV3 is better than HRRR at the 95% significance level

No statistically significant difference between NSSLFV3 and HRRR NSSLFV3 is worse than HRRR at the 95% significance level

NSSLFV3 is worse than HRRR at the 99% significance level

Not statistically relevant

NSSLFV3 is worse than HRRR at the 99.9% significance level

- ▲ HREFv2 is better than HRRRE at the 99.9% significance level
- HREFv2 is better than HRRRE at the 99% significance level

HREFv2 is better than HRRRE at the 95% significance level

No statistically significant difference between HREFv2 and HRRRE

HREFv2 is worse than HRRRE at the 95% significance level

- HREFv2 is worse than HRRRE at the 99% significance level
- ▼ HREFv2 is worse than HRRRE at the 99.9% significance level
- Not statistically relevant

METViewer CAM Scorecard

for HREFv2 and HRRRE

2018-04-30 00:00:00 - 2018-06-01 00:00:00

To eliminate biases
Percentile Thresholding
applied

Specific Threshold

Threshold used is associated with Percentiles (75, 80, 85, 90, 95) computed from climatology

			Dally Domain	CONUS
			Daily	
		>=0.02		•
		>=0.05		
		>=0.10		
	75%	>=0.15		
		>=0.30		
		>=0.45		
		>=0.60		
		>=0.02		
		>=0.05		
		>=0.10		
	80%	>=0.15		
		>=0.30		
		>=0.45		
		>=0.60		
		>=0.02		
		>=0.05		
	85%	0.10		
CSI		>=0.15		
		>=0.30	*	
		>=0.45	*	*
		>=0.60		
		>=0.02		
		>=0.05	*	
		>=0.10	A	
	90%	>=0.15		
		>=0.30	*	*
		>=0.45	*	*
		>=0.60		
		>=0.02	A	
		>=0.02 >=0.05	A	A
		>=0.02 >=0.05 >=0.10		*
	95%	>=0.02 >=0.05 >=0.10 >=0.15	A A	A
	95%	>=0.02 >=0.05 >=0.10 >=0.15 >=0.30		*
	95%	>=0.02 >=0.05 >=0.10 >=0.15	A A	A

■ HREFv2 is better than HRRRE at the 99% significance level
HREFv2 is better than HRRRE at the 95% significance level
No statistically significant difference between HREFv2 and HRRRE
HREFv2 is worse than HRRRE at the 95% significance level

HREFv2 is worse than HRRRE at the 99% significance level
Not statistically relevant



2018-04-30 00:00:00 - 2018-05-22 00:00:00

			Daily Domain	CONUS
			Dail	y
	NBR 50	>=0.02		
		>=0.05		
		>=0.10		
		>=0.15		
		>=0.30		
		>=0.45		
		>=0.60		
		>=0.02		
		>=0.05		
	NBR 75	>=0.10		
		>=0.15		
		>=0.30		
		>=0.45		
CSI		>=0.60	4	*
5	NBR 100	>=0.02		
		>=0.05		
		>=0.10		
		>=0.15		
		>=0.30		
		>=0.45		
		>=0.60		
	NBR 125	>=0.02		*
		>=0.05		
		>=0.10		
		>=0.15		
		>=0.30		
		>=0.45		
		>=0.60	A	A

- HREFv2 is better than HRRRE at the 99.9% significance level
- HREFv2 is better than HRRRE at the 99% significance level

HREFv2 is better than HRRRE at the 95% significance level

No statistically significant difference between HREFv2 and HRRRE

- HREFv2 is worse than HRRRE at the 95% significance level
- ▼ HREFv2 is worse than HRRRE at the 99% significance level
- ▼ HREFv2 is worse than HRRRE at the 99.9% significance level

Not statistically relevant

