Model Evaluation Tools

Winter 2011 Tutorial

Weather Research & Forecasting

Basic Verification Concepts

Barbara Brown
National Center for Atmospheric Research
Boulder Colorado USA

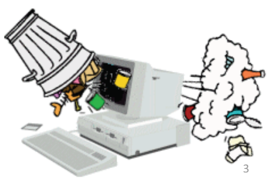
DTC Developmental Testbed Center

Basic concepts - outline

- What is verification?
- Why verify?
- Identifying verification goals
- Forecast "goodness"
- Designing a verification study
- Types of forecasts and observations
- Matching forecasts and observations
- Verification attributes
- Miscellaneous issues
- **Questions to ponder: Who? What? When? Where? Which? Why?**

How do you do verification?

- Using MET is the easy part, scientifically speaking.
- Good verification depends mostly on what you do before and after MET.
  - What do you want to know?
  - Good forecasts.
  - Good observations.
  - Well matched.
  - Thorough and correct interpretation of results.



What is verification?

- Verification is the process of comparing forecasts to relevant observations
  - Verification is one aspect of measuring forecast ***goodness***
- Verification measures the ***quality*** of forecasts (as opposed to their ***value***)
- For many purposes a more appropriate term is "***evaluation***"

Why verify?

- Purposes of verification (traditional definition)

  - Administrative purpose
    - Monitoring performance
    - Choice of model or model configuration (has the model improved?)

  - Scientific purpose
    - Identifying and correcting model flaws
    - Forecast improvement

  - Economic purpose
    - Improved decision making
    - "Feeding" decision models or decision support systems

Why verify?

- What are some other reasons to verify weather forecasts?
  - Help operational forecasters understand model biases and select models for use in different conditions
  - Help "users" interpret forecasts (e.g., "What does a temperature forecast of 0 degrees really mean?")
  - Identify forecast weaknesses, strengths, differences

Identifying verification goals

What *questions* do we want to answer?

- Examples:
  - ✓ In what locations does the model have the best performance?
  - ✓ Are there regimes in which the forecasts are better or worse?
  - ✓ Is the probability forecast well calibrated (i.e., reliable)?
  - ✓ Do the forecasts correctly capture the natural variability of the weather?

Other examples?

Identifying verification goals (cont.)

- What forecast performance ***attribute*** should be measured?
  - Related to the *question* as well as the type of forecast and observation
- Choices of verification statistics, measures, graphics
  - Should match the type of forecast and the attribute of interest
  - Should measure the quantity of interest (i.e., the quantity represented in the question)

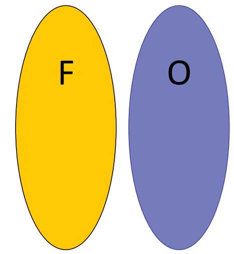
Forecast “goodness”

- Depends on the quality of the forecast

AND

- The user and his/her application of the forecast information

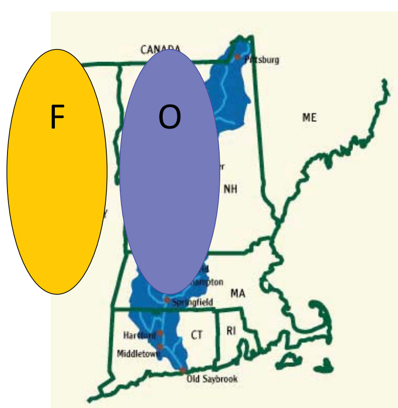
Good forecast or bad forecast?



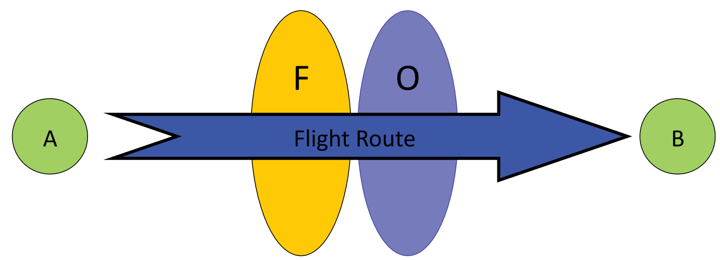
Many verification approaches would say that this forecast has NO skill and is very inaccurate.

Good forecast or Bad forecast?

If I’m a water manager for this watershed, it’s a pretty bad forecast…



Good forecast or Bad forecast?



If I’m an aviation traffic strategic planner…

It might be a pretty good forecast

Different users have different ideas about what makes a forecast good

Different verification approaches can measure different types of “goodness”

Forecast "goodness"

- Forecast quality is only one aspect of forecast "goodness"
- Forecast value is related to forecast quality through complex, non-linear relationships

  In some cases, *improvements in forecast quality (according to certain measures) may result in a degradation in forecast value for some users!*
- ***However*** - Some approaches to measuring forecast quality can help understand goodness

  Examples
  - ✓ Diagnostic verification approaches
  - ✓ New features based approaches
  - ✓ Use of multiple measures to represent more than one attribute of forecast performance
  - ✓ Examination of multiple thresholds

Basic guide for developing verification studies

Consider the users…
- – … of the forecasts
- – … of the verification information
- What aspects of forecast quality are of interest for the user?
  - – Typically (always?) need to consider multiple aspects

Develop verification questions to evaluate those aspects/ attributes
- *Exercise*: What verification questions and attributes would be of interest to …
  - – … operators of an electric utility?
  - – … a city emergency manager?
  - – … a mesoscale model developer?
  - – … aviation planners?

Basic guide for developing verification studies

Identify *observations* that represent the ***event*** being forecast, including the

Element (e.g., temperature, precipitation)

Temporal resolution

Spatial resolution and representation

Thresholds, categories, etc.

Observations are not truth

- We can't know the complete "truth".
- Observations generally are more "true" than a model analysis (at least they are relatively more independent)
- Observational uncertainty should be taken into account in whatever way possible
  - ✓ In other words, how well do adjacent observations match each other?

Observations might be garbage if

- Not Independent (of forecast or each other)
- Biased
  - Space
  - Time
  - Sampling bias
  - Reporting bias
- Not enough of them

Basic guide for developing verification studies

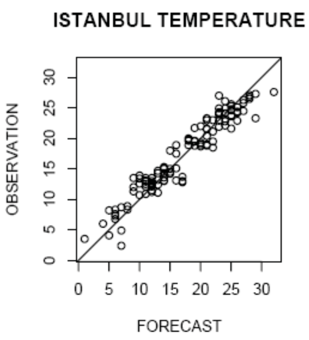
Identify multiple *verification attributes* that can provide answers to the questions of interest

Select *measures and graphics* that appropriately measure and represent the attributes of interest

Identify a *standard of comparison* that provides a reference level of skill (e.g., persistence, climatology, old model)

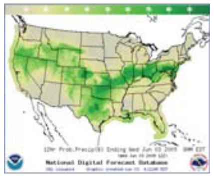
Types of forecasts, observations

- Continuous
  - Temperature
  - Rainfall amount
  - 500 mb height
- Categorical
  - Dichotomous
    - ✓ Rain vs. no rain
    - ✓ Strong winds vs. no strong wind
    - ✓ Night frost vs. no frost
    - ✓ Often formulated as Yes/No
  - Multi category
    - ✓ Cloud amount category
    - ✓ Precipitation type
  - May result from *subsetting* continuous variables into categories
    - ✓ *Ex: Temperature categories of 0-10, 11-20, 21-30, etc.*



Types of forecasts, observations

- Probabilistic
  - Observation can be dichotomous, multi-category, or continuous
    - Precipitation occurrence – Dichotomous (Yes/No)
    - Precipitation type – Multi-category
    - Temperature distribution - Continuous
  - Forecast can be
    - Single probability value (for dichotomous events)
    - Multiple probabilities (discrete probability distribution for multiple categories)
    - Continuous distribution
  - For dichotomous or multiple categories, probability values may be limited to certain values (e.g., multiples of 0.1)



2-category precipitation forecast (PoP) for US

- Ensemble
  - Multiple iterations of a continuous or categorical forecast
    - May be transformed into a probability
  - Observations may be continuous, dichotomous or multi-category

ECMWF 2-m temperature meteogram for Helsinki

Matching forecasts and observations

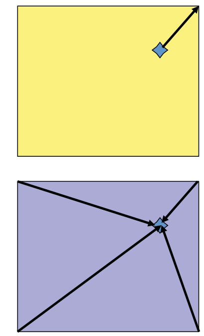
- May be the *most difficult* part of the verification process!
- Many factors need to be taken into account
  - Identifying observations that represent the forecast event
    - ✓ Example: Precipitation accumulation over an hour at a point
  - For a gridded forecast there are many options for the matching process
    - Point to grid
      - Match obs to closest gridpoint
    - Grid to point
      - Interpolate?
      - Take largest value?

Matching forecasts and observations

- Point-to-Grid and

  Grid-to-Point

- Matching approach can impact the results of the verification



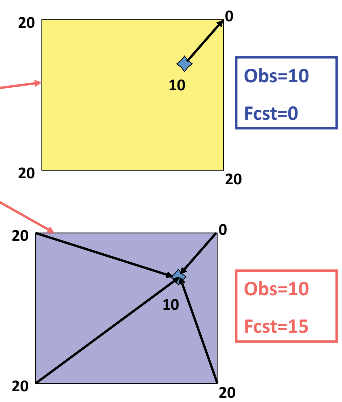
Matching forecasts and observations

Example:

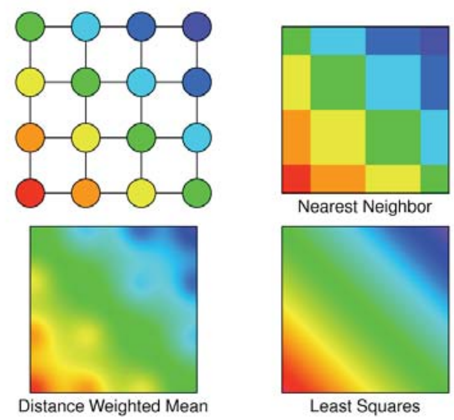
- Two approaches:
  - Match rain gauge to nearest gridpoint ***or***
  - Interpolate grid values to rain gauge location
    - Crude assumption: equal weight to each gridpoint
- Differences in results associated with matching:

"Representativeness" difference

Will impact most verification scores



Interpolation Examples



© 2010 University Corporation for Atmospheric Research. All Rights Reserved.

Matching forecasts and observations

<u>Final point:</u>

- It is not advisable to use the model analysis as the verification "observation".
- Why not??
- Issue: Non-independence!!

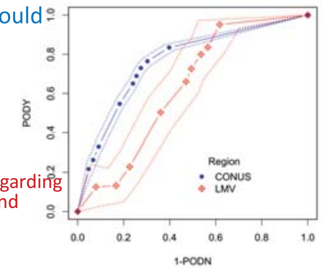
Comparison and inference

Uncertainty in scores and measures should be estimated whenever possible!

Uncertainty arises from
- Sampling variability
- Observation error
- Representativeness differences
- Others?

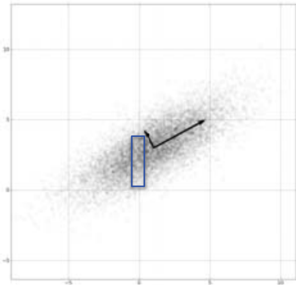
Erroneous conclusions can be drawn regarding improvements in forecasting systems and models

Methods for *confidence intervals* and *hypothesis tests*
- Parametric (i.e., depending on a statistical model)
- Non parametric (e.g., derived from re sampling procedures, often called "bootstrapping")



Verification attributes

- Verification attributes measure different aspects of forecast quality
  - Represent a range of characteristics that should be considered
  - Many can be related to joint, conditional, and marginal distributions of forecasts and observations



Joint : The probability of two events in conjunction.
P (Yellow & Hit) = 0.09

Conditional : The probability of one variable given that the second is already determined.
P (Hit | Yellow) = 0.09 / (0.09+0.01)

Marginal : The probability of one variable without regard to the other.

	Red	Yellow	Green	Marginal Probability
Hit	**0.693**	**0.09**	**0.04**	**0.823**
Not Hit	**0.007**	**0.01**	**0.16**	**0.177**
Total				1

Verification attribute examples

- Bias
  (Marginal distributions)
- Correlation
  Overall association (Joint distribution)
- Accuracy
  Differences (Joint distribution)
- Calibration
  Measures conditional bias (Conditional distributions)
- Discrimination
  Degree to which forecasts discriminate between different observations (Conditional distribution)

Miscellaneous issues

- In order to be *verified*, forecasts must be formulated so that they are *verifiable*!
  - Corollary: All forecasts should be verified if something is worth forecasting, it is worth verifying
- Stratification and aggregation
  - Aggregation can help increase sample sizes and statistical robustness but can also hide important aspects of performance
    - ✓ Most common regime may dominate results, mask variations in performance.
  - Thus it is very important to *stratify results into meaningful, homogeneous sub-groups*

Some key things to think about …

Who…
- …wants to know?

What…
- … does the user care about?
- … kind of parameter are we evaluating? What are its characteristics (e.g., continuous, probabilistic)?
- … thresholds are important (if any)?
- … forecast resolution is relevant (e.g., site specific, area average)?
- … are the characteristics of the obs (e.g., quality, uncertainty)?
- … are appropriate methods?

Why…
- …do we need to verify it?

Some key things to think about…

How…
- …do you need/want to present results (e.g., stratification/aggregation)?

Which…
- …methods and metrics are appropriate?
- … methods are required (e.g., bias, event frequency, sample size)

Categorical verification

Eric Gilleland

Slides by Matt Pocernich via ??

Finley Tornado Dat (1884)

		Observed		
Forecast		Yes	No	Total
	Yes	28	72	100
	No	23	2680	2703
	Total	51	2752	2803

A success?

		Observed		
Forecast		Yes	No	Total
	Yes	28	72	100
	No	23	2680	2703
	Total	51	2752	2803

•Percent Correct (28+2680)/2803 96.6% !!!!

Maybe not.

		Observed		
Forecast		Yes	No	Total
	Yes	0	0	0
	No	51	2752	2803
	Total	51	2752	2803

•Percent Correct (0+2752)/2803 98.2%

2 x 2 Contingency Table

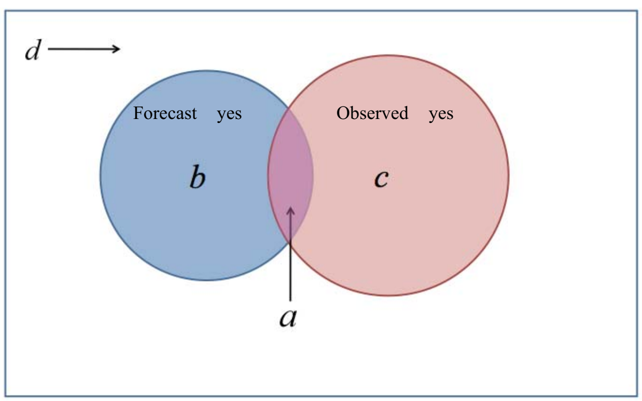
		Observed		
		Yes	No	Total
Forecast	Yes	Hit	False Alarm	Forecast Yes
	No	Miss	Correct Negative	Forecast No
	Total	Obs. Yes	Obs. No	Total

Common – though not universal - notation

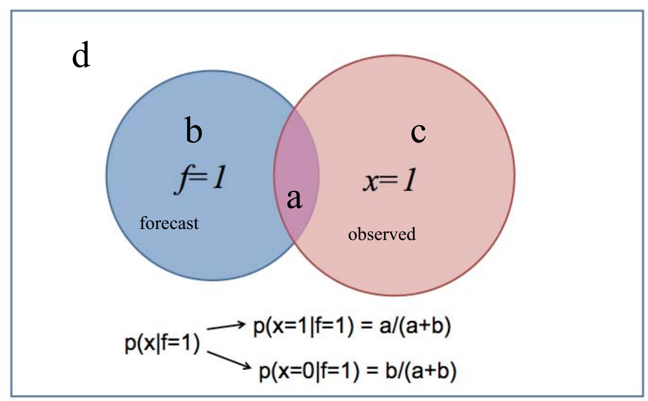
		Observed		
		Yes	No	Total
Forecast	Yes	a	b	a+b
	No	c	d	c+d
	Total	a+c	b+d	n

Base Rate (aka sample climatology) = (a+c)/n

Alternative Perspective on Contingency Table



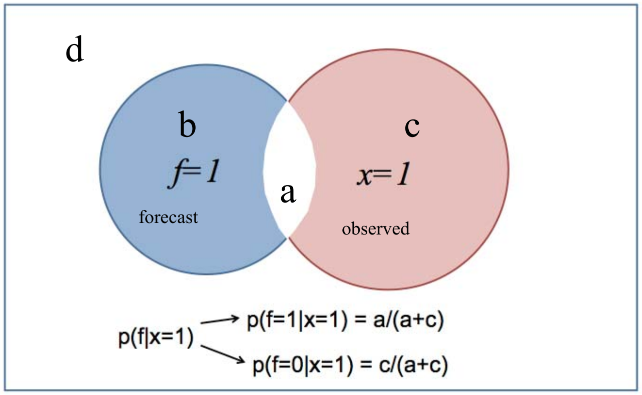
Conditioning on Forecast



“good” forecasts

- p(x=1|f=1) = a/(a+b) to be as large as possible
  - fraction of “hits” in forecast region
- p(x=0|f=1) = b/(a+b) to be a small as possible
  - fraction of “false alarms” in forecast region

Conditioning on outcome

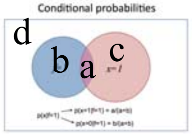


“good” forecasts

- p(f=1|x=1) to be as large as possible
  - fraction of “hits” in observed region
- p(f=0|x=1) to be a small as possible
  - fraction of “missed” in observed region
- provides information regarding the ability of the forecast system to “discriminate” observed events vs. non-events
- these conditional probabilities are also called the “likelihoods” associated with the forecast

A laundry list of scores – most based on conditioning

- Hit Rate = a/(a+c)
- False Alarm Rate (POFD) = b/(b+d)
- False Alarm **Ratio** (FAR) = b/(a+b)
- (frequency) Bias = (a+b)/(a+c)
- Threat Score or Critical Success Index = a/(a+b+c)
- PODn = d/(b+d) = (1 POFD)



		Observed		
		Yes	No	Total
Forecast	Yes	a	b	a+b
	No	c	d	c+d
	Total	a+c	b+d	n

Alternative Statistics

		Observed		
Forecast		Yes	No	Total
	Yes	28	72	100
	No	23	2680	2703
	Total	51	2752	2803

Threat Score 28 / (28 + 72+ 23) 0.228
Probability of Detection 28 / (28 + 23) 0.55
False Alarm Ratio 72/(28 + 72) 0.720

Uncertainty 

A normal approximation confidence interval (CI) can be used to obtain a $(1-\alpha)100\%$ CI around these estimates. For $\hat{s}$ an estimate of a statistic, θ, it is given by

$$\hat{s} \pm z[\alpha/2]\mathrm{se}(\theta)$$

using $\mathrm{sqrt}(\hat{s}(1-\hat{s})/n)$ as an estimate for $\mathrm{se}(\theta)$ (for proportions such as hit rate and false alarm ratio). A better approximation exists, but is messier for slide purposes.

95% normal approximation CI shown in red

Uncertainty

		Observed		
Forecast		Yes	No	Total
	Yes	28	72	100
	No	23	2680	2703
	Total	51	2752	2803

Hit rate 0.55 ≈ (0.41, 0.69)
FAR 0.72 ≈ (0.63, 0.81)
False Alarm Rate 0.03 ≈ (0.02, 0.03)

Skill Scores

(How do you compare the skill of easy to predict events with difficult to predict events?)

- Single value to summarize performance.
- Reference forecast - best naive guess; persistence, climatology.
- Proper skill scores reflect forecaster true intent.
- A perfect forecast implies that the object can be perfectly observed.
- Reference forecast must be comparable.

Generic Skill Score

$$SS = \frac{(A - A_{ref})}{(A_{perf} - A_{ref})} -$$

$$MSESS = 1 - \frac{MSE}{MSE_{climo}}$$

- Positively oriented – Positive is good

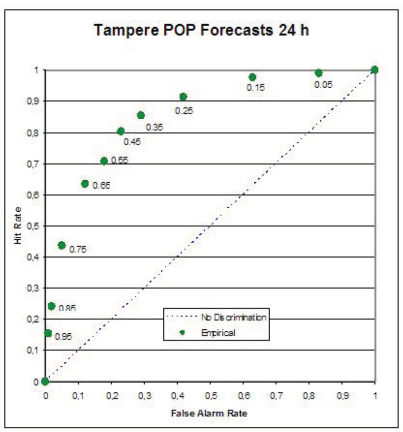
Calculation of Empirical ROC

Does not need to be a probability!

Does not need to be calibrated!

PROB	# YES	# NO
0.05	6	32
0.15	7	8
0.25	2	8
0.35	7	9
0.45	4	9
0.55	15	5
0.65	10	10
0.75	12	3
0.85	16	8
0.95	146	14

Empirical ROC



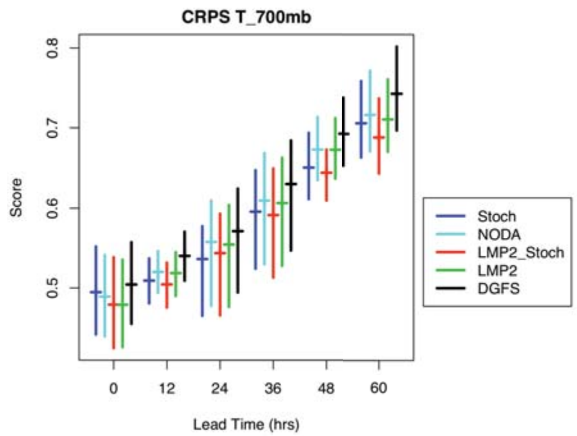
Area under the curve is a useful measure;

Perfect = 1, random = 0.5

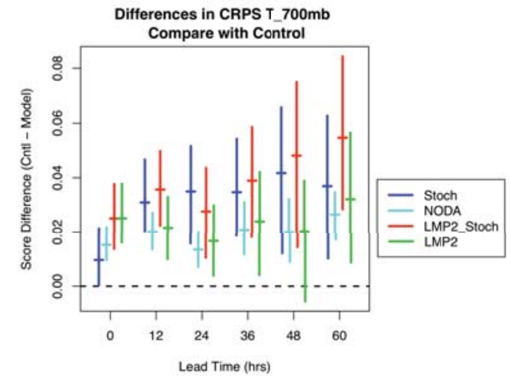
Sampling variability can be used

- When you are comparing two forecasts of the same event, evaluate the differences.
- Sampling variability is large and can quickly overwhelm small, but significant differences.

Scores in absolute terms



Examining the distribution of differences



Bootstrap CI's

Assumptions: *Sample* distribution represents *population* distribution.

1. Sample with replacement from data;
2. Calculate statistic(s) of interest;
3. Repeat steps 1 and 2 several times to obtain a sample of the statistic(s);
4. Calculate CI's based on the sample obtained in step 3.

Bootstrap CI's

Assumptions: *Sample* distribution represents *population* distribution.

1. Sample with replacement from data;

- Certain values will appear more than once (possibly several times);
- Others will not appear at all;
- Referred to as replicate samples;
- Replicate sample size usually taken to be same size as original sample, but there are cases where it is preferable to use a smaller sample size (usually on the order of the square root of the original size).

Bootstrap CI's

Assumptions: *Sample* distribution represents *population* distribution.
Calculate statistic(s) of interest;
3. Repeat steps 1 and 2 several times to obtain a sample of the statistic(s);

- Number of replicate samples is a trade-off between computational efficiency and accuracy;
- Try with a small number (e.g., 25), then do it again. If answer does not change drastically, then 25 is sufficient. Otherwise, try this experiment again with a larger number (e.g., 50). For some statistics, there are recommendations for how large it should be.

Bootstrap CI's

Assumptions: *Sample* distribution represents *population* distribution.
4. Calculate CI's based on the sample obtained in step 3.

- Many ways to do this;
- Based on bootstrap "tables" (not recommended);
- Based on bootstrap percentiles (recommended);
- Simplest choice is percentile method, but two further assumptions about the distribution of the statistic are implied, and violations of either assumption will lead to intervals that are too narrow.
- One method that accounts for these assumption violations is the BCa. It is highly accurate, but computationally intensive.
- ABC is a faster approximation to BCa, but only works for certain types of statistics.
- MET has percentile and BCa.

References and further reading

- Gilleland, E., 2010: Confidence intervals for forecast verification. NCAR Technical Note NCAR/TN 479+STR, 71pp. *(Available at:* http://nldr.library.ucar.edu/collections/technotes/asset 000 000 000 846.pdf)
- Jolliffe and Stehenson (2003): Forecast verification: A practitioner's guide, Wiley & sons
- JWGFVR (2009): Recommendation on verification of precipitation forecasts. WMO/TD report, no.1485 WWRP 2009 1
- Nurmi (2003): Recommendations on the verification of local weather forecasts. ECMWF Techical Memorandum, no. 430
- Wilks (2006): Statistical methods in the atmospheric sciences, ch. 7. Academic Press

See also

- http://tinyurl.com/verif training
- http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif web page.html

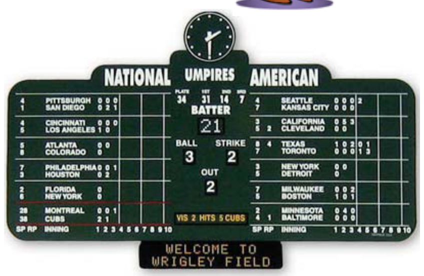
Verification of Continuous Forecasts

Presented by
Tressa L. Fowler

Adapted from presentations created by Barbara Casati and Barbara Brown

Developmental Testbed Center

- Exploratory methods
  - Scatter plots
  - Discrimination plots
  - Box plots
- Statistics
  - Bias
  - Error statistics
  - Robustness
  - Comparisons

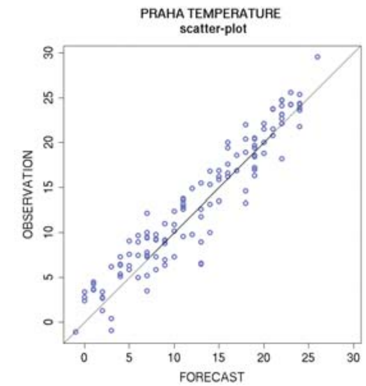


Exploratory methods: joint distribution

Scatter-plot: plot of observation versus forecast values

Perfect forecast = obs, points should be on the 45º diagonal

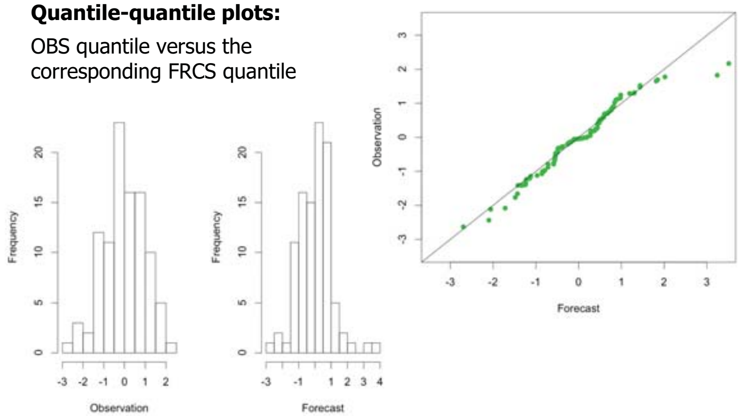
Provides information on: bias, outliers, error magnitude, linear association, peculiar behaviours in extremes, misses and false alarms (link to contingency table)



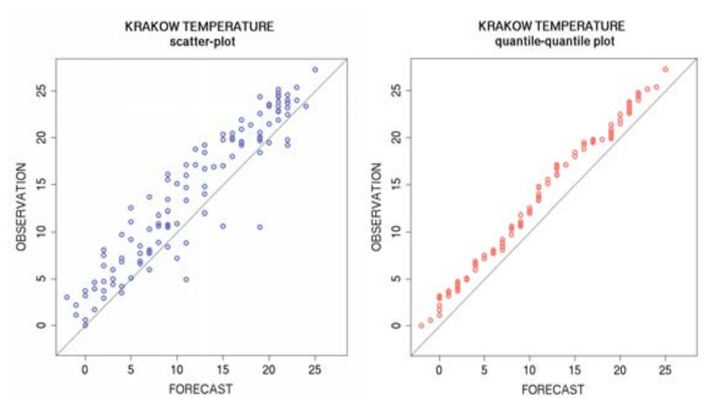
Exploratory methods: marginal distribution

Quantile-quantile plots:

OBS quantile versus the corresponding FRCS quantile

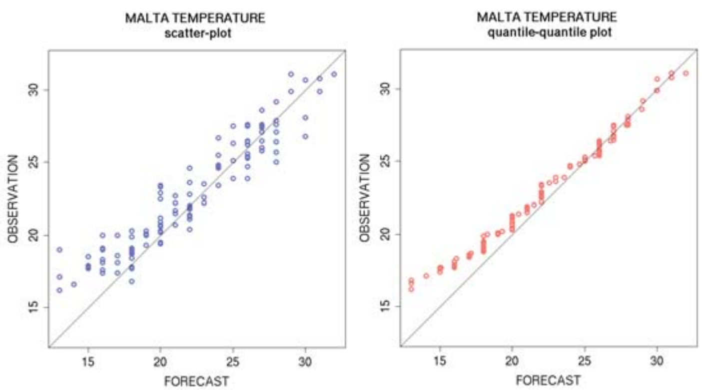


Scatter-plot and qq-plot: example 1
Q: is there any bias? Positive (over-forecast) or negative (under-forecast)?



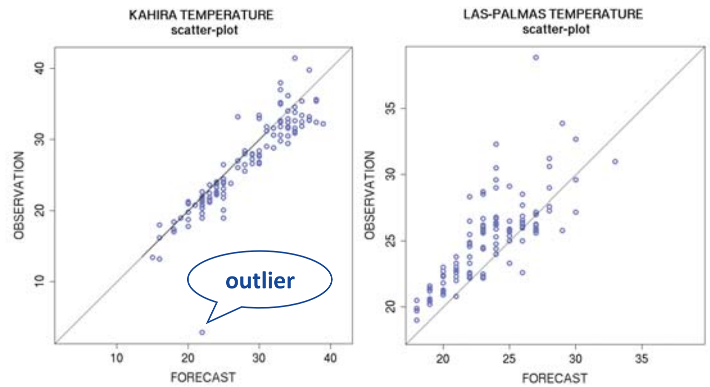
Scatter-plot and qq-plot: example 2

Describe the peculiar behaviour of low temperatures



Scatter-plot: example 3

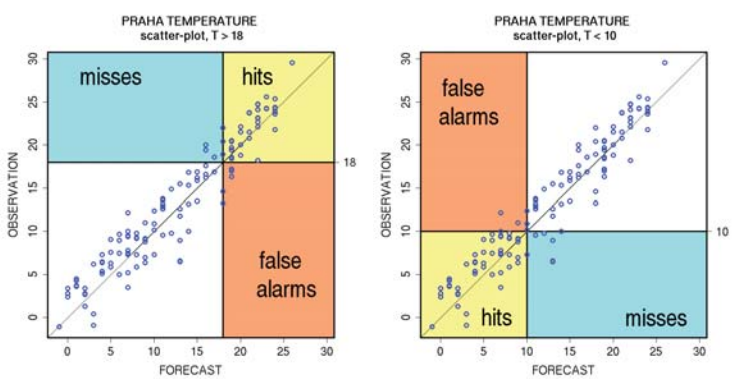
Describe how the error varies as the temperatures grow

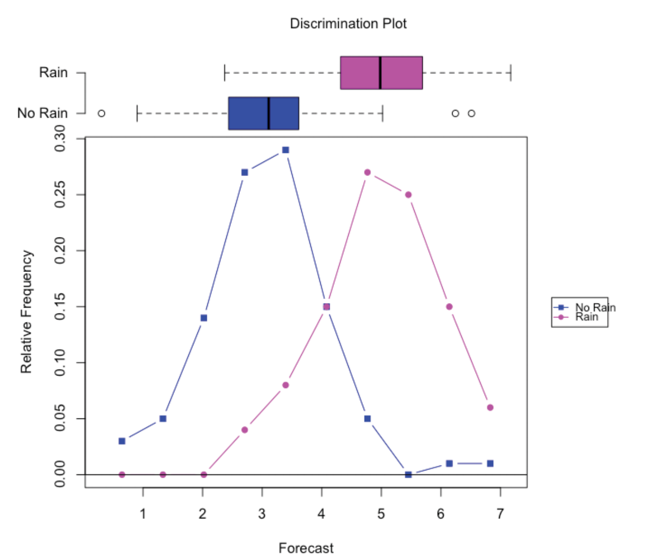


Scatter-plot and Contingency Table

Does the forecast detect correctly temperatures above 18 degrees ?

Does the forecast detect correctly temperatures below 10 degrees ?





Exploratory methods: marginal distributions

Visual comparison:
Histograms, box-plots, …

Summary statistics:

- ***Location:***

$$\text{mean} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

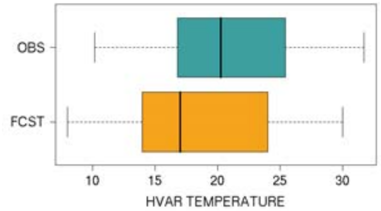
$$\text{median} = q_{0.5}$$

- ***Spread:***

$$\text{st dev} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}$$

Inter Quartile Range =

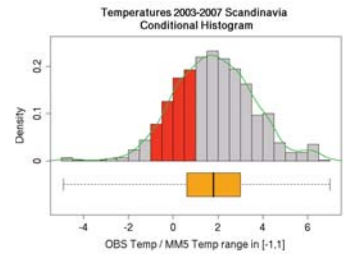
$$\text{IQR} = q_{0.75} - q_{0.25}$$

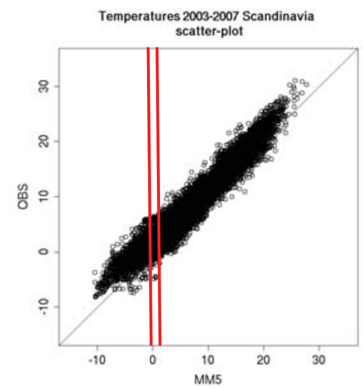


	MEAN	MEDIAN	STDEV	IQR
OBS	20.71	20.25	5.18	8.52
FRCS	18.62	17.00	5.99	9.75

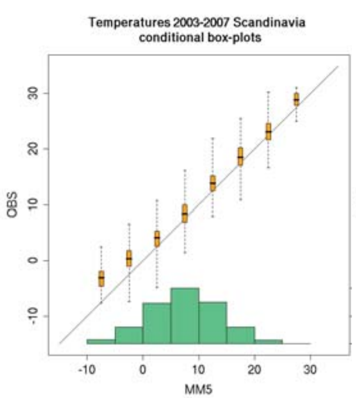
Exploratory methods: conditional distributions

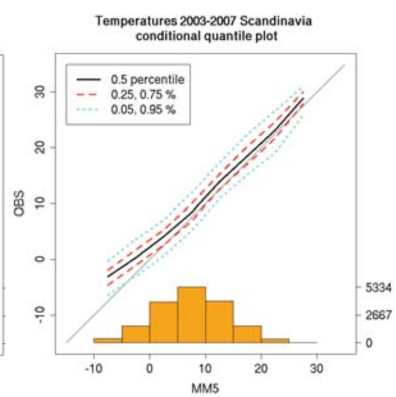
Conditional histogram and conditional box plot





Exploratory methods: conditional qq-plot





Continuous scores: linear bias

$$\text{linear bias} = \text{Mean Error} = \frac{1}{n}\sum_{i=1}^{n}\left(f_i - o_i\right) = \overline{f} - \overline{o}$$

Attribute: measures the bias

Mean Error = average of the errors = difference between the means

It indicates the average direction of error: positive bias indicates over-forecast, negative bias indicates under-forecast (y=forecast, x=observation)

Does not indicate the magnitude of the error (positive and negative error can cancel outs)

Bias correction: misses (false alarms) improve at the expenses of false alarms (misses). Q: If I correct the bias in an over-forecast, do false alarms grow or decrease ? And the misses ?

Good practice rules: sample used for evaluating bias correction should be consistent with sample corrected (e.g. winter separated by summer); for fair validation, cross validation should be adopted for bias corrected forecasts

Mean Absolute Error

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|f_i - o_i\right|$$

Attribute: measures accuracy

Average of the magnitude of the errors

Linear score = each error has same weight

It does not indicates the direction of the error, just the magnitude

Median Absolute Deviation

$$MAD = median\left\{\left|f_i - o_i\right|\right\}$$

Attribute: measures accuracy

Median of the magnitude of the errors

Very robust

Extreme errors have no effect

Continuous scores: MSE

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(f_i - o_i\right)^2$$

Attribute: measures accuracy

Average of the squares of the errors: it measures the magnitude of the error, weighted on the squares of the errors

it does not indicate the direction of the error

Quadratic rule, therefore large weight on large errors:
→ good if you wish to penalize large error
→ sensitive to large values (e.g. precipitation) and outliers; sensitive to large variance (high resolution models); encourage conservative forecasts (e.g. climatology)

Continuous scores: RMSE

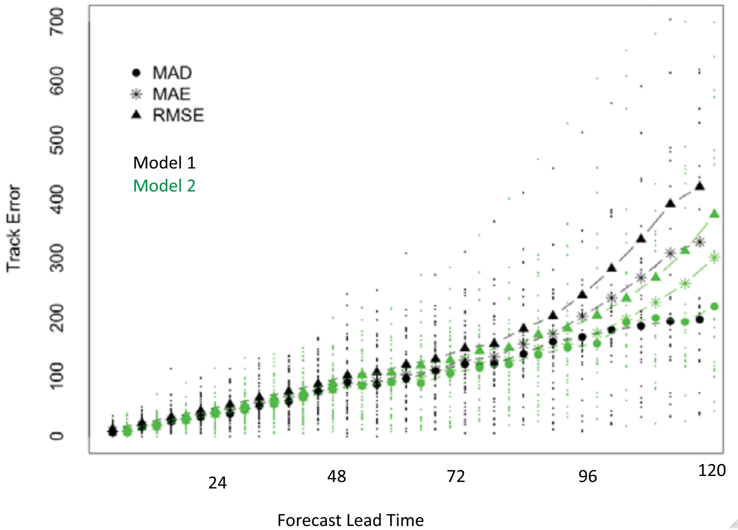
$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(f_i - o_i\right)^2}$$

Attribute: measures accuracy

RMSE is the squared root of the MSE: measures the magnitude of the error retaining the variable unit (e.g. ^{o}C)

Similar properties of MSE: it does not indicate the direction the error; it is defined with a quadratic rule = sensitive to large values, etc.

NOTE: RMSE is always larger or equal than the MAE



Continuous scores: linear correlation

$$r_{XY} = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \cdot \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \frac{\mathrm{cov}(Y,X)}{s_Y s_X}$$

Attribute: measures association

Measures linear association between forecast and observation

Y and X rescaled (non-dimensional) covariance: ranges in [-1,1]

It is not sensitive to the bias

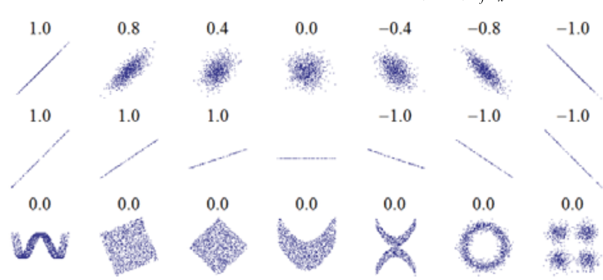
The correlation coefficient alone does not provide information on the inclination of the regression line (it says only is it is positively or negatively tilted); observation and forecast variances are needed; the slope coefficient of the regression line is given by $b = (s_X/s_Y)r_{XY}$

Not robust = better if data are normally distributed
Not resistant = sensitive to large values and outliers

Scores for continuous forecasts

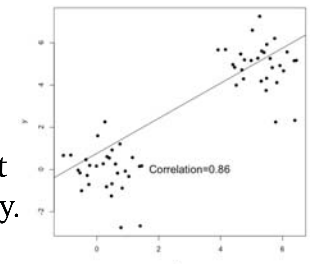
Simplest overall measure of performance:
Correlation coefficient

$$\rho_{fx} = \frac{Cov(f,x)}{\sqrt{Var(f)Var(x)}} \qquad r_{fx} = \frac{\sum_{i=1}^{n}(f_i - \bar{f})(x_i - \bar{x})}{(n-1)s_f s_x}$$



Continuous scores: anomaly correlation

- Correlation calculated on anomaly.
- Anomaly is difference between what was forecast (observed) and climatology.
- Centered or uncentered versions.



MSE and bias correction

$$MSE = \left(\bar{f} \quad \bar{o}\right)^2 + s_f^2 + s_o^2 \quad 2s_f s_o r_{fo}$$

$$MSE = ME^2 + \mathrm{var}(f \quad o)$$

- MSE is the sum of the squared bias and the variance. So ↑ bias = ↑ MSE

Continuous skill scores: MAE skill score

$$SS_{MAE} = \frac{MAE - MAE_{ref}}{MAE_{perf} - MAE_{ref}} = 1 - \frac{MAE}{MAE_{ref}}$$

Attribute: measures skill

Skill score: measure the forecast accuracy with respect to the accuracy of a reference forecast: positive values = skill; negative values = no skill

Difference between the score and a reference forecast score, normalized by the score obtained for a perfect forecast minus the reference forecast score (for perfect forecasts MAE=0)

Reference forecasts:

- **persistence:** appropriate when time-correlation > 0.5
- **sample climatology:** information only a posteriori
- **actual climatology:** information a priori

Continuous skill scores: MSE skill score

$$SS_{MSE} = \frac{MSE - MSE_{ref}}{MSE_{perf} - MSE_{ref}} = 1 - \frac{MSE}{MSE_{ref}}$$

Attribute: measures skill

Same definition and properties as the MAE skill score: measure accuracy with respect to reference forecast, positive values = skill; negative values = no skill

Sensitive to sample size (for stability) and sample climatology (e.g. extremes): needs large samples

Reduction of Variance: MSE skill score with respect to climatology.
If sample climatology is considered:

linear correlation

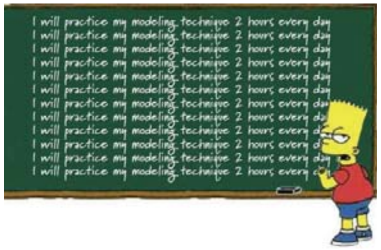
bias

$$Y = \bar{X}; \quad MSE_{cli} = s_X^2 \quad \text{and} \quad RV = 1 - \frac{MSE}{s_X^2} = r_{XY}^2 - \left(r_{XY} - \frac{s_Y}{s_X}\right)^2 - \left(\frac{\bar{Y} - \bar{X}}{s_X}\right)^2$$

reliability: regression line slope coeff $b=(s_X/s_Y)r_{XY}$

Continuous skill scores: good practice rules

- Use same climatology for the comparison of different models.
- When evaluating the Reduction of Variance, **sample climatology** gives always worse skill score than **long-term climatology**: ask always which climatology is used to evaluate the skill.



Continuous skill scores: good practice rules

- If the climatology is calculated pulling together data from many different stations and times of the year, the skill score will be better than if **a different climatology for each station and month of the year are used**.
  - In the former case the model gets credit from forecasting correctly seasonal trends and specific locations climatologies.
  - In the latter case the specific topographic effects and long term trends are removed and the forecast discriminating capability is better evaluated. Choose the appropriate climatology for fulfilling your verification purposes.
- Persistence forecast: use same time of the day to avoid diurnal cycle effects.

Continuous Scores of **Ranks**

Problem: Continuous scores sensitive to large values or non robust.

Solution: Use the **ranks** of the variable, rather than its actual values.

Temp °C	27.4	21.7	24.2	23.1	19.8	25.5	24.6	22.3
rank	8	2	5	4	1	7	6	3

The value-to-rank transformation:
- diminish effects due to large values
- transform distribution to a Uniform distribution
- remove bias

Rank correlation is the most common.

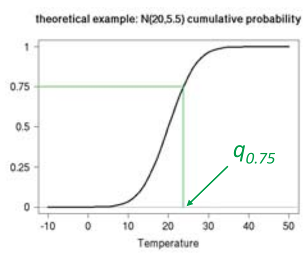
Linear Error in Probability Space

$$LEPS = \frac{1}{n}\sum_{i=1}^{n}\left|F_X(y_i) - F_X(x_i)\right|$$

The LEPS is a MAE evaluated by using the cumulative frequencies of the observation

Errors in the tail of the distribution are penalized less than errors in the centre of the distribution

MAE and LEPS are minimized by the median correction



Thank you!

References:

Jolliffe and Stephenson (2003): Forecast Verification: a practitioner's guide, Wiley & Sons, 240 pp.

Wilks (2005): Statistical Methods in Atmospheric Science, Academic press, 467 pp.

Stanski, Burrows, Wilson (1989) Survey of Common Verification Methods in Meteorology

http://www.eumetcal.org.uk/eumetcal/verification/www/english/courses/msgcrs/index.htm

http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif web page.html

Release History

- **METv0.9**: Beta release – July, 2007
- **METv1.0**: First official release – January, 2008
- **METv1.1**: Incremental upgrades – July, 2008
- **METv2.0**: Probabilistic forecasts – April, 2009
- **METv3.0**: Current Version – September, 2010
  - Pre-installed on tutorial machines
  - 1100+ registered users from 94 countries
  - 50/50 University/Non-University users
  - On-line tutorial available for METv3.0
  - Hands-on tutorial offered with the WRF-Tutorial

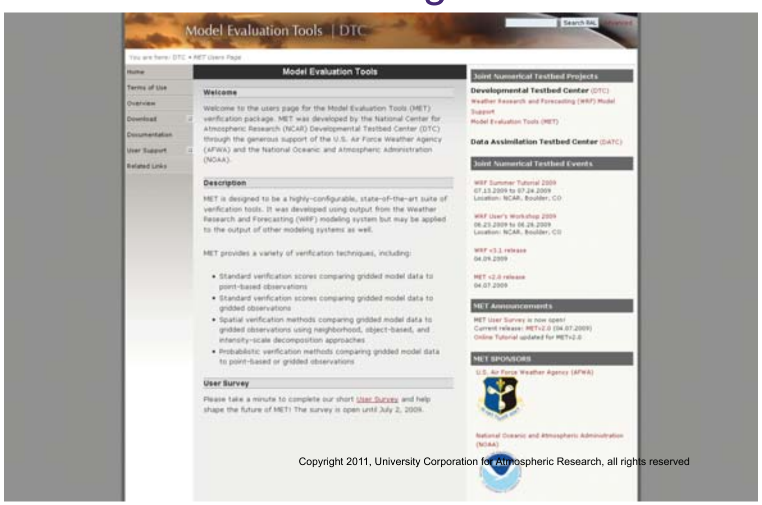
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Downloading MET

- Download MET release and compile locally.
  - Register and download: **www.dtcenter.org/met/users**
- Language:
  - Primarily in C++ with calls to a Fortran library
- Supported Platforms and Compilers:
  1. Linux with GNU compilers
  2. Linux with Portland Group (PGI) compilers
  3. Linux with Intel compilers
  4. IBM machines with IBM compilers

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

www.dtcenter.org/met/users



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Dependencies

- REQUIRED:
  - GNU Make Utility
  - C++/Fortran Compilers (GNU, PGI, Intel, or IBM)
  - NetCDF Library
  - BUFRLIB Library
  - GNU Scientific Library (GSL)
  - *F2C or G2C Library (only for some compilers)*
- RECOMMENDED:
  - WRF Post-Processor
  - COPYGB (included with WRF-Post)
  - R statistics and graphics package

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Directory Structure

File or Directory	Contents
README	Installation instructions and release notes.
Makefile_gnu (pgi, intel, ibm)	Top level Makefile to be configured prior to building MET.
lib/	Source code for internal MET libraries.
src/	Source code for the MET applications.
doc/	MET User's Guide.
bin/	Built MET executables.
scripts/	Test scripts to be run after building MET.
data/	Sample data used by the test scripts.
out/	Output generated by the test scripts.
tools/	Source code for optional MET tools.

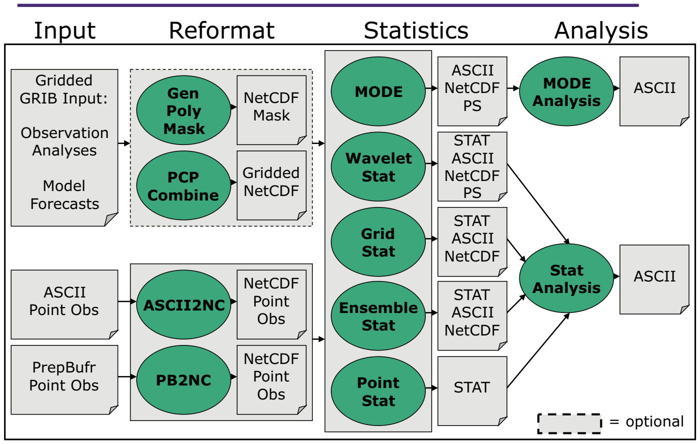
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Building MET

- Steps for building MET:
  1. Build required libraries.
     - Same family of compilers for MET
  2. Download and unpack latest MET patches.
  3. Select the appropriate Makefile.
     - GNU, PGI, Intel, or IBM
  4. Edit the Makefile.
     - C++ and Fortran compilers
     - Paths for NetCDF, BUFRLIB, and GSL libraries
  5. Run Make to build all of the MET tools.
  6. Run the test script and check for runtime errors.
     - Runs each of the MET tools at least once.
     - Uses sample data distributed with the tarball.

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

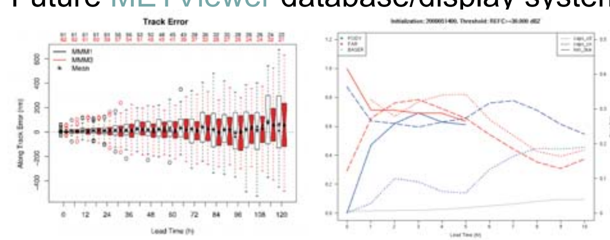
MET v3.0 Flowchart



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Graphics

- Limited graphics incorporated into MET
- Options for plotting MET statistical output
  - R, NCL, IDL, GNUPlot, and many others
- Sample plotting scripts on MET website
- Future METViewer database/display system



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

R Statistics and Graphics

- The R Project for Statistical Computing (www.r-project.org)
  - Powerful statistical analysis and plotting tools
  - Large and growing user community
  - Freely available and well supported for Linux/ Windows/Mac
- Sample R plotting and analysis scripts posted on the MET website
- Use R to plot data in the practical sessions

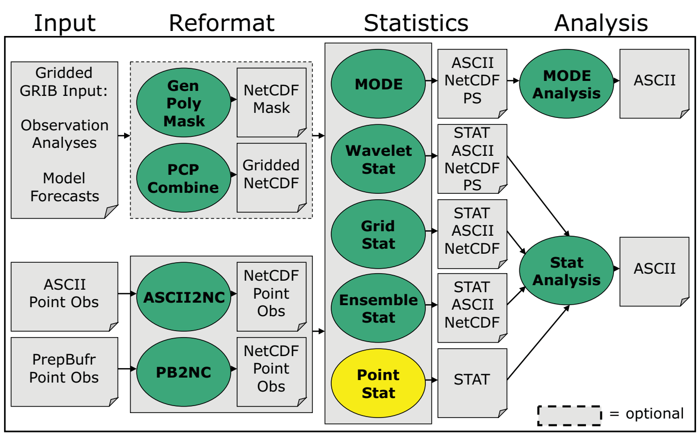
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Configuration Files

- MET tools controlled using command line options and ASCII configuration files
  - Well commented and documented in MET User's Guide
  - Easy to modify
  - Distributed with the tarball
- Configuration files control things such as:
  - Fields/levels to be verified
  - Thresholds to be applied
  - Interpolation methods to be used
  - Verification methods to be applied
  - Regions over which to accumulate statistics

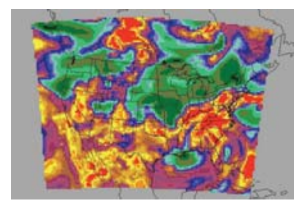
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat Tool



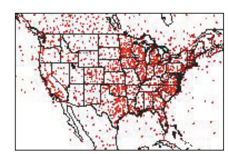
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat: Overview



- Compare **gridded forecasts** to **point observations.**
- Accumulate matched pairs over a defined area at a single point in time.
- Verify one or more variables/ levels.
- Analysis tool provided to aggregate through time.

- Verification methods:
  - Continuous statistics for raw fields.
  - Single and Multi Category counts and statistics for thresholded fields.
  - Parametric and non parametric confidence intervals for statistics.
  - Compute partial sums for raw fields and/or the raw matched pair values.
  - Methods for probabilistic forecasts.



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat: Input/Output

- Input Files
  - Gridded forecast file
    - GRIB output of WRF Post Processor (or other)
    - NetCDF output of PCP Combine or p_interp
  - Point observation file
    - NetCDF output of PB2NC
    - NetCDF output of ASCII2NC
  - ASCII configuration file
- Output Files
  - ASCII statistics file with all output lines (end with ".stat")
  - Optional ASCII files sorted by line type with a header row (ends with "_TYPE.txt")

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat: Usage

Usage: point_stat
fcst_file
obs_file
config_file
[-climo climo_file]
[-point_obs netcdf_file]
[-fcst_valid time]
[-fcst_lead time]
[-obs_valid_beg time]
[-obs_valid_end time]
[-outdir path]
[-v level]

fcst file	GRIB or NC forecast file
obs file	NC point observation file (PB2NC or ASCII2NC)
config file	ASCII configuration file
climo	Climo file for computing anomaly partial sums
point obs	Additional NC point observation files
fcst valid	Forecast valid time
fcst lead	Forecast lead time
obs valid beg	Beginning of valid time window for matching
obs valid end	End of valid time window for matching
outdir	Output directory to be used
v	Level of logging

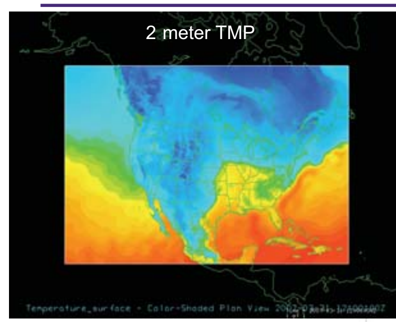
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

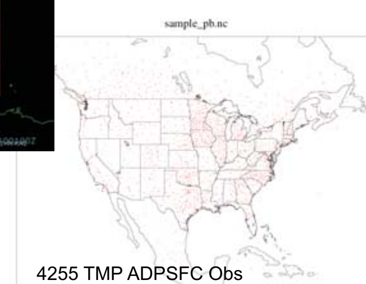
Point-Stat: Configuration

- 28 configurable parameters – only set a few:
  - Temperature at the surface (2-meter).
    - **fcst_field[] = ["TMP/Z2"];**
  - Temperature below freezing.
    - **fcst_thresh[] = ["gt273 gt283 gt293"];**
  - Match to observations at the surface.
    - **message_type[] = ["ADPSFC"];**
  - Look at all the points in my domain.
    - **mask_grid[] = ["FULL"];**
  - Match observation to the nearest forecast value.
    - **interp_wdth[] = [1];**
  - Generate all possible statistic types, except probabilistic.
    - **output_flag[] = [2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 0, 0, 0, 2];**

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat: Input





Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat: Run

- METv3.0/bin/point_stat \
  sample_fcst.grb sample_pb.nc \
  PointStatConfig_TMPZ2 -outdir out -v 2

```
Forecast File: sample fcst grb
Climatology File: none
Configuration File: PointStatConfig TMPZ2
Observation File: sample pb nc

Reading records for TMP/Z2
For TMP/Z2 found 1 forecast levels and 0 climatology levels

Searching 89759 observations from 9716 PrepBufr messages

Processing TMP/Z2 versus TMP/Z2, for observation type ADPSFC, over region FULL, for interpolation method UW MEAN(1),
using 4250 pairs
Computing Categorical Statistics
Computing Multi Category Statistics
Computing Continuous Statistics
Computing Scalar Partial Sums

Output file: out/point stat 360000L 20070331 120000V stat
Output file: out/point stat 360000L 20070331 120000V fho txt
Output file: out/point stat 360000L 20070331 120000V ctc txt
Output file: out/point stat 360000L 20070331 120000V cts txt
Output file: out/point stat 360000L 20070331 120000V mctc txt
Output file: out/point stat 360000L 20070331 120000V mcts txt
Output file: out/point stat 360000L 20070331 120000V cnt txt
Output file: out/point stat 360000L 20070331 120000V sl1l2 txt
Output file: out/point stat 360000L 20070331 120000V sal1l2 txt
Output file: out/point stat 360000L 20070331 120000V vl1l2 txt
Output file: out/point stat 360000L 20070331 120000V val1l2 txt
Output file: out/point stat 360000L 20070331 120000V mpr txt
```

Point-Stat: ASCII Output Types

- Statistics line types: 15 possible
  - Categorical   Single Threshold
    - Contingency table counts and stats (FHO, CTC, CTS)
  - Categorical   Multiple Thresholds
    - NxN Contingency table counts and stats (MCTC, MCTS)
  - Continuous   raw fields
    - Continuous statistics (CNT)
    - Partial Sums (SL1L2, SAL1L2, VL1L2, VAL1L2)
  - Probabilistic
    - Nx2 Contingency table counts and stats (PCT, PSTD)
    - Continuous statistics and ROC curve (PJC, PRC)
  - Matched pairs
    - Raw matched pairs   a lot of data! (MPR)
- 21 header columns common to all line types
- Remaining columns specific to each line type

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat: Sample Output

1. STAT file output for sample run:
   - 1 line each for CNT, SL1L2 , MCTC, MCTS
   - 3 lines each for FHO, CTC, CTS
   - 4,250 lines for MPR!
2. Additional TXT files for each line type

```
Output file: out/point_stat_360000L_20070331_120000V.stat
Output file: out/point_stat_360000L_20070331_120000V_fho.txt
Output file: out/point_stat_360000L_20070331_120000V_ctc.txt
Output file: out/point_stat_360000L_20070331_120000V_cts.txt
Output file: out/point_stat_360000L_20070331_120000V_mctc.txt
Output file: out/point_stat_360000L_20070331_120000V_mcts.txt
Output file: out/point_stat_360000L_20070331_120000V_cnt.txt
Output file: out/point_stat_360000L_20070331_120000V_sl1l2.txt
Output file: out/point_stat_360000L_20070331_120000V_sal1l2 txt
Output file: out/point_stat_360000L_20070331_120000V_vl1l2.txt
Output file: out/point_stat_360000L_20070331_120000V_val1l2 txt
Output file: out/point_stat_360000L_20070331_120000V_mpr.txt
```

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat: CTC Output Line

VERSION	V3.0
MODEL	WRF
FCST LEAD	360000
FCST VALID BEG	20070331 120000
FCST VALID END	20070331 120000
OBS LEAD	000000
OBS VALID BEG	20070331 103000
OBS VALID END	20070331 133000
FCST VAR	TMP
FCST LEV	Z2
OBS VAR	TMP
OBS LEV	Z2
OBTYPE	ADPSFC

VX MASK	FULL
INTERP MTHD	UW MEAN
INTERP PNTS	1
FCST THRESH	>273.000
OBS THRESH	>273.000
COV THRESH	NA
ALPHA	NA
LINE TYPE	CTC
TOTAL	4250
FY OY	3275
FY ON	245
FN OY	102
FN ON	628

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

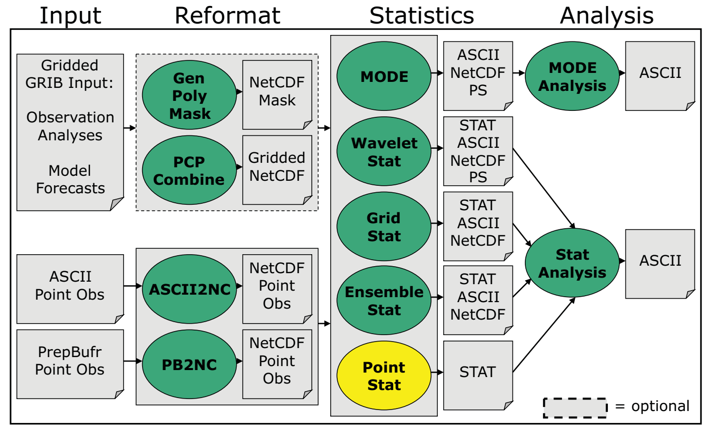
Point-Stat: Matched Pairs

- Matched Pair (MPR) line type contains 1 line for each matched pair.
  - Data overload!

TOTAL	INDEX	OBS SID	OBS LAT	OBS LON	OBS LVL	OBS ELV	FCST	OBS	CLIMO
4250	1	71600	43.93000	-60.01000	1010.79999	4.01053	271.99994	271.54999	NA
4250	2	71616	46.43000	-71.93000	1016.09998	102.04903	269.00017	269.45001	NA
4250	3	71629	44.23000	-78.36000	1004.50000	191.44466	273.00041	272.35001	NA
4250	4	71028	51.67000	-124.40000	916.50000	872.82202	264.00027	264.95001	NA
4250	5	71066	58.61000	-117.16000	973.90002	337.50449	271.99994	270.95001	NA
4250	6	71104	52.18000	-122.04000	906.50000	938.08594	271.00029	264.35001	NA
4250	7	71109	50.68000	-127.36000	1020.20001	22.03931	274.99971	275.04999	NA
4250	8	71150	50.45000	-100.59000	949.09998	562.38477	271.99994	271.75000	NA
4250	9	71177	57.13000	-61.47000	899.70001	834.87476	257.99991	254.64999	NA
4250	10	71197	47.56000	-59.16000	1000.90002	40.06803	271.99994	269.54999	NA
4250	11	71378	47.41000	-72.79000	1006.90002	169.37592	266.00040	265.95001	NA
4250	12	71415	45.76000	-62.68000	1014.00000	1.99518	269.00017	268.64999	NA
4250	13	71425	49.24000	-65.33000	1014.90002	28.96468	264.00027	267.25000	NA
4250	14	71437	43.29000	-79.79000	1017.79999	77.03765	274.00006	275.85001	NA
4250	15	71473	48.78000	-123.04000	1015.70001	23.93772	278.00031	280.25000	NA

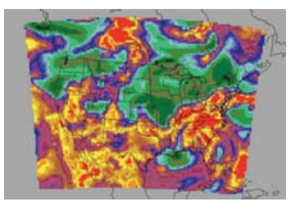
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat Tool



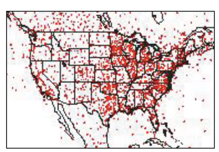
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat: Overview



- Compare **gridded forecasts** to **point observations.**
- Accumulate matched pairs over a defined area at a single point in time.
- Verify one or more variables/ levels.
- Analysis tool provided to aggregate through time.

- Verification methods:
  - Continuous statistics for raw fields.
  - Single and Multi Category counts and statistics for thresholded fields.
  - Parametric and non parametric confidence intervals for statistics.
  - Compute partial sums for raw fields and/or the raw matched pair values.
  - Methods for probabilistic forecasts.



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat: Input/Output

- Input Files
  - Gridded forecast file
    - GRIB output of WRF Post Processor (or other)
    - NetCDF output of PCP Combine or p_interp
  - Point observation file
    - NetCDF output of PB2NC
    - NetCDF output of ASCII2NC
  - ASCII configuration file
- Output Files
  - ASCII statistics file with all output lines (end with ".stat")
  - Optional ASCII files sorted by line type with a header row (ends with "_TYPE.txt")

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat: Usage

Usage: point_stat

fcst_file

obs_file

config_file

[-climo climo_file]

[-point_obs netcdf_file]

[-fcst_valid time]

[-fcst_lead time]

[-obs_valid_beg time]

[-obs_valid_end time]

[-outdir path]

[-v level]

fcst file	GRIB or NC forecast file
obs file	NC point observation file (PB2NC or ASCII2NC)
config file	ASCII configuration file
climo	Climo file for computing anomaly partial sums
point obs	Additional NC point observation files
fcst valid	Forecast valid time
fcst lead	Forecast lead time
obs valid beg	Beginning of valid time window for matching
obs valid end	End of valid time window for matching
outdir	Output directory to be used
v	Level of logging

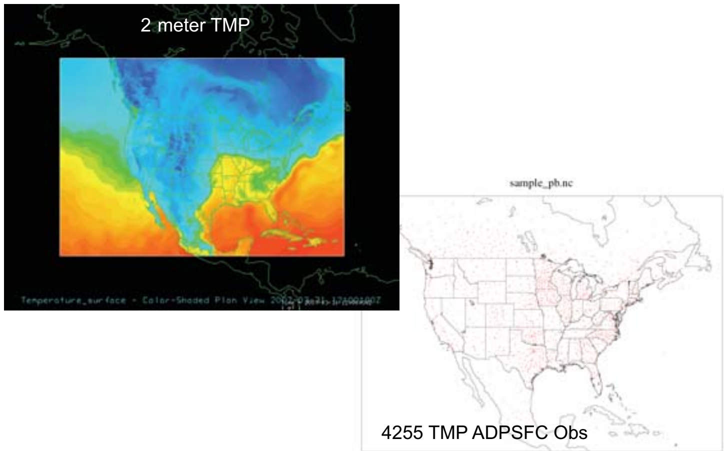
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat: Configuration

- 28 configurable parameters – only set a few:
  - Temperature at the surface (2-meter).
    - **fcst_field[] = ["TMP/Z2"];**
  - Temperature below freezing.
    - **fcst_thresh[] = ["gt273 gt283 gt293"];**
  - Match to observations at the surface.
    - **message_type[] = ["ADPSFC"];**
  - Look at all the points in my domain.
    - **mask_grid[] = ["FULL"];**
  - Match observation to the nearest forecast value.
    - **interp_wdth[] = [1];**
  - Generate all possible statistic types, except probabilistic.
    - **output_flag[] = [2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 0, 0, 0, 2];**

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat: Input



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat: Run

- METv3.0/bin/point_stat \
  sample_fcst.grb sample_pb.nc \
  PointStatConfig_TMPZ2 -outdir out -v 2

```
Forecast File: sample fcst grb
Climatology File: none
Configuration File: PointStatConfig TMPZ2
Observation File: sample pb nc

Reading records for TMP/Z2
For TMP/Z2 found 1 forecast levels and 0 climatology levels

Searching 89759 observations from 9716 PrepBufr messages

Processing TMP/Z2 versus TMP/Z2, for observation type ADPSFC, over region FULL, for interpolation method UW MEAN(1),
using 4250 pairs
Computing Categorical Statistics
Computing Multi Category Statistics
Computing Continuous Statistics
Computing Scalar Partial Sums

Output file: out/point stat 360000L 20070331 120000V stat
Output file: out/point stat 360000L 20070331 120000V fho txt
Output file: out/point stat 360000L 20070331 120000V ctc txt
Output file: out/point stat 360000L 20070331 120000V cts txt
Output file: out/point stat 360000L 20070331 120000V mctc txt
Output file: out/point stat 360000L 20070331 120000V mcts txt
Output file: out/point stat 360000L 20070331 120000V cnt txt
Output file: out/point stat 360000L 20070331 120000V sl1l2 txt
Output file: out/point stat 360000L 20070331 120000V sal1l2 txt
Output file: out/point stat 360000L 20070331 120000V vl1l2 txt
Output file: out/point stat 360000L 20070331 120000V val1l2 txt
Output file: out/point stat 360000L 20070331 120000V mpr txt
```

Point-Stat: ASCII Output Types

- Statistics line types: 15 possible
  - Categorical Single Threshold
    - Contingency table counts and stats (FHO, CTC, CTS)
  - Categorical Multiple Thresholds
    - NxN Contingency table counts and stats (MCTC, MCTS)
  - Continuous raw fields
    - Continuous statistics (CNT)
    - Partial Sums (SL1L2, SAL1L2, VL1L2, VAL1L2)
  - Probabilistic
    - Nx2 Contingency table counts and stats (PCT, PSTD)
    - Continuous statistics and ROC curve (PJC, PRC)
  - Matched pairs
    - Raw matched pairs a lot of data! (MPR)
- 21 header columns common to all line types
- Remaining columns specific to each line type

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat: Sample Output

1. STAT file output for sample run:
   - 1 line each for CNT, SL1L2 , MCTC, MCTS
   - 3 lines each for FHO, CTC, CTS
   - 4,250 lines for MPR!
2. Additional TXT files for each line type

```
Output file: out/point_stat_360000L_20070331_120000V.stat
Output file: out/point_stat_360000L_20070331_120000V_fho.txt
Output file: out/point_stat_360000L_20070331_120000V_ctc.txt
Output file: out/point_stat_360000L_20070331_120000V_cts.txt
Output file: out/point_stat_360000L_20070331_120000V_mctc.txt
Output file: out/point_stat_360000L_20070331_120000V_mcts.txt
Output file: out/point_stat_360000L_20070331_120000V_cnt.txt
Output file: out/point_stat_360000L_20070331_120000V_sl1l2.txt
Output file: out/point_stat_360000L_20070331_120000V_sal1l2.txt
Output file: out/point_stat_360000L_20070331_120000V_vl1l2.txt
Output file: out/point_stat_360000L_20070331_120000V_val1l2.txt
Output file: out/point_stat_360000L_20070331_120000V_mpr.txt
```

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Point-Stat: CTC Output Line

VERSION	V3.0
MODEL	WRF
FCST LEAD	360000
FCST VALID BEG	20070331 120000
FCST VALID END	20070331 120000
OBS LEAD	000000
OBS VALID BEG	20070331 103000
OBS VALID END	20070331 133000
FCST VAR	TMP
FCST LEV	Z2
OBS VAR	TMP
OBS LEV	Z2
OBTYPE	ADPSFC

VX MASK	FULL
INTERP MTHD	UW MEAN
INTERP PNTS	1
FCST THRESH	>273.000
OBS THRESH	>273.000
COV THRESH	NA
ALPHA	NA
LINE TYPE	CTC
TOTAL	4250
FY OY	3275
FY ON	245
FN OY	102
FN ON	628

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

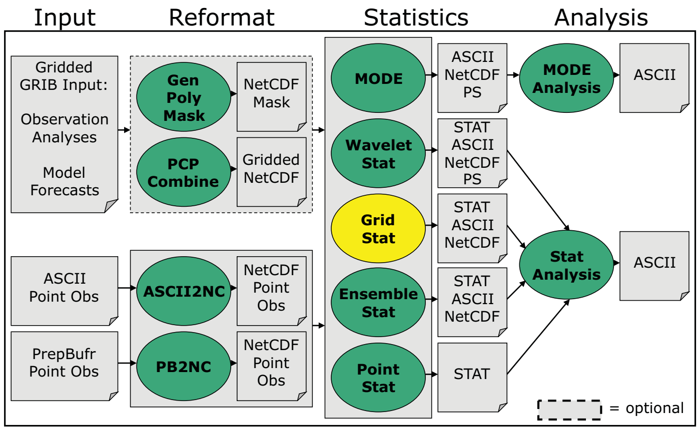
Point-Stat: Matched Pairs

- Matched Pair (MPR) line type contains 1 line for each matched pair.
  - Data overload!

TOTAL	INDEX	OBS SID	OBS LAT	OBS LON	OBS LVL	OBS ELV	FCST	OBS	CLIMO
4250	1	71600	43.93000	-60.01000	1010.79999	4.01053	271.99994	271.54999	NA
4250	2	71616	46.43000	-71.93000	1016.09998	102.04903	269.00017	269.45001	NA
4250	3	71629	44.23000	-78.36000	1004.50000	191.44466	273.00041	272.35001	NA
4250	4	71028	51.67000	-124.40000	916.50000	872.82202	264.00027	264.95001	NA
4250	5	71066	58.61000	-117.16000	973.90002	337.50449	271.99994	270.95001	NA
4250	6	71104	52.18000	-122.04000	906.50000	938.08594	271.00029	264.35001	NA
4250	7	71109	50.68000	-127.36000	1020.20001	22.03931	274.99971	275.04999	NA
4250	8	71150	50.45000	-100.59000	949.09998	562.38477	271.99994	271.75000	NA
4250	9	71177	57.13000	-61.47000	899.70001	834.87476	257.99991	254.64999	NA
4250	10	71197	47.56000	-59.16000	1000.90002	40.06803	271.99994	269.54999	NA
4250	11	71378	47.41000	-72.79000	1006.90002	169.37592	266.00040	265.95001	NA
4250	12	71415	45.76000	-62.68000	1014.00000	1.99518	269.00017	268.64999	NA
4250	13	71425	49.24000	-65.33000	1014.90002	28.96468	264.00027	267.25000	NA
4250	14	71437	43.29000	-79.79000	1017.79999	77.03765	274.00006	275.85001	NA
4250	15	71473	48.78000	-123.04000	1015.70001	23.93772	278.00031	280.25000	NA

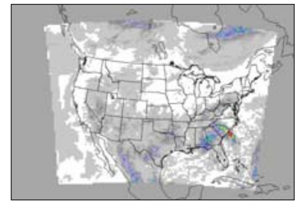
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Grid-Stat Tool



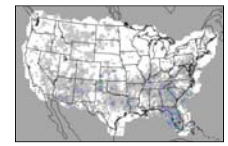
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Grid-Stat: Overview



- Compare **gridded forecasts** to **gridded observations** on the same grid.
- Accumulate matched pairs over a defined area at a single point in time.
- Verify one or more variables/ levels.
- Analysis tool provided to aggregate through time.

- Verification methods:
  - Continuous statistics for raw fields.
  - Single and Multi Category counts and statistics for thresholded fields.
  - Parametric and non parametric confidence intervals for statistics.
  - Compute partial sums for raw fields.
  - Methods for probabilistic forecasts.
  - Continuous statistics and categorical counts/statistics using neighborhood verification method.



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Grid-Stat: Input/Output

- Input Files
  - Gridded forecast and observation files
    - GRIB output of WRF Post-Processor (or other)
    - NetCDF output of PCP-Combine or p_interp
  - ASCII configuration file
- Output Files
  - ASCII statistics file with all output lines (end with ".stat")
  - Optional ASCII files sorted by line type with a header row (ends with "_TYPE.txt")
  - Optional NetCDF matched pairs file

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Grid-Stat: Usage

Usage: grid_stat

fcst_file
obs_file
config_file
[-fcst_valid time]
[-fcst_lead time]
[-obs_valid time]
[-obs_lead time]
[-outdir path]
[-v level]

fcst_file	Forecast file in GRIB or NetCDF
obs_file	Observation file in GRIB or NetCDF
config_file	ASCII configuration file
fcst_valid	Forecast valid time
fcst_lead	Forecast lead time
obs_valid	Observation valid time
obs_lead	Observation lead time
outdir	Output directory to be used
v	Level of logging

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Grid-Stat: Configuration

- 28 configurable parameters – only set a few:
  - Precipitation accumulated over 24 hours.
    - **fcst_field[] = [“APCP/A24”];**
  - Any rain and moderate rain.
    - **fcst_thresh[] = [“gt0.00 gt20.00”]; (mm)**
  - Look at all the points and only the eastern United States.
    - **mask_grid[] = [“FULL”];**
    - **mask_poly[] = [“EAST.poly”];**
  - Compute neighborhood statistics with two sizes.
    - **nbr_width[] = [3, 5];**
  - Generate all possible statistic types, except probabilistic.
    - **output_flag[] = [2, 2, 2, 2, 2, 2, 2, 2, 0, 0, 0, 0, 2, 2, 2, 1];**

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Grid-Stat: Run

- METv3.0/bin/grid_stat \
  sample_fcst.grb sample_obs.nc \
  GridStatConfig_APCP24 -outdir out -v 2

```
Forecast File: sample fcst grb
Observation File: sample obs nc
Configuration File: GridStatConfig APCP24

Processing APCP/A24 versus APCP/A24, for interpolation method UW MEAN(1), over region FULL, using 6412 pairs
Computing Categorical Statistics
Computing Multi Category Statistics
Computing Continuous Statistics
Processing APCP/A24 versus APCP/A24, for interpolation method UW MEAN(1), over region EAST, using 2586 pairs
Computing Categorical Statistics
Computing Multi Category Statistics
Computing Continuous Statistics
Processing APCP/A24 versus APCP/A24, for interpolation method NBRHD(9), raw thresholds of >0 000 and >0 000, over
region EAST, using 5829 pairs
    MORE NEIGHBORHOOD COMPUTATIONS

Output file: out/grid stat 240000L 20050808 000000V stat
Output file: out/grid stat 240000L 20050808 000000V fho txt
Output file: out/grid stat 240000L 20050808 000000V ctc txt
Output file: out/grid stat 240000L 20050808 000000V cts txt
Output file: out/grid stat 240000L 20050808 000000V mctc txt
Output file: out/grid stat 240000L 20050808 000000V mcts txt
Output file: out/grid stat 240000L 20050808 000000V cnt txt
Output file: out/grid stat 240000L 20050808 000000V sl1l2 txt
Output file: out/grid stat 240000L 20050808 000000V nbrctc txt
Output file: out/grid stat 240000L 20050808 000000V nbrcts txt
Output file: out/grid stat 240000L 20050808 000000V nbrcnt txt
Output file: out/grid stat 240000L 20050808 000000V pairs nc
```

Copyright 2011 University Corporation for Atmospheric Research all rights reserved

Grid-Stat: ASCII Output Types

- Statistics line types: 15 possible
  - Same as Point-Stat
    - FHO, CTC, CTS, MCTC, MCTS, CNT, SL1L2, and VL1L2
    - PCT, PSTD, PJC, and PRC
  - Omitted for Grid-Stat
    - *SAL1L2*, *VL1L2*, or *VAL1L2*
  - Neighborhood – apply threshold, define neighborhood
    - Neighborhood continuous statistics (NBRCNT)
    - Neighborhood contingency table counts (NBRCTC)
    - Neighborhood contingency table statistics (NBRCTS)
- 21 header columns common to all line types
- Remaining columns specific to each line type

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Grid-Stat: Sample Output

1. STAT file output for sample run:
   - 2 lines each for CNT, MCTC, MCTS, and SL1L2
     - = 2 verification regions (FULL and EAST)
   - 4 lines each for FHO, CTC, and CTS
     - = 2 regions * 2 thresholds
   - 8 lines each for NBRCNT, NBRCTC, NBRCTS
     - = 2 regions * 2 thresholds * 2 neighborhood sizes
2. Additional TXT files for each line type
3. NetCDF file containing matched pairs

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

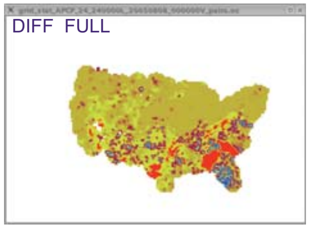
Grid-Stat: CTC Output Line

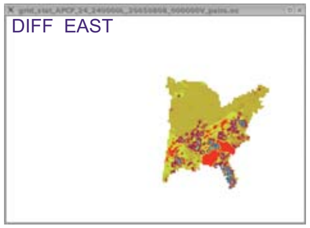
VERSION	V3.0	VX_MASK	EAST
MODEL	WRF	INTERP_MTHD	UW_MEAN
FCST_LEAD	240000	INTERP_PNTS	1
FCST_VALID_BEG	20050808_000000	FCST_THRESH	>=20.000
FCST_VALID_END	20050808_000000	OBS_THRESH	>=20.000
OBS_LEAD	000000	COV_THRESH	NA
OBS_VALID_BEG	20050808_000000	ALPHA	NA
OBS_VALID_END	20050808_000000	LINE_TYPE	CTC
FCST_VAR	APCP_24	TOTAL	2586
FCST_LEV	A24	FY_OY	5
OBS_VAR	APCP_24	FY_ON	104
OBS_LEV	A24	FN_OY	70
OBTYPE	MC_PCP	FN_ON	2407

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Grid-Stat: NetCDF Matched Pairs

- Forecast, observation, and difference fields for each combination of…
  - Variable, level, masking region, and interpolation method (smoothing)
- Sample output contains 6 fields:
  - FCST, OBS, and DIFF for FULL and EAST





Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Verifying Probabilities

- Probabilistic verification methods added for:
  - Grid-Stat, Point-Stat, and Stat-Analysis
- Define Nx2 contingency table using:
  - Multiple forecast probability thresholds
  - One observation threshold

Forecast	Observation: o = 1 (e.g., "Yes")	Observation: o = 0 (e.g., "No")	Total
p_1 = midpoint of (0 and threshold1)	n_{11}	n_{10}	$n_{1.} = n_{11} + n_{10}$
p_2 = midpoint of (threshold1 and threshold2)	n_{21}	n_{20}	$n_{2.} = n_{21} + n_{20}$
⋮	⋮	⋮	⋮
p_j = midpoint of (threshold$i$ and 1)	n_{j1}	n_{j0}	$n_j = n_{j1} + n_{j0}$
Total	$n_{.1} = \Sigma n_{j1}$	$n_{.0} = \Sigma n_{j0}$	$T = \Sigma n_j$

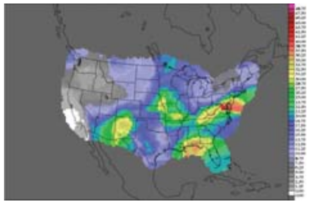
Example:

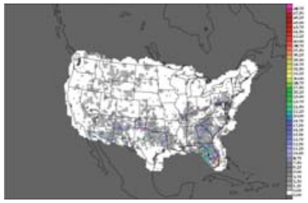
- FCST: Probability of precip [0.00, 0.25, 0.50, 0.75, 1.00]
- OBS: Accumulated precip > 0.00

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Verifying Probabilities: Example

- Verify probability of precip with total precip:



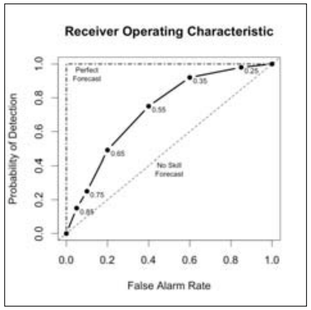


- Configuration file settings:
  - fcst_field[] = ["POP/Z0/PROB"];
  - obs_field[] = ["APCP/A12"];
  - fcst_thresh[] = ["ge0.00 ge0.25 ge0.50 ge0.75 ge1.00"];
  - obs_thresh[] = ["gt0.00"];

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Verifying Probabilities: Output

- Statistical Output (Line Type):
  - Nx2 Table Counts (PCT)
  - Joint/Conditional factorization table with calibration, refinement, likelihood, and base rate by threshold (PJC)
  - Receiver Operating Characteristic (ROC) plot points by threshold (PRC)
  - Reliability, resolution, uncertainty, area under ROC Curve, and Brier Score (PSTD)



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Comparing Different Fields

- Grid-Stat and Point-Stat may be used to compare two different variables.
  - User must interpret results.
  - Example: Convective Precip vs. Total Precip
  - Configuration file settings:
    - Selecting variable/levels:
      - **fcst_field[] = [“ACPCP/A24”];**
      - **obs_field[] = [“APCP/A24”];**
    - Selecting thresholds:
      - **fcst_thresh[] = [“gt0.0 ge20.0”];**
      - **obs_thresh[] = []; (leave blank to use fcst setting)**

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Verifying Winds

- Verify u, v, and speed, but not wind direction.
- Incremental support for wind direction:
  - Enhancements for Point-Stat and Grid-Stat:
    - Add wind speed thresholds to determine which u/v pairs are included in the vector partial sums (VL1L2).
  - Enhancements for Stat-Analysis:
    - Support new job to aggregate one or more vector partial sum lines and compute statistics for the wind direction errors.
      - Mean forecast and observation wind directions, mean error (F O), and mean absolute error

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Wind Direction: Example

Point-Stat: VL1L2 Lines

VX MASK	THRESH	LINE TYPE	TOTAL	UFBAR	VFBAR	UOBAR	VOBAR	UVFOBAR	UVFFBAR	UVOOBAR
DTC 165	> 1.000	VL1L2	653	1.91117	0.07900	1.40658	-0.06126	13.01039	18.12575	20.31649
DTC 165	> 3.000	VL1L2	279	3.13561	-0.35096	2.87061	-0.30072	26.50472	30.03257	38.25362
DTC 165	> 5.000	VL1L2	96	5.21268	-2.74580	5.47813	-2.01667	49.90791	51.10427	70.78802
DTC 166	> 1.000	VL1L2	2431	-1.62742	0.25391	-1.23402	-0.04393	18.48309	29.70179	21.89615
DTC 166	> 3.000	VL1L2	1610	-1.84581	0.16061	-1.47491	-0.11217	24.45214	36.67400	29.36032
DTC 166	> 5.000	VL1L2	520	-0.93518	-0.45435	-0.25923	-0.49558	37.21821	52.51917	47.26483

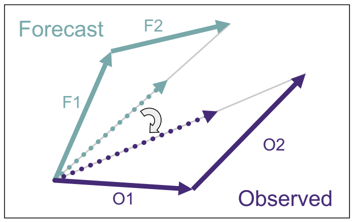
Stat-Analysis: aggregate stat jobs

```
JOB LIST:      -job aggregate stat -fcst thresh > 1.000 -line type VL1L2 -out line type WDIR
     COL NAME: TOTAL FBAR      OBAR     ME        MAE
ROW MEAN WDIR: 2     183.25038 0.22749  -3.02289  7.88372
    AGGR WDIR: 3084  103.87238 85.96574 -17.90663 NA
-----------------------------------------------------------------------------------------------
JOB LIST:      -job aggregate stat -fcst thresh > 3.000 -line type VL1L2 -out line type WDIR
     COL NAME: TOTAL FBAR     OBAR     ME        MAE
ROW MEAN WDIR: 2     5.67967  0.81565  -4.86402  4.86402
    AGGR WDIR: 1889  94.38140 80.45939 -13.92200 NA
-----------------------------------------------------------------------------------------------
JOB LIST:      -job aggregate stat -fcst thresh > 5.000 -line type VL1L2 -out line type WDIR
     COL NAME: TOTAL FBAR      OBAR      ME        MAE
ROW MEAN WDIR: 2     0.93288   338.91179 -22.02109 22.02109
    AGGR WDIR: 616   358.38152 319.08761 -39.29391 NA
```

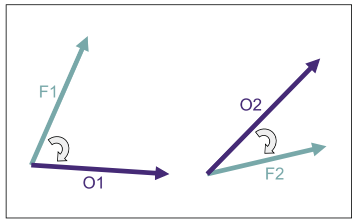
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Wind Direction: Output

- AGGR_WDIR
  1. Aggregate VL1L2 partial sums lines
  2. Derive wind directions and errors



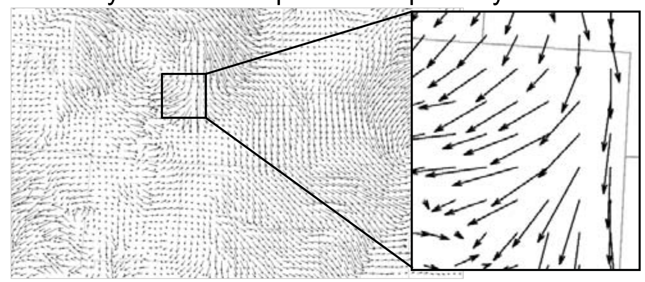
- ROW_MEAN_WDIR
  1. Derive wind directions and errors for each VL1L2 line
  2. Compute mean of errors



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Wind Direction: Suggestions

- When aggregating, wind directions can cancel out.
  - Verify over regions with unimodal wind direction.
  - Verify u and v components separately.



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

MET Config File GUI

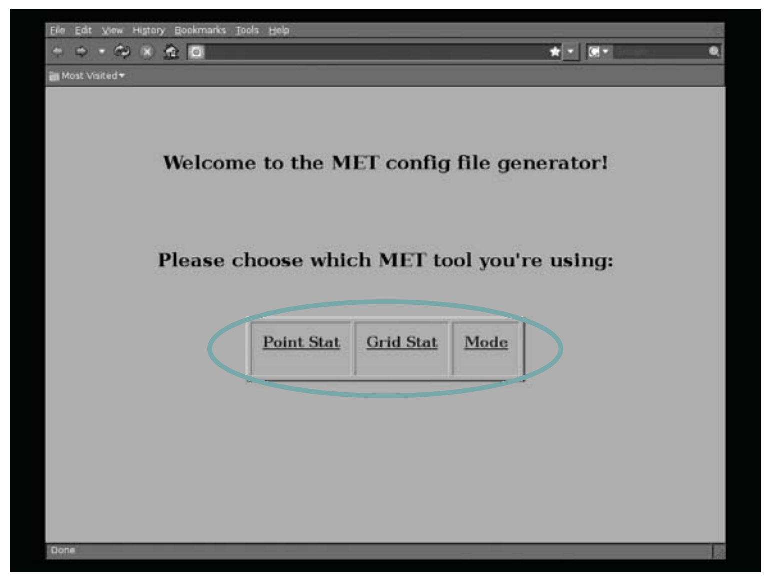
- Now a web GUI exists to assist in creation of some of these files:

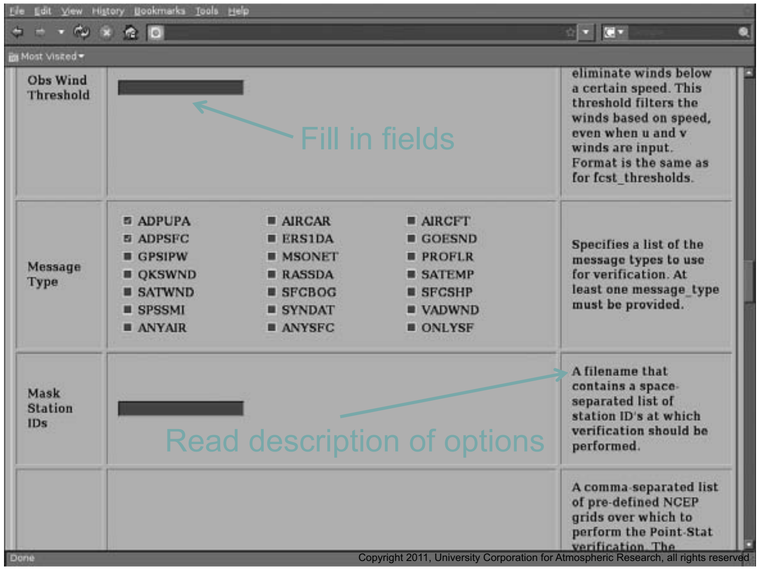
http://verif.rap.ucar.edu/cgi/metgui/base.cgi

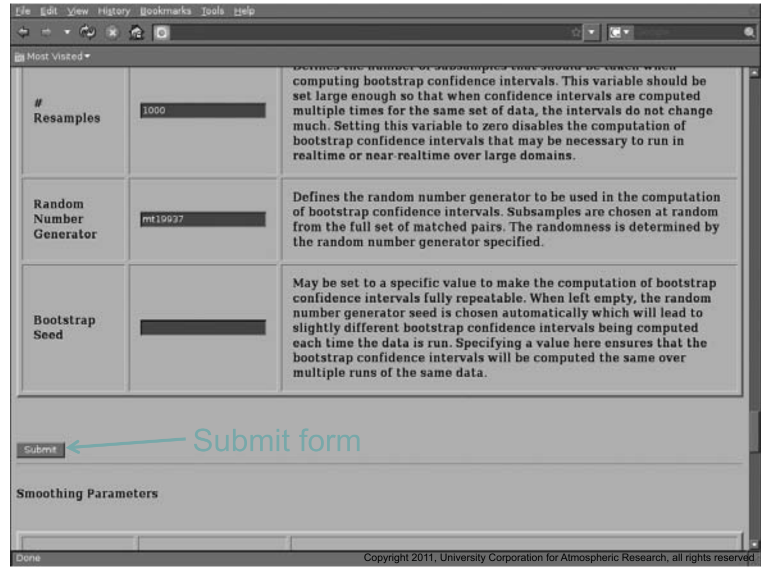
- All fields are described.
- Fill in blank fields.
- Submit to get ASCII config file, then save from browser to your machine.

```
////////////////////////////////////////////////////////////////////
//
// Point Stat Config File
//
//   Generated by MET Config File Web Utility
//
//   June 23, 2009   12:33 pm  MDT
//
////////////////////////////////////////////////////////////////////

model = "WRF";

beg_ds = -5400;

end_ds = -5400;

fcst_field [] = [ "TMP/P750-900", "UGRD/Z10", "VGRD/Z10" ];

obs_field [] = [];

fcst_thresh [] = [ "lt273 ge273", "ge5", "ge5" ];

obs_thresh [] = [];

fcst_wind_thresh [] = [ "NA" ];

obs_wind_thresh [] = [];

message_type [] = [ "ADPUPA", "ADPSFC" ];

mask_grid [] = [ "DTC165", "DTC166" ];

mask_poly [] = [ "MET_BASE/data/poly/LMV.poly" ];

mask_sid = "";

ci_alpha [] = [ 0.10, 0.05 ];

boot_interval = 1;

boot_rep_prop = 1.00;

n_boot_rep = 1000;

boot_rng = "mt19937";
```

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved








```
////////////////////////////////////////////////////////////////////////
//
// Point Stat Config File
//
//   Generated by MET Config File Web Utility
//
//   June 23, 2009   12:33 pm  MDT
//
////////////////////////////////////////////////////////////////////////

model = "WRF";

beg_ds = -5400;

end_ds = -5400;

fcst_field [] = [ "TMP/P750-900", "UGRD/Z10", "VGRD/Z10" ];

obs_field [] = [];

fcst_thresh [] = [ "lt273 ge273", "ge5", "ge5" ];

obs_thresh [] = [];

fcst_wind_thresh [] = [ "NA" ];

obs_wind_thresh [] = [];

message_type [] = [ "ADPUPA", "ADPSFC" ];

mask_grid [] = [ "DTC165", "DTC166" ];

mask_poly [] = [ "MET_BASE/data/poly/LMV.poly" ];

mask_sid = "";

ci_alpha [] = [ 0.10, 0.05 ];

boot_interval = 1;

boot_rep_prop = 1.00;

n_boot_rep = 1000;
```

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Stat-Analysis Tool

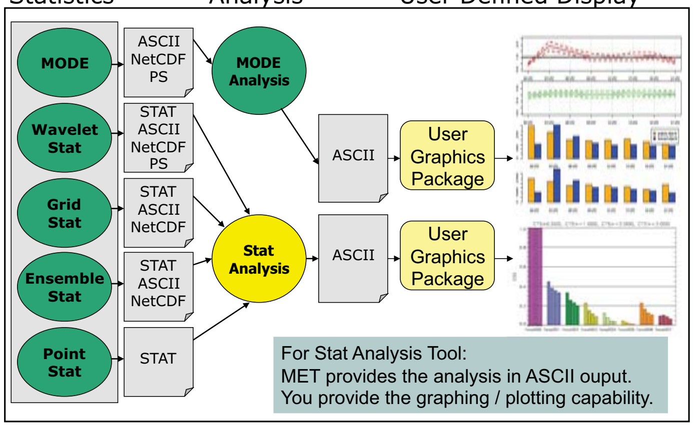
- Filtering
- Summarizing
- Aggregating

of Grid-Stat, Point-Stat, Ensemble-Stat & Wavelet-Stat output

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Stat Analysis Tool

Statistics Analysis User Defined Display



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Stat Analysis Jobs

- **Filtering**
  - **filter -** filters out lines from one or more stat files based on user-specified filtering options.
- **Summarizing**
  - **summary -** produces summary information from a single column of data including:

    mean, standard deviation, min, max, and the 10th, 25th, 50th, 75th, and 90th percentiles.
- **Customized tool for AFWA**
  - **go_index -** computes the GO Index, a performance statistic used primarily by the United States Air Force.

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Stat Analysis Jobs

- **Aggregation**
  - **aggregate -** aggregates stat data across multiple time steps or masking regions. Output line type is the same as input line type.
  - **aggregate_stat –** aggregates across multiple times/ regions then calculates statistics. Output line is different from input line types.

Valid line type combinations include:
-line_type FHO, CTC, -out_line_type CTS
-line_type MCTC, -out_line_type MCTS
-line_type SL1L2, SAL1L2, -out_line_type CNT
-line_type VL1L2, VAL1L2, -out_line_type WDIR (wind direction)
-line_type PCT, -out_line_type PSTD, PJC, PRC
-line_type NBRCTC, -out_line_type NBRCTS
-line_type MPR, -out_line_type FHO, CTC, CTS, MCTC, MCTS, CNT, SL1L2, SAL1L2,PCT, PSTD, PJC, PRC

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Stat Analysis Tool: Usage

Usage: stat_analysis
-lookin path
[-out filename]
[-tmp_dir path]
[-v level]
-config config_file
or –job at command line options with associated arguments
[filter]
[summary]
[aggregate]
[aggregate_stat]
[go_index]

-lookin	**Path to *.stat files – this can be a directory or a single file name (Use one or more times)**
-out	**Output name for ASCII file**
-tmp_dir	**Folder for temporary files**
-v	**Level of logging**
-config	**StatAnalysisConfig file**
filter	**See previous 2 slides**
summary	**See previous 2 slides**
aggregate	**See previous 2 slides**
aggregate_stat	**See previous 2 slides**
go_index	**See previous 2 slides**

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Stat Analysis Tool: Configuration

- 37 configurable parameters   only set a few:
  - Apply NAM G212 mask
    - **vx_mask[ ] = [“G212”];**
  - Using only the Temperature variable
    - **fcst_var[ ] = [“TMP”];**
      obs_var[ ] = [];
  - Filter on CTC lines in which fcst_var[ ] > 278 K
    - **line_type[ ] = [“CTC"];**
    - **fcst_thresh[ ] = [“>278”];**
      obs_thresh[ ] = [];
  - Dump the filtered stat data to a file
    AND sum contingency table count (CTC) lines of data
    for pressure levels between 850 and 750
    - **jobs[ ] = [*"-job filter -dump_row out/filter_job.stat", \
      "-job aggregate   line_type CTC \
      -dump_row out/aggr_ctc_job.stat   fcst_lev P850-750"*];**

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Stat Analysis Tool: Run –job aggregate

***"-job aggregate –line type CTC -dump row out/aggr ctc job.stat \
-fcst lev P850-750"***

Point Stat Output ***(i.e. pointstat out.stat)***

```
V2.0    WRF     ADPUPA G212   TMP
   P850-750   >278.00 CTC
   401     192     11     24     174
   UW_MEAN 1
```

		OBS		
FCST		Y	N	
	Y	192	11	203
	N	24	174	198
		216	185	401

```
V2.0    WRF     ADPSFC G212   TMP
   P850-750   >278.00 CTC
   167      25     23     0     119
   UW_MEAN  1
```

		OBS		
FCST		Y	N	
	Y	25	23	48
	N	0	119	119
		25	142	167

(NOTE: header modified to show only pertinent info)

Copyright 2011, Univers ty Corporation for Atmospheric Research, all rights reserved

Stat Analysis Tool: Run –job aggr

Stat Analysis Output ***(i.e. aggr ctc job.stat)***

```
FILTER:  -job filter
   -vx_mask G212 -line_type CTC
   -fcst_thresh >278.000 -var TMP
   -dump_row out/filter_job.stat

JOB_LIST: -job aggregate
-vx_mask G212 -line_type CTC
   -fcst_thresh >278.000 -var TMP
   -level P850-750 -dump_row out/
   aggr_ctc_job.stat
COL_NAME:          TOTAL
   FY_OY             FY_ON
   FN_OY           FN_ON
   INTERP_MTHD       INTERP_PNTS
CTC:               568             217
   34                24           293

    9999                         9999
```

		OBS		
FCST		Y	N	
	Y	217	34	251
	N	24	293	317
		241	327	568

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Stat Analysis Tool: Run –job summary

***"-job summary –line type CNT -fcst var TMP \
–dump row out/job summary RMSE.stat –column RMSE"***

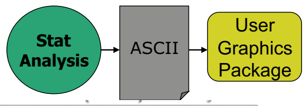
Column Number	Description
1	Summary (job type)
2	Total
3-7	Mean* *Includes normal and bootstrap upper and lower confidence limits*
8-10	Standard deviation** *Includes bootstrap upper and lower confidence limits*
11	Minimum value
12	10^{th} percentile
13	25^{th} percentile
14	Median (50^{th} percentile)
15	75^{th} percentile
16	90^{th} percentile
17	Maximum value

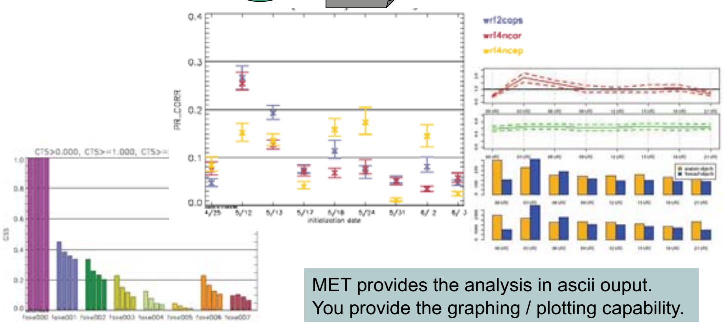
(stat analysis.out cont.)

```
JOB_LIST: -job summary -
line_type CNT  .
COL_NAME: TOTAL MEAN
MEAN_NCL MEAN_NCU MEAN_BCL
MEAN_BCU STDEV STDEV_BCL
STDEV_BCU MIN  P10
P25 P50 P75
P90 MAX
SUMMARY: 4  1.98438
1.33219 2.63656  1.58837
2.29289 0.40986 0.04574
0.55950 1.41291 1.59671
1.87241 2.07130 2.18328
2.18328 2.30251
```

Copyright 2011, Univers ty Corporation for Atmospheric Research, all rights reserved

Use your favorite plotting software

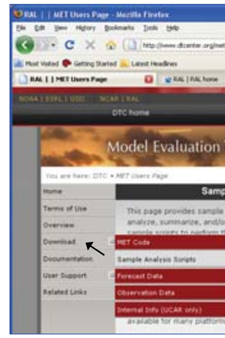


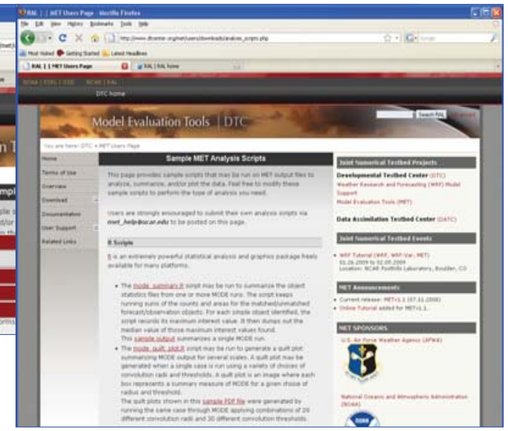


MET provides the analysis in ascii ouput.
You provide the graphing / plotting capability.

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

User Contributed Plotting Scripts





Please feel free to send your contributions to met help@ucar.edu

Copyright 2011, Univers ty Corporation for Atmospheric Research, all rights reserved

Interpretation of Output – Grid and Point Stat

Tressa L. Fowler

Point and Grid Stat Output Lines

- CTC - Contingency Table Counts
- CTS - Contingency Table Statistics
- CNT - Continuous Statistics
- FHO - Forecast, Hit, Observation Rates
- PCT - Contingency Table counts for Probabilistic forecasts
- PSTD Contingency table Statistics for Probabilistic forecasts with Dichotomous outcomes
- PJC - Joint and Conditional factorization for Probabilistic forecasts
- PRC - Receiver Operating Characteristic for Probabilistic forecasts
- SL1L2 - Scalar L1L2 Partial Sums
- SAL1L2 - Scalar Anomaly L1L2 Partial Sums when climatological data is supplied
- VL1L2 - Vector L1L2 Partial Sums
- VAL1L2 - Vector Anomaly L1L2 Partial Sums when climatological data is supplied
- MPR - Matched Pair data

Copyright 2011, UCAR, all rights reserved

Point and Grid stat output example

Header Line:

```
TOTAL BASER   BASER_NCL BASER_NCU BASER_BCL BASER_BCU FMEAN
FMEAN_NCL FMEAN_NCU FMEAN_BCL FMEAN_BCU ACC     ACC_NCL
ACC_NCU ACC_BCL ACC_BCU FBIAS   FBIAS_BCL FBIAS_BCU PODY
PODY_NCL PODY_NCU PODY_BCL PODY_BCU PODN PODN_NCL PODN_NCU
PODN_BCL PODN_BCU POFD POFD_NCL POFD_NCU POFD_BCL POFD_BCU FAR
FAR_NCL FAR_NCU FAR_BCL FAR_BCU CSI     CSI_NCL CSI_NCU CSI_BCL
CSI_BCU GSS     GSS_BCL GSS_BCU HK HK_NCL HK_NCU HK_BCL HK_BCU HSS
HSS_BCL HSS_BCU ODDS ODDS_NCL ODDS_NCU ODDS_BCL ODDS_BCU
```

Data Line:

```
CTS     5     1.00000 0.56552   1.00000   NA       NA       0.40000 0.11762   0.76928   NA
NA       0.40000 0.11762 0.76928 NA      NA      0.40000 NA       NA       0.40000 0.11762
0.76928  NA       NA       NA   NA       NA       NA       NA       NA   NA       NA       NA       NA
0.00000 0.00000 0.43448 NA       NA       0.40000 0.11762 0.76928 NA       NA       0.00000 NA
NA      NA NA      NA      NA      NA      0.00000 NA       NA       NA   NA       NA       NA       NA
```

Copyright 2011, UCAR, all rights reserved

Example Contingency Table Statistics

See Appendix C of MET documentation for equations and details

- TOTAL – Count of total pairs included in the stats for this line
- BASER – base rate (e.g. sample climatological rate of event)
- ACC - accuracy
- FBIAS – frequency bias
- PODY – probability of detection (events)
- PODN – probability of detection (non-events)
- POFD – probability of false detection
- FAR – False Alarm Ratio
- CSI – Critical success index
- GSS – Gilbert Skill Score
- HK - Hanssen-Kuipers Discriminant
- HSS - Heidke Skill Score
- ODDS – Odds Ratio
- Etc., etc., etc. . .

Confidence Intervals for some measures are included with the following suffixes:
_NCL = Normal Confidence Lower
_NCU = Normal Confidence Upper
_BCL = Bootstrap Confidence Lower
_BCU = Bootstrap Confidence Upper

Copyright 2011, UCAR, all rights reserved

Example Continuous Statistics

See Appendix C of MET documentation for equations and details

Confidence Intervals for some measures are included with the following suffixes:
_NCL = Normal Confidence Lower
_NCU = Normal Confidence Upper
_BCL = Bootstrap Confidence Lower
_BCU = Bootstrap Confidence Upper

- FBAR – Forecast average
- OBAR– Observation average
- ME – Mean Error
- ESTDEV– Estimated standard deviation of the error
- MAE – Mean absolute error
- MSE – Mean squared error
- BCMSE – Bias corrected mean squared error
- RMSE – Root mean squared error
- E50 – Median (50% percentile) of the error.
- Etc., etc., etc. . .

Copyright 2011, UCAR, all rights reserved

Example Probability Forecast Statistics

See Appendix C of MET documentation for equations and details

- RELIABILITY – conditional bias
- RESOLUTION – discrimination ability of forecast
- UNCERTAINTY – variability of the observations
- ROC AUC - Area under the receiver operating characteristic curve
- BRIER – Brier Score
- CALIBRATION i - conditional probability of an event given each probability forecast
- REFINEMENT i – probability of each forecast category
- LIKELIHOOD i - conditional probability for each forecast category given the event
- BASER i - probability of an event for each forecast category

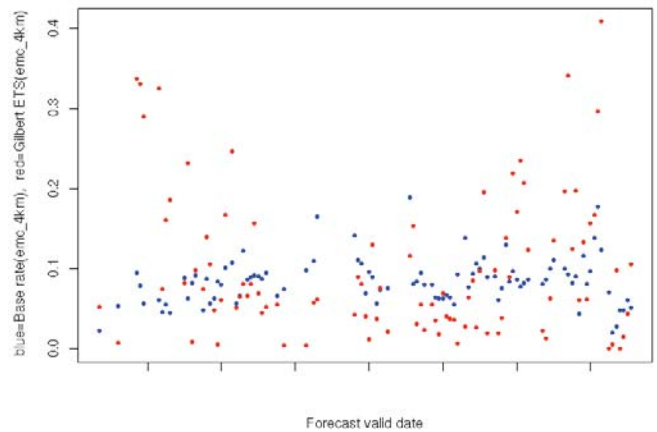
Copyright 2011, UCAR, all rights reserved

Example of Appendix C info

- **Accuracy -** for a 2x2 contingency table is defined as $\frac{n_{11} + n_{00}}{T}$
- That is, it is the proportion of forecasts that were either hits or correct rejections – the fraction that were correct.
- Accuracy ranges from 0 to 1; a perfect forecast would have an accuracy value of 1.
- Accuracy should be used with caution, especially for rare events, because it can be strongly influenced by large values of n_{00}.

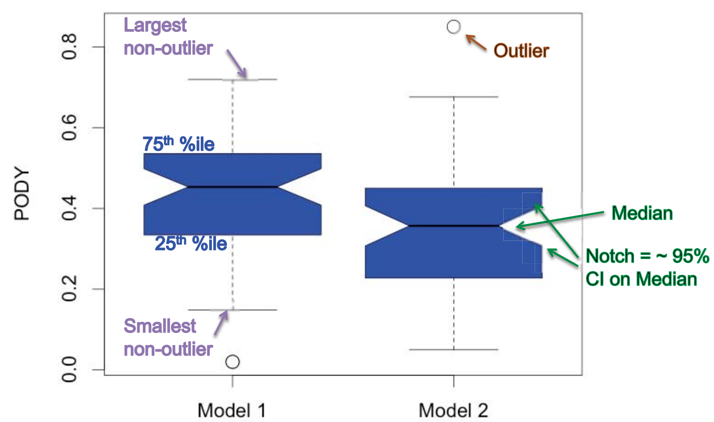
Copyright 2011, UCAR, all rights reserved

Example Scatter Plot



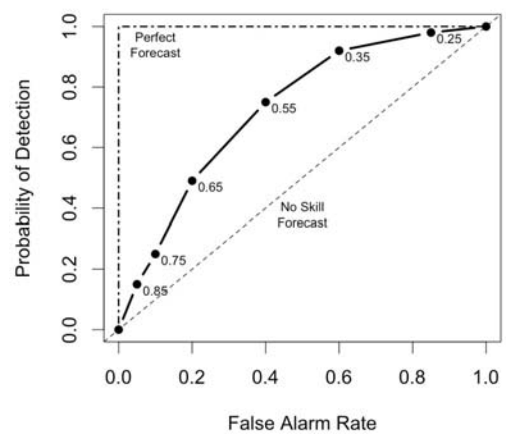
Copyright 2011, UCAR, all rights reserved

Example Box (and Whisker) Plot



Copyright 2011, UCAR, all rights reserved

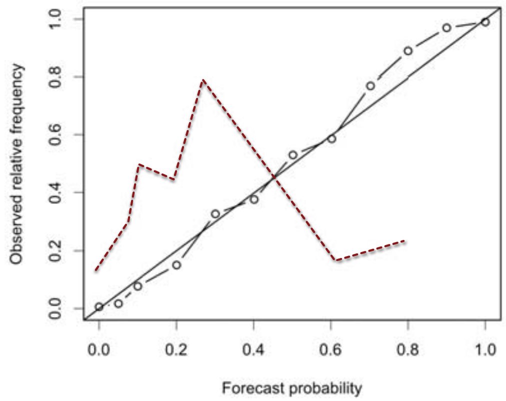
Example Receiver Operating Characteristic Plot



Create with points from PRC line type.

Copyright 2011, UCAR, all rights reserved

Example Reliability Diagram



A measure of conditional bias.

Do the forecast and observed probabilities match?

One to one line (solid) shows perfect reliability.

Forecast 1 (the line with dots) has good reliability.

Forecast 2 (the dashed line) has very poor reliability.

Information from PJC line type.

Copyright 2011, UCAR, all rights reserved

Customizing Point-Stat and Grid-Stat Output

We'll restrict this discussion to an explanation of masking and interpolation.

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

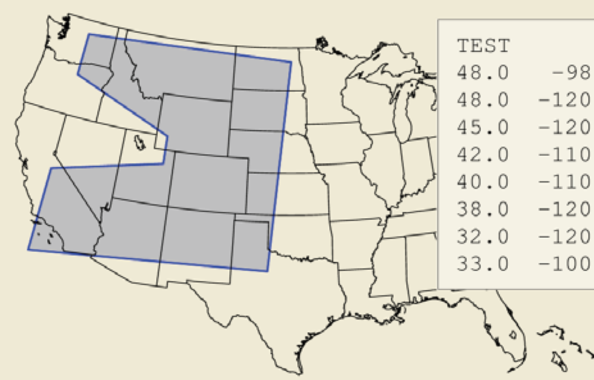
Masking Methods

	Polyline	Grid	Stations	On/Off Bit Map
Point Stat	✓	✓	✓	✓
Grid Stat	✓	✓	N/A	✓

Masking by stations doesn't apply to Grid Stat.

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

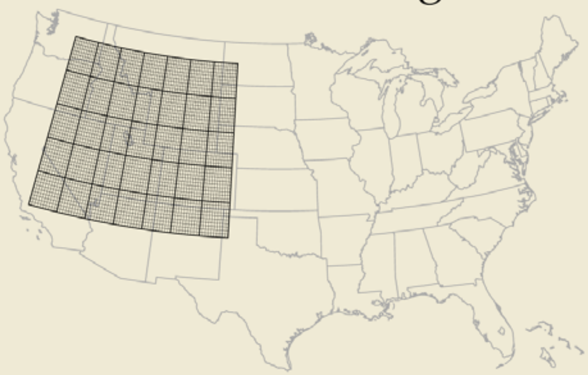
Polyline Masking



```
TEST
48.0   -98.0
48.0  -120.0
45.0  -120.0
42.0  -110.0
40.0  -110.0
38.0  -120.0
32.0  -120.0
33.0  -100.0
```

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Grid Masking



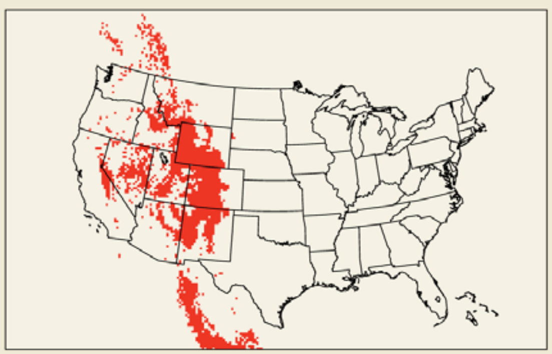
NCEP grids link: www.nco.ncep.noaa.gov/pmb/docs/on388/tableb.html

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Station Masking

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Data Threshold Mask



Topography > 6000 feet

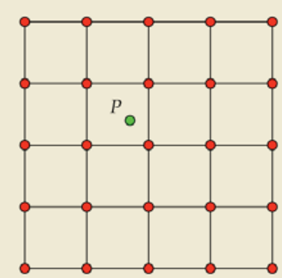
© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Interpolation

Need to Choose:

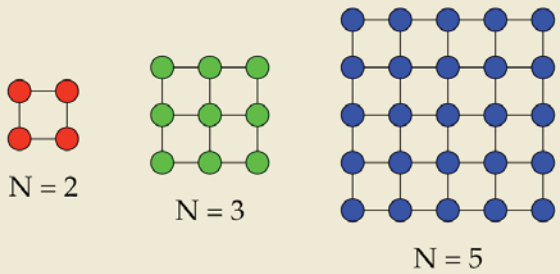
(1) Method

(2) Width



© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Interpolation Widths



© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Interpolation Methods

	Min	Max	Median	UW Mean	DW Mean	Nearest Nbr	Least Squares
Point Stat	✓	✓	✓	✓	✓	✓	✓
Grid Stat	✓	✓	✓	✓	N/A	N/A	N/A

For Grid Stat, these are smoothing methods.

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Min, Max, Median

Takes minimum, maximum or median of values in interpolation square.

Median separates the upper half of data values from the lower half. This is different from the mean, which is an average.

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Nearest Neighbor

Essentially, no interpolation is performed.

Value at interpolation point is simply the data value at the closest grid point.

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Unweighted Mean
Distance-Weighted Mean

Unweighted Mean is the average.

Distance-Weighted Mean is an average weighted according to distance from nearby grid points.

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

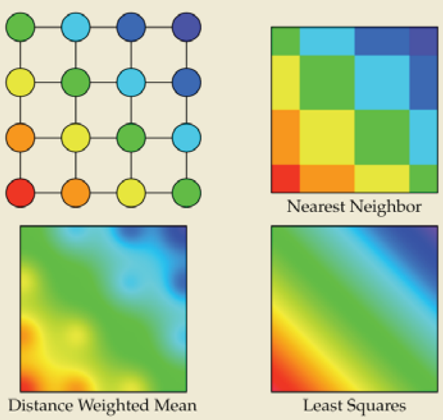
Least Squares

Performs a local Least-Squares linear fit in interpolation square.

$$z = Ax + By + C$$

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Interpolation Examples



© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

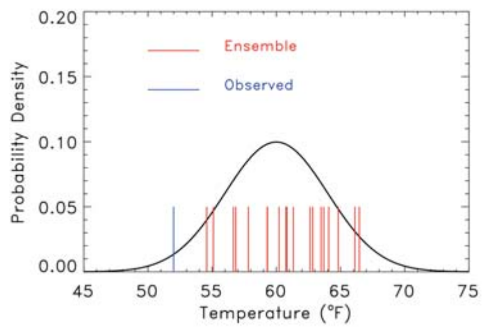
Verification of ensembles

Tressa Fowler

Thanks to Tom Hamill and Laurie Wilson whose slides I have stolen without permission.

Copyright UCAR 2011, all rights reserved.

How good is this ensemble forecast?



Copyright UCAR 2011, all rights reserved.

Questions to ask before beginning?

- How were the ensembles constructed?
  - Poor man's ensemble (distinct members)
  - Multi-physics (distinct members)
  - Random perturbation of initial conditions (anonymous members)
- How are your forecasts used?
  - Improved point forecast (ensemble mean)
  - Probability of an event
  - Full distribution

Copyright UCAR 2011, all rights reserved.

Verifying a probabilistic forecast

- You cannot verify a probabilistic forecast with a single observation.
- The more data you have for verification, (as with other statistics) the more certain you are.
- Rare events (low probability) require more data to verify.
- These comments refer to probabilistic forecasts developed by methods other than ensembles as well.

Copyright UCAR 2011, all rights reserved.

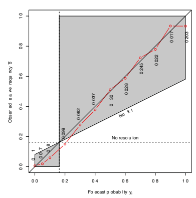
Evaluate each member as a separate, deterministic forecast

- Why? Because it is easy and important
    - If members are unique, it might provide useful diagnostics.
    - If members are bias, verification statistics might be skewed.
    - If members have different levels of bias, should you calibrate?
  - Do these results conform to your understanding of how the ensemble members were created?

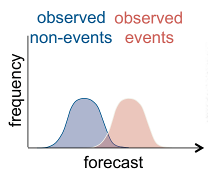
Copyright UCAR 2011, all rights reserved.

Properties of a perfect probabilistic forecast of a binary event.

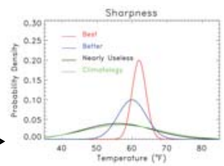
Reliability



Resolution

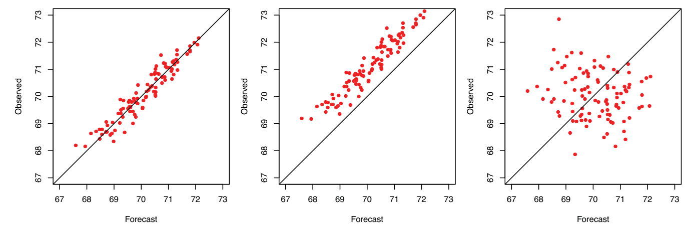


Sharpness



Copyright UCAR 2011, all rights reserved.

Our friend, the scatterplot



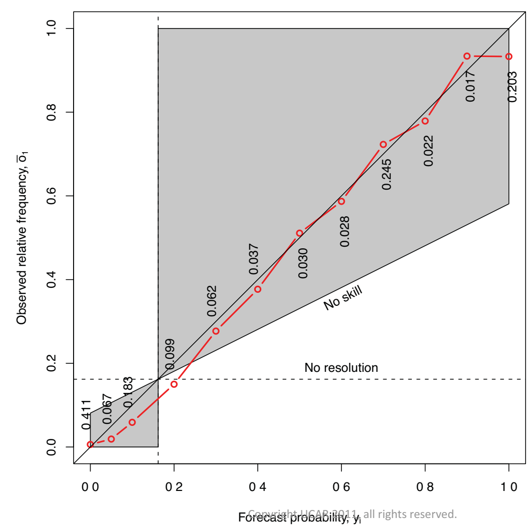
Copyright UCAR 2011, all rights reserved.

Introducing the attribute diagram!

(close relative to the reliability 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.

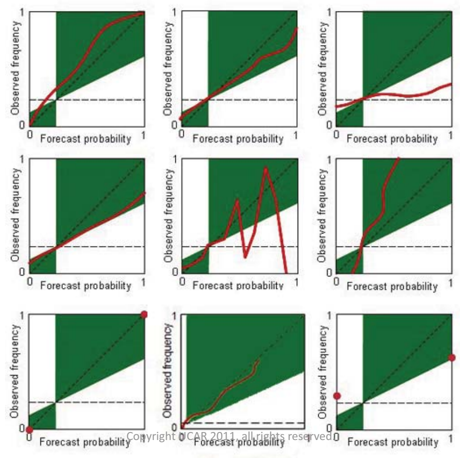
Copyright UCAR 2011, all rights reserved.



Attribute Diagram

Copyright UCAR 2011, all rights reserved.

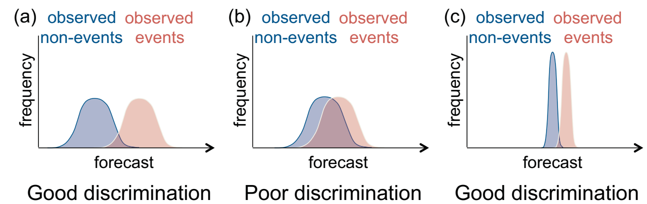
Reliability Diagram Exercise



Copyright UCAR 2011, all rights reserved.

Discrimination

- *Discrimination*: The ability of the forecast system to clearly distinguish situations leading to the occurrence of an event of interest from those leading to the non-occurrence of the event.
- Depends on:
  - Separation of means of conditional distributions
  - Variance within conditional distributions



Copyright UCAR 2011, all rights reserved.

Brier score

- Appropriate for probabilities of a binary event occurring.
- Analogous to mean square error.
- Can be decomposed to resolution, reliability and uncertainty components.
- Geometrically relates to attribute diagram.
- Fair comparisons require common sample climatology.

Copyright UCAR 2011, all rights reserved.

The Brier Score

- Mean square error of a probability forecast

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2$$

- Weights larger errors more than smaller ones

Copyright UCAR 2011, all rights reserved.

Components of probability error

The Brier score can be decomposed into 3 terms (for K probability classes and a sample of size N):

$$BS = \frac{1}{N}\sum_{k=1}^{K} n_k (p_k - \bar{o}_k)^2 \; - \; \frac{1}{N}\sum_{k=1}^{K} n_k (\bar{o}_k - \bar{o})^2 \; + \; \bar{o}(1-\bar{o})$$

reliability

If for all occasions when forecast probability p_k is predicted, the observed frequency of the event is $\bar{o}_k = p_k$ then the forecast is said to be reliable. Similar to bias for a continuous variable

resolution

The ability of the forecast to distinguish situations with distinctly different frequencies of occurrence.

uncertainty

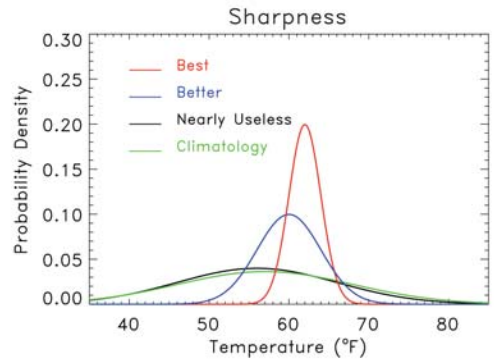
The variability of the observations. Maximized when the climatological frequency (*base rate*) =0.5

Has nothing to do with forecast quality! Use the Brier skill score to overcome this problem.

The presence of the uncertainty term means that Brier Scores should not be compared on different samples.

Copyright UCAR 2011, all rights reserved.

Sharpness also important



"**Sharpness**" measures the specificity of the probabilistic forecast. Given two reliable forecast systems, the one producing the sharper forecasts is preferable.

But: don't want sharp if not reliable. Implies unrealistic confidence.

Copyright UCAR 2011, all rights reserved.

Sharpness ≠ resolution

- Sharpness is *a property of the forecasts alone*; a measure of sharpness in Brier score decomposition would be how populated the extreme N_i*'s are* is.

$$\mathrm{BS} = \frac{1}{n}\sum_{i\;1}^{I} N_i (y_i - \bar{o}_i)^2 - \frac{1}{n}\sum_{i\;1}^{I} N_i (\bar{o}_i - \bar{o})^2 + \bar{o}(1-\bar{o})$$

("reliability") ("resolution") ("uncertainty")

Copyright UCAR 2011, all rights reserved.

Forecasts of a full distribution

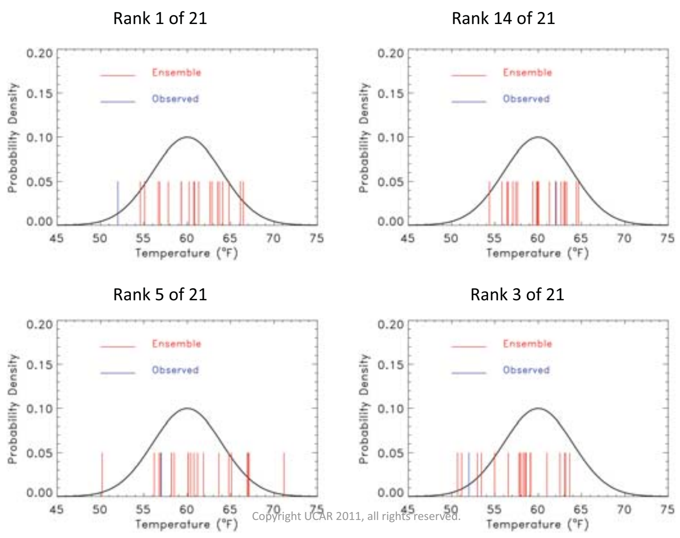
- How is it expressed?
  - Discretely by providing forecasts from all ensemble members
  - A parametric distribution normal(ensemble mean, spread)
  - Smoothed function kernel smoother

Copyright UCAR 2011, all rights reserved.

Assuming the forecast is reliable (calibrated)

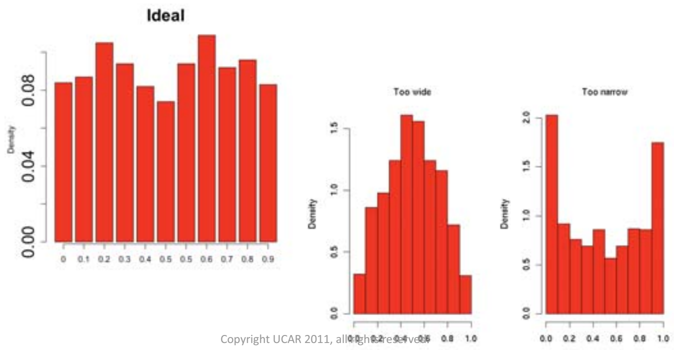
- By default, we assume all ensemble forecasts have the same number of members. Comparing forecasts with different number of members is an advanced topic.
- For a perfect ensemble, the observation comes from the same distribution as the ensemble.
- Huh?

Copyright UCAR 2011, all rights reserved.



Copyright UCAR 2011, all rights reserved.

Rank Histograms



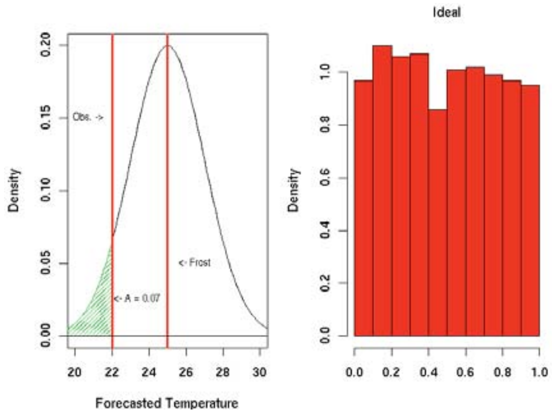
Copyright UCAR 2011, all rights reserved.

Verifying a continuous expression of a distribution (i.e. normal, Poisson, beta)

- Probability of any observation occurring is on [0,1] interval.
- Probability Integral Transformed (PIT) - fancy word for how likely is a given forecast
- Still create a rank histogram using bins of probability of observed events.

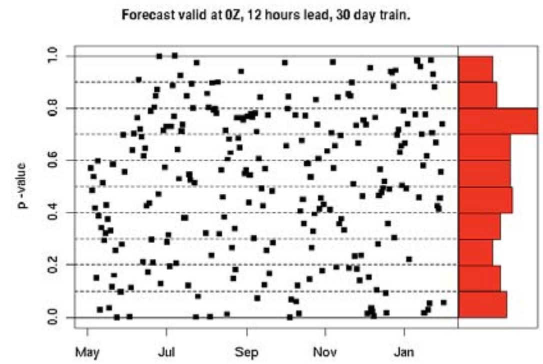
Copyright UCAR 2011, all rights reserved.

Verifying a distribution forecast



Copyright UCAR 2011, all rights reserved.

Evaluate order of probabilistic forecasts



Copyright UCAR 2011, all rights reserved.

Warnings about rank histograms

- Assume all samples come from the same climotology!
- A flat rank histogram can be derived by combining forecasts with offsetting biases
- See Hamill, T. M., and J. Juras, 2006: Measuring forecast skill: is it real skill or is it the varying climatology? *Quart. J. Royal Meteor. Soc.*, Jan 2007 issue
- Techniques exist for evaluating "flatness", but they mostly require much data.

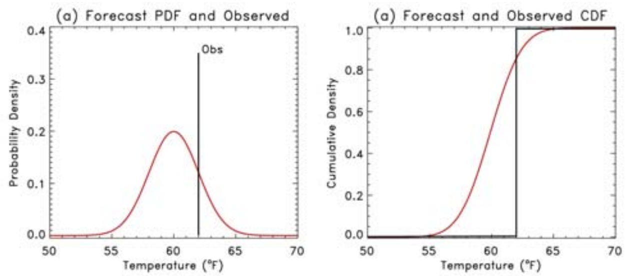
Copyright UCAR 2011, all rights reserved.

Continuous and discrete rank probability scores.

- Introduce pdf -> cdf High, low, no variance (event)
- Area of wide, narrow
- Perfect forecast with bias …
- Aggregate
- Relates to Brier score – for a forecast of a binary event, the RPS score is equivalent to the Brier score.

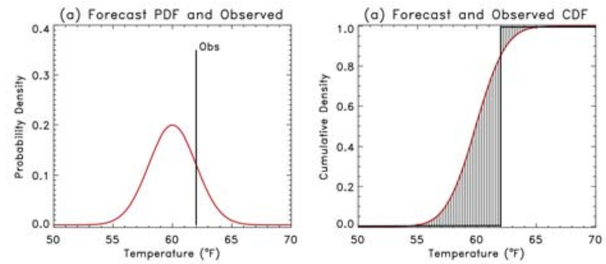
Copyright UCAR 2011, all rights reserved.

Rank Probability Scores



Copyright UCAR 2011, all rights reserved.

A good RPS score minimizes area



Copyright UCAR 2011, all rights reserved.

Final comments

- Know how and why your ensemble is being created.
- Use a combination of graphics and scores.
- Areas of more research
  - Verification of spatial forecasts

Copyright UCAR 2011, all rights reserved.

Useful references

- **Good overall references** for forecast verification:
  - (1): Wilks, D.S., 2006: *Statistical Methods in the Atmospheric Sciences (2nd Ed).* Academic Press, 627 pp.
  - (2) Beth Ebert's forecast verification web page, http://tinyurl.com/y97c74
- **Rank histograms**: Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550-560.
- **Spread-skill relationships**: Whitaker, J.S., and A. F. Loughe, 1998: The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev.*, **126**, 3292-3302.
- **Brier score, continuous ranked probability score, reliability diagrams**: Wilks text again.
- **Relative operating characteristic**: Harvey, L. O., Jr, and others, 1992: The application of signal detection theory to weather forecasting behavior. *Mon. Wea. Rev.*, **120**, 863-883.
- **Economic value diagrams**:
  - (1)Richardson, D. S., 2000: Skill and relative economic value of the ECMWF ensemble prediction system. *Quart. J. Royal Meteor. Soc.*, **126**, 649-667.
  - (2) Zhu, Y, and others, 2002: The economic value of ensemble-based weather forecasts. *Bull. Amer. Meteor. Soc.*, **83**, 73-83.
- **Overforecasting skill**: Hamill, T. M., and J. Juras, 2006: Measuring forecast skill: is it real skill or is it the varying climatology? *Quart. J. Royal Meteor. Soc.*, Jan 2007 issue. http://tinyurl.com/kxtct

Copyright UCAR 2011, all rights reserved.

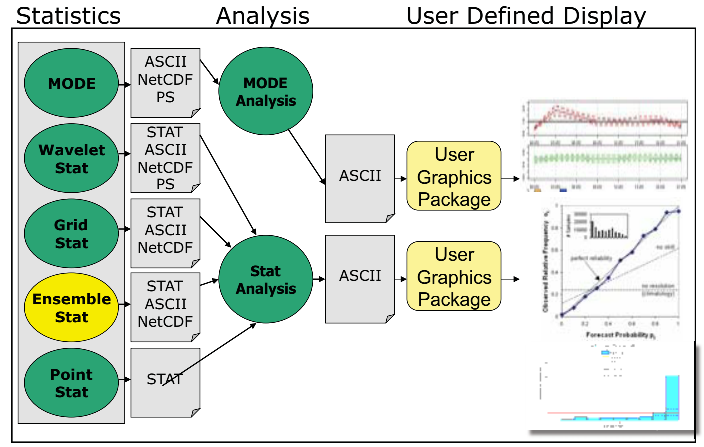
Verifying Ensembles with MET

- Ensemble Stat Tool
  - Ensemble Mean Fields
  - Probability Fields
  - Rank Histograms
- Point Stat and Grid Stat Tool
  - Brier Score + Decomposition
  - Reliability Diagrams
  - Receiver Operating Characteristic Diagram + Area Under the Curve
  - Joint/Conditional factorization table

Presenter: Tara Jensen

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Ensemble Stat Tool



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Ensemble Stat Capabilities

Calculates:

- Ensemble Mean field
- Ensemble Standard Deviation field
- Ensemble Mean 1 Standard Deviation field
- Ensemble Mean + 1 Standard Deviation field
- Ensemble Minimum field
- Ensemble Maximum field
- Ensemble Range field
- Ensemble Valid Data Count field
- Ensemble Relative Frequency by threshold fields
- Ranked Histograms (if Obs Field Provided)

Writes:

- Gridded field of Observation Ranks to a NetCDF file
- Stat file with Rank Histogram and Ensemble information
- Observation Rank Matched Pairs

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Ensemble Stat Tool: Usage

Usage: ensemble_stat

n_ens ens_file_1 \ ... ens_file_n | ens_file_list	Number of Ensemble members followed by list of ensemble member names OR ens file list (the name of an ASCII file with names of members)
config_file	Config file name
[-grid_obs file] [-point_obs file]	Name of gridded or point observed file – Required if Rank Histograms desired (optional)
[-ens_valid time] [-ens_lead time]	YYYYMMDD[HH[MMSS]] format sets the ensemble valid time or lead time to be used (optional)
[-obs_valid_beg time] [-obs_valid_end time]	YYYYMMDD[HH[MMSS]] format to set the beginning and end of the matching observation time window (optional)
[-obs_lead time]	HH[MMSS] format - observation lead time (opt.)
[-outdir path]	Set output directory (optional)
[-v level]	Set level of verbosity (optional)

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Ensemble Stat Tool: Configuration

- 23 configurable parameters   set only a few:
  - Set Ensemble Field to Precip for calculating Ensemble Mean, Std. Dev, Ens Relative Freq fields, etc..
    - **ens_field[ ] = [“61/A24”];**
  - Using only the following thresholds
    - **ens_thresh[ ] = [“gt0.0 ge12.7 ge25.4”];**
  - Set ratio of valid members (data values) to total ensemble members required for processing
    - **valid_ens_thresh = 0.5;**
    - **valid_data_thresh = 0.5;**
  - Set fcst_field and obs_field to Precip for calculating Rank Histograms
    - **fcst_field[ ] = [“61/A24”];**
    - **obs_field[ ] = [“ ”];**

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Ensemble Stat Tool: Run

***ensemble_stat ***
***6 sample_fcst/2009123112/*gep*/d01_2009123112_02400.grib ***
***config/EnsembleStatConfig ***
***-grid_obs sample_obs/ST4/ST4.2010010112.24h ***
***-point_obs out/ascii2nc/precip24_2010010112.nc ***
-outdir out/ensemble_stat -v 2

NOTE:
You can pass in a path with wildcards to pull out the files you would like to process or you can pass in an ASCII filename that contains a list of ensemble members

Gridded and Obs field are included for use in calculating Rank Histogram

Copyright 2011, Univers ty Corporation for Atmospheric Research, all rights reserved

Ensemble Stat Tool: Run

```
dakota:/d3/personal/jensen> test  ensemble  stat.sh
*** Running Ensemble-Stat on APCP using GRIB forecasts, point observations, and gridded observations
***
GSL  RNG  TYPE  mt19937
GSL  RNG  SEED  1814440524
Configuration File:
  EnsembleStatConfig  default
Ensemble Files[6]:
  data/sample  fcst/2009123112/arw-fer-gep1/d01  2009123112  02400.grib
  data/sample  fcst/2009123112/arw-fer-gep5/d01  2009123112  02400.grib
  data/sample  fcst/2009123112/arw-sch-gep2/d01  2009123112  02400.grib
  data/sample  fcst/2009123112/arw-sch-gep6/d01  2009123112  02400.grib
  data/sample  fcst/2009123112/arw-tom-gep3/d01  2009123112  02400.grib
  data/sample  fcst/2009123112/arw-tom-gep7/d01  2009123112  02400.grib
Gridded Observation Files[1]:
  data/sample  obs/ST4/ST4.2010010112.24h
Point Observation Files[1]:
  out/ascii2nc/precip24  2010010112.nc
--------------------------------------------------------------------------------
Processing ensemble field: APCP  24/A24
--------------------------------------------------------------------------------
…
Processing gridded verification APCP  24/A24 versus APCP  24/A24, for observation type MC  PCP, over
region FULL, for interpolation method UW  MEAN(1), using 15480 pairs.
--------------------------------------------------------------------------------
Output file: test/ensemble  stat  20100101  120000V.stat
Output file: test/ensemble  stat  20100101  120000V  rhist.txt
Output file: test/ensemble  stat  20100101  120000V  orank.txt
Output file: test/ensemble  stat  20100101  120000V  ens.nc
Output file: test/ensemble  stat  20100101  120000V  orank.nc
```

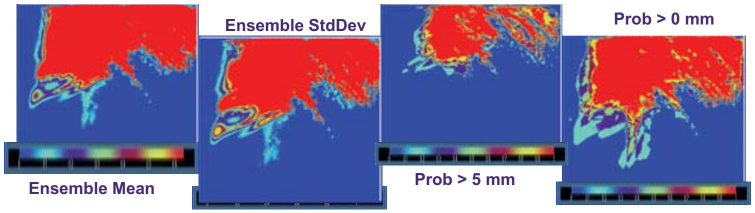
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Ensemble Stat Tool: Run

Output from   out/ensemble stat/ensemble stat 20100101 120000V rhist.txt

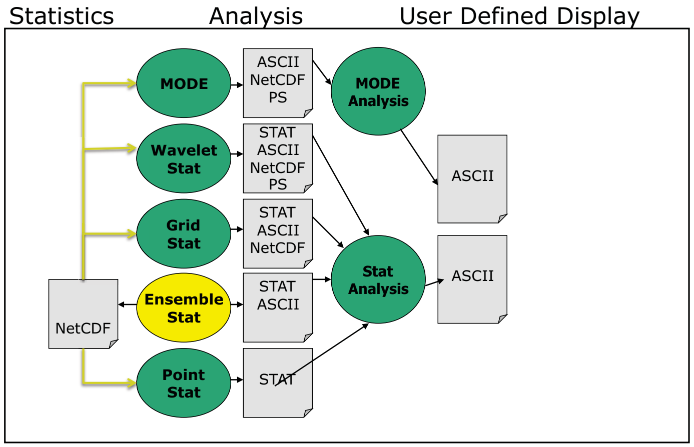
```
VERSION MODEL FCST  LEAD FCST  VALID  BEG  FCST  VALID  END
OBS  LEAD OBS  VALID  BEG   OBS  VALID  END   FCST  VAR FCST  LEV
OBS  VAR OBS  LEV OBTYPE VX  MASK INTERP  MTHD INTERP  PNTS
FCST  THRESH OBS  THRESH COV  THRESH ALPHA LINE  TYPE TOTAL
N  RANK RANK  1 RANK  2 RANK  3 RANK  4 RANK  5 RANK  6 RANK  7
V3.0    WRF   240000    20100101  120000 20100101  120000 240000
20100101  120000 20100101  120000 APCP  24  A24     APCP  24 A24    MC  PCP
FULL    UW  MEAN    1        NA        NA        NA        NA    RHIST     15480 7
4517   2133   1885   1681   1759   1519   1986
```

Output from   out/ensemble stat/ensemble stat 20100101 120000V ens.nc   *(using ncview)*



Copyright 2011, Univers ty Corporation for Atmospheric Research, all rights reserved

V3.0 Output from Ensemble Stat

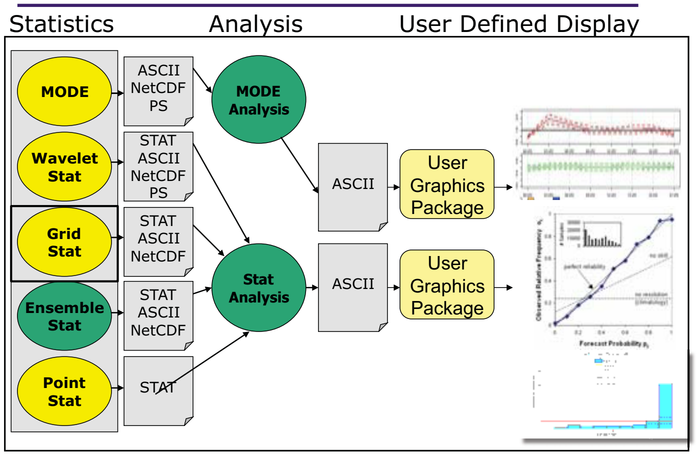


Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Verifying Probabilistic Fields and Ensemble Relative Frequency

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Probability Fields in Other Tools



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Grid Stat Probability Configuration

- Some of the key settings include:
  - Set fcst_field and obs_field to Precip for calculating Rank Histograms
    - **fcst field[] = [“APCP/A6/PROB", “APCP/A6/PROB", “APCP/A6/PROB"];**
    - **obs field[] = ["APCP/A6", "APCP/A6", "APCP/A6"];**
  - Set fcst_thresh to multiple “bins” for nx2 contingency table. *We have 3 probabilistic fields so bins must be set for each one.*
    - **fcst thresh[] = ["ge0.0 ge0.1 ge0.2 ge0.3 ge0.4 ge0.5 ge0.6 ge0.7 ge0.8 ge0.9 ge1.0", **
      **"ge0.0 ge0.1 ge0.2 ge0.3 ge0.4 ge0.5 ge0.6 ge0.7 ge0.8 ge0.9 ge1.0", **
      "ge0.0 ge0.1 ge0.2 ge0.3 ge0.4 ge0.5 ge0.6 ge0.7 ge0.8 ge0.9 ge1.0"];
  - Set obs_thresh to appropriate observed thresholds
    - **obs thresh[] = ["ge12.7", "ge25.4", "ge50.8"];**

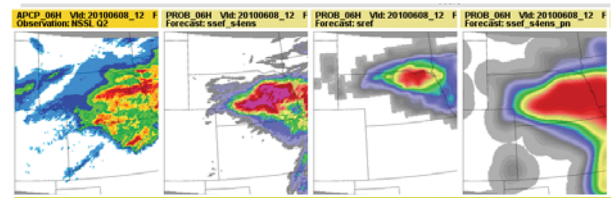
Copyright 2011, Univers ty Corporation for Atmospheric Research, all rights reserved

Grid Stat Probability: Run

- Make sure Probability Output flags set correctly in Grid Stat Config file (output flags 7 10)
- Run Grid Stat as previously described
- Output written to .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
  - PRC ROC Curve Points for Probabilistic Variables

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Grid Stat Probability: Example



Observed Precip Field

Ensemble Relative Frequency

4 km Simple

32 35 km Simple

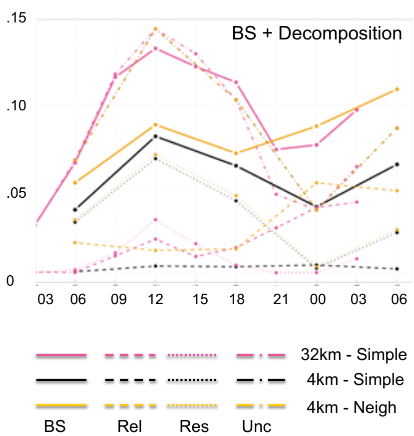
4 km Nbrhood Method Applied

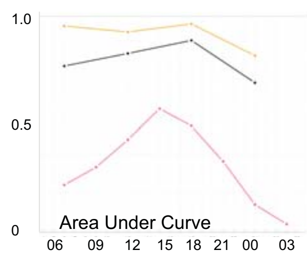
Relative frequency calculated from raw field

Gaussian smoother applied to each member before calculating relative frequency.

Copyright 2011, Univers ty Corporation for Atmospheric Research, all rights reserved

Grid Stat Probability: Example



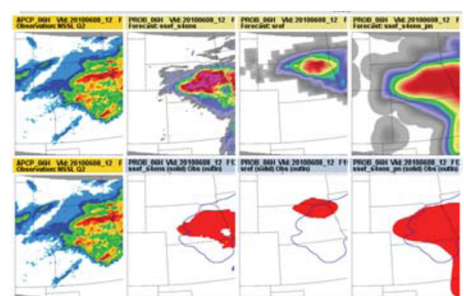


Optimal:
Brier Score ↓
Reliability ↓
Resolution ↑
Uncertainty *
Area Under Curve ↑

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

A teaser… Spatial Methods Application

You can use MODE on probability fields also…



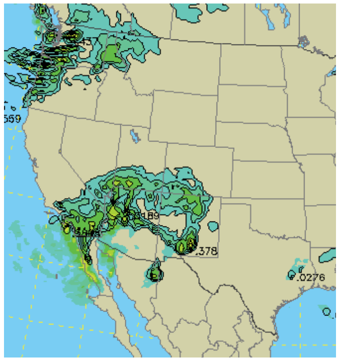
In this case: Probability field threshold = 50%
Observed field threshold > 12.7 mm (or 0.5")

Copyright 2011, Univers ty Corporation for Atmospheric Research, all rights reserved

Why Use Objects to Verify?

Tressa L. Fowler

Typical situation



Forecast

Observation

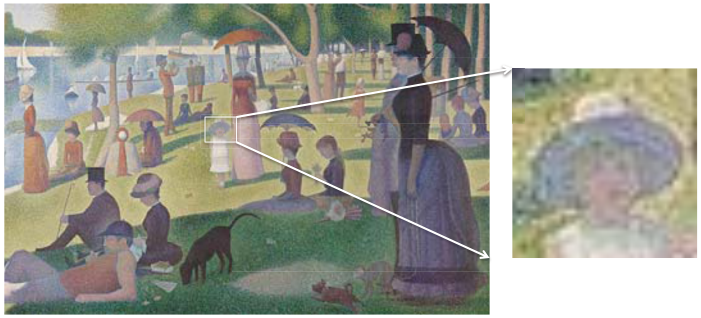
copyright 2011, UCAR, all rights reserved

Traditional verification matches up points, then sums them up.

Many forecasts are more than the sums of their parts.

copyright 2011, UCAR, all rights reserved

Pixels or Pictures?



copyright 2011, UCAR, all rights reserved

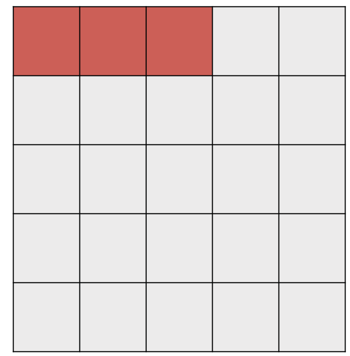
Object verification
is more like what
humans
do.

Objects recognize
the spatial relationship
between points.

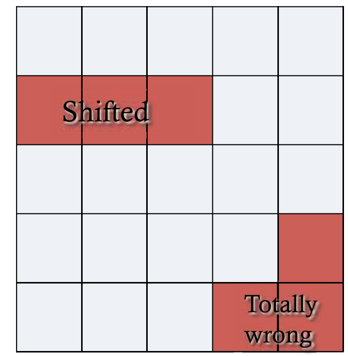
copyright 2011, UCAR, all rights reserved

Simple example

Observed



Forecast



copyright 2011, UCAR, all rights reserved

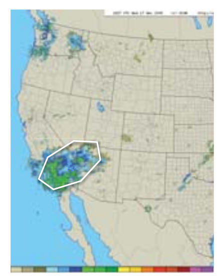
REAL - observed

Forecast 1 –
Distorted view of reality

Forecast 2 –
Another distorted
view of reality

copyright 2011, UCAR, all rights reserved

REAL - observed



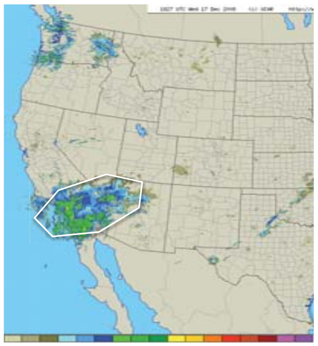
Forecast 1 –
Distorted view of reality

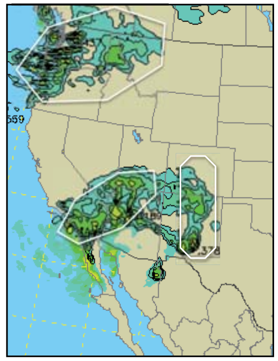
Forecast 2 –
Another distorted
view of reality

copyright 2011, UCAR, all rights reserved

We can compare attributes of forecast and observations even when they are not in the same place!

- Is the object in the right place? — Centroid distance = 25
- Does the size of these objects match? — Area ratio = 85%
- Is the intensity within the objects similar? — 50 dBZ vs 40 dBZ





copyright 2011, UCAR, all rights reserved

This is not really a new idea . . .

Analytic cubists "analyzed" natural forms and reduced the forms into basic geometric parts on the two-dimensional picture plane.

Analytic cubism was developed between 1908 and 1912 . . .

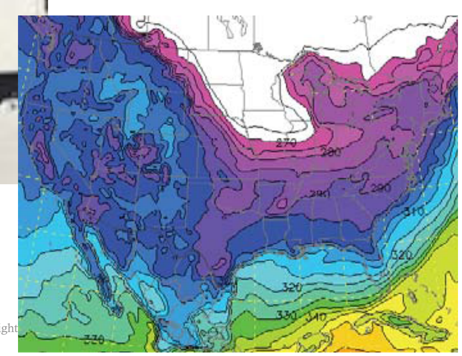
copyright 2011, UCAR, all rights reserved

Comparing objects can tell you things about your forecast like . . .

This:	Instead of this:
30% Too Big (area ratio=1.3)	POD = 0.35
Shifted west 1 km (centroid distance = 1km)	FAR = 0.7235
Rotated 15° (angle diff = 15%)	CSI = 0.1587
Peak Rain 1/2" too much (diff in 90^{th} percentile of intensities = 0.5)	

copyright 2011, UCAR, all rights reserved

Verifying with objects doesn't always make sense . . .



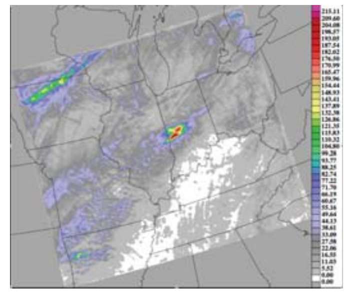
copyright

- In MET, object based verification is done using the MODE (Method for Object-Based Diagnostic Evaluation) tool.
  - Define objects
  - Compute attributes (e.g. area, centroid, axis angle, intensity)
  - Merge objects (e.g. thunderstorm cells merge into line)
  - Match forecast and observed objects
  - Compare attributes between matches
  - Output summary statistics

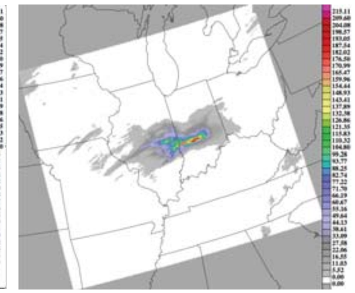
copyright 2011, UCAR, all rights reserved

Object verification

Forecast
6 hour accumulated precip



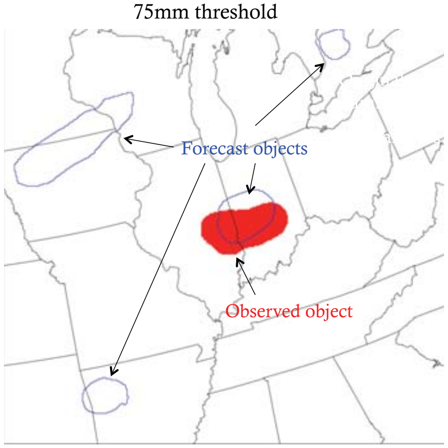
Observed
6 hour accumulated precip



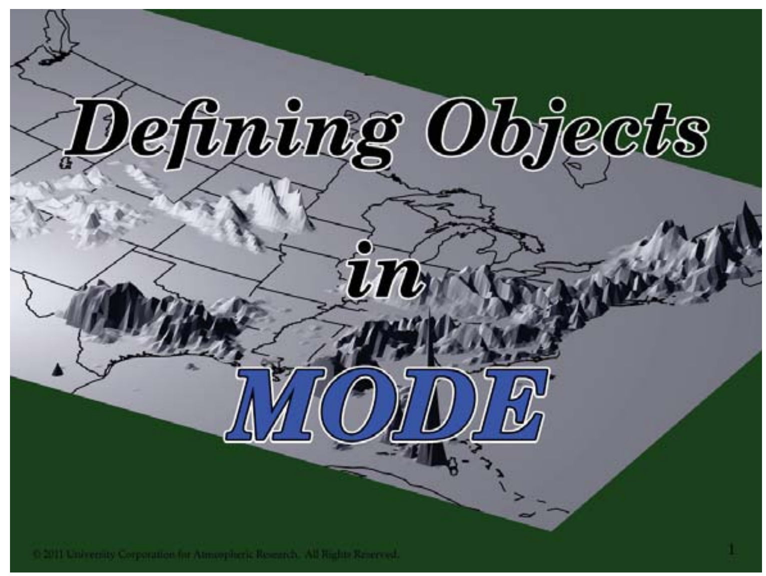
copyright 2011, UCAR, all rights reserved

Overlaid objects

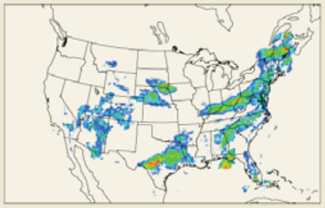
75mm threshold



copyright 2011, UCAR, all rights reserved



What are Objects?

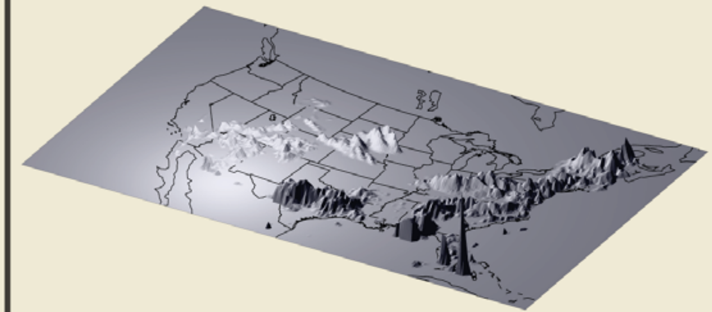


Raw Field

Object Field

Objects are Regions of Interest

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

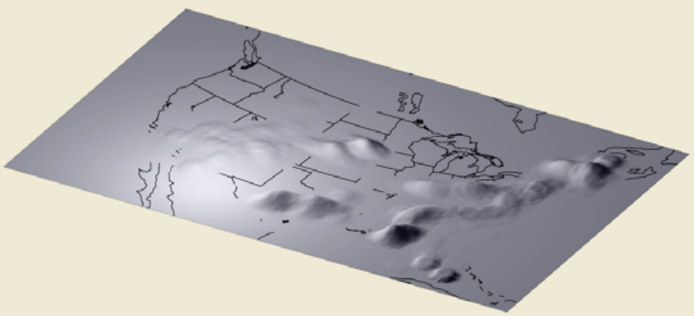


Step #1

Start with the raw data field.

In this case, a precipitation field.

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.



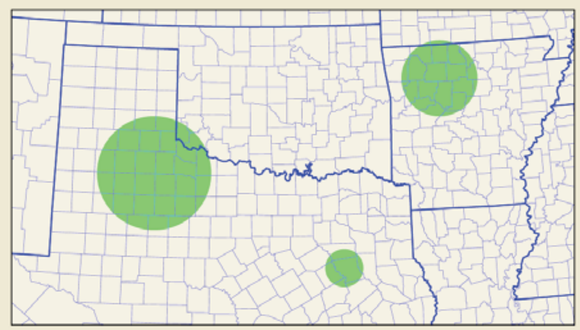
Step #2

Apply convolution operator.

This is basically a smoothing operation.

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Convolution Radius



Radius of Influence

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

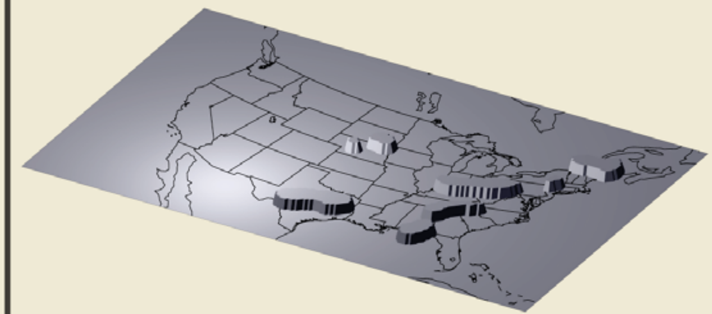
Convolution

Uses raw field $f(x, y)$ and

filter function $\phi(x, y)$

$$C(x, y) = \sum_{(\hat{x}, \hat{y}) \in G} \phi(\hat{x}, \hat{y})\, f(x - \hat{x}, y - \hat{y})$$

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.



Step #3

Threshold the smoothed field.

This produces an on/off mask field.

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

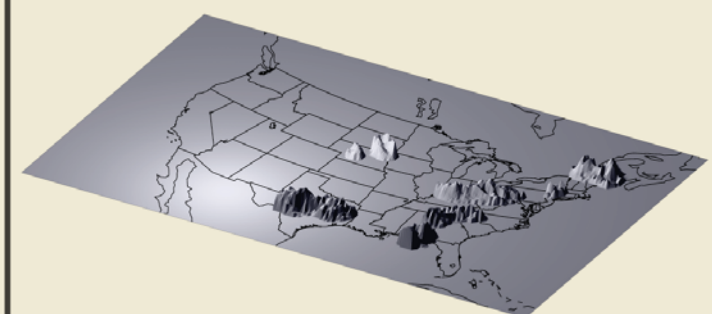
Masking

Uses convolved field $C(x, y)$

and threshold T

$$M(x, y) = \begin{cases} 1 & \text{if } C(x, y) \geq T \\ 0 & \text{else} \end{cases}$$

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.



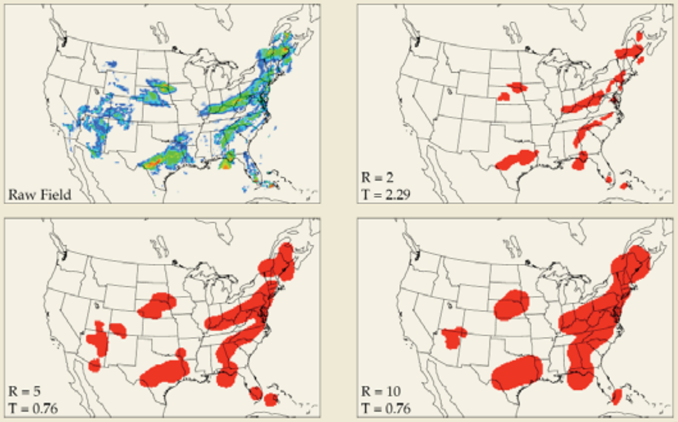
Step #4

Restore original data to object interiors.

This gives us our objects.

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Changing Object-Definition Parameters



© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Matching & Merging

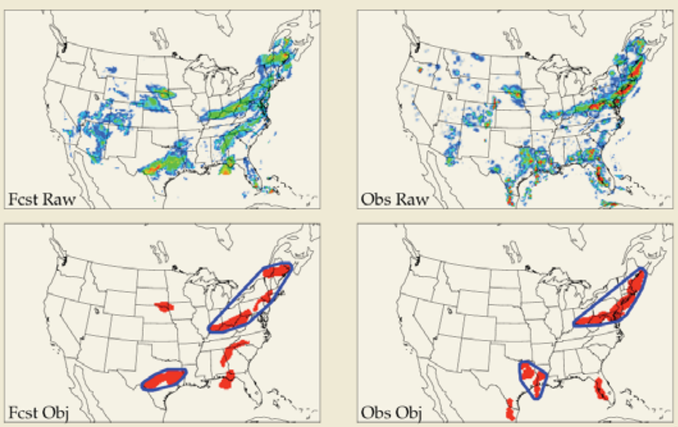
Merging: Associating objects in the same field.

Matching: Associating objects in different fields.

MODE does this using a Fuzzy-Logic engine.

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Example of Matching & Merging



© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Object Attributes

Single:

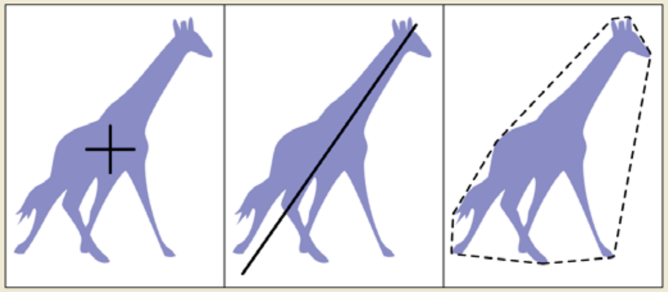
- Area
- Centroid
- Axis Angle
- Median Intensity
- Complexity
- Aspect Ratio
- Curvature

Pair:

- Centroid Distance
- Angle Difference
- Median Intensity Ratio
- Intersection Area
- Convex Hull Distance
- Boundary Distance
- Area Ratio

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

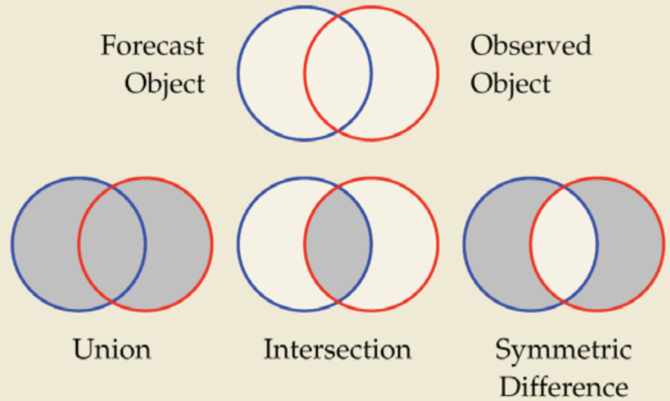
Example Single Attributes



Centroid | Axis | Convex Hull

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Example Pair Attributes



© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Interest Maps

Map attributes to interest values.

Example: Centroid Distance

All interest maps can be changed in the config file.

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Weights

Express relative importance of different attributes in matching and merging.

All weights can be changed in the config file.

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Total Interest

Calculated from weights, attributes, and interest maps.

$$T(\alpha) = \frac{\sum_i w_i \, C_i(\alpha) \, I_i(\alpha_i)}{\sum_i w_i \, C_i(\alpha)}$$

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

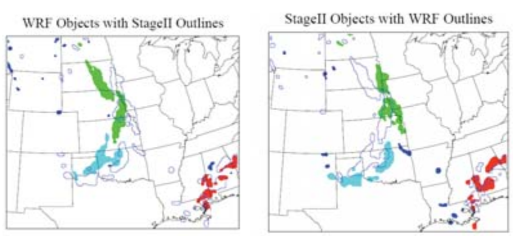
Alternative Merging Method

Double Thresholding

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Interpreting MODE Output

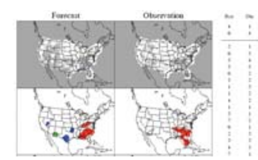
Verifying with Objects



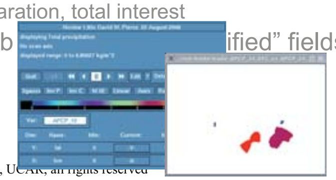
Presenter: Tara Jensen

copyright 2011, UCAR, all rights reserved

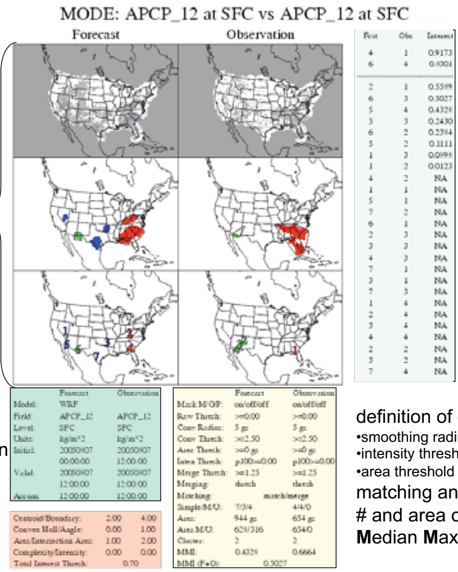
MODE Output



- PostScript
  - object pictures, definitions
  - matching/merging strategy
  - total interest for each object pair
- ASCII Text
  - attributes of simple, paired objects and clusters
    - size, shape, position, separation, total interest
  - verification scores (CSI, b … ified" fields
- netCDF
  - gridded object fields
  - view with ncview



copyright 2011, UCAR, all rights reserved



object pictures

total interest of object pairs

field names
model description

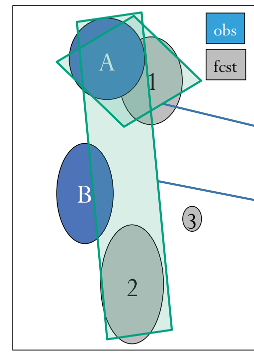
definition of objects
- •smoothing radius
- •intensity threshold
- •area threshold

matching and/or merging
and area of objects
Median **M**ax. **I**nterest (**MMI**)

weight of object attributes

Scoring Object Forecasts

Median of the Max. Interest (MMI*)



Interest Matrix

observed / forecast

	A	B
1	0.90	0.65
2	0.50	0.80
3	0.40	0.55

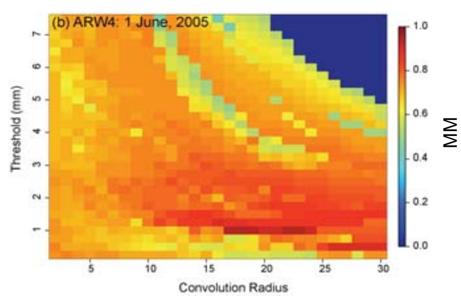
* Davis et al , 2009 The Method for Object based Diagnostic Evaluat on (MODE) Applied to WRF Forecasts from the 2005 SPC Spring Program Weather and Forecasting

maximum interest

$$MMI = \text{median}\{ 0\ 90, 0\ 80, 0\ 90, 0\ 80, 0\ 55 \} = 0\ 80$$

copyright 2011, UCAR, all rights reserved

Median of the Max. Interest (MMI) Quilt Plot

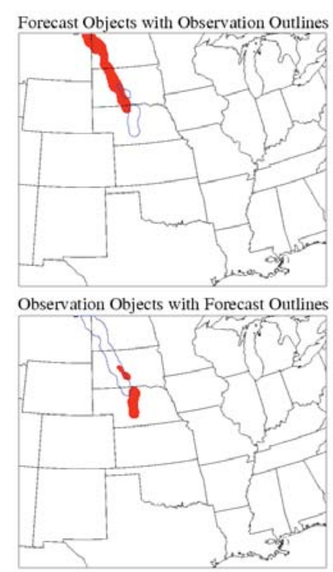


MMI as a function of convolution radius (grid squares) and threshold (mm) for 24 h forecast of 1 h rainfall

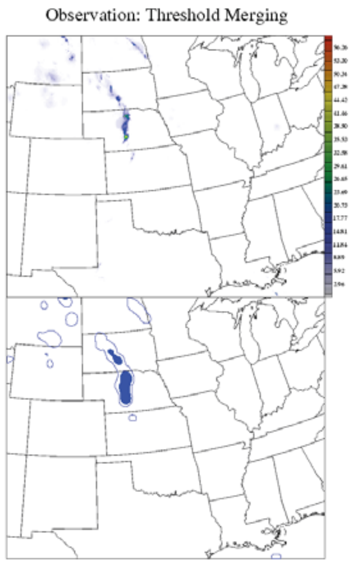
Each pixel is a MODE run.

This graphic is not in MET, but R code on MET website.

copyright 2011, UCAR, all rights reserved

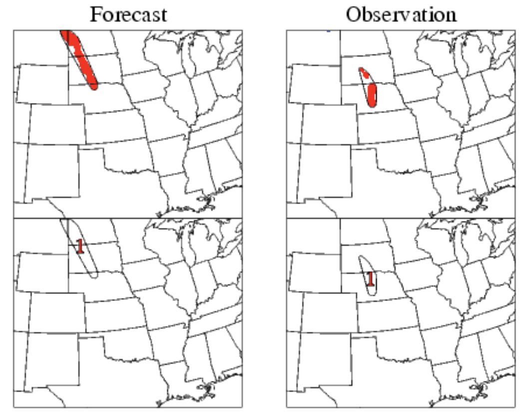


Merging Objects with Double Threshold



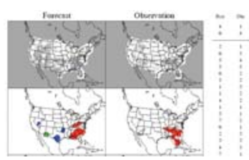
copy

Cluster Object Information

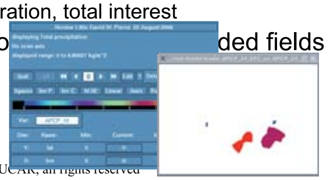


CLUS PAIR	CEN DIST	ANG DIFF	FCST AREA	OBS AREA	INTER AREA	UNION AREA	SYM DIFF	FCST INT50	OBS INT50	FCST INT90	OBS INT90	TOT INTR
1	97.32	12.16	4184	1762	358	5588	5230	5.33	4.06	16.51	15.57	0.7622

MODE Output



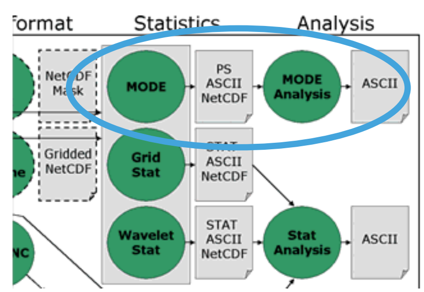
- PostScript
  - object pictures, definitions
  - matching/merging strategy
  - total interest for each object pair
- ASCII Text
  - attributes of simple, paired objects and clusters
    - size, shape, position, separation, total interest
  - verification scores on smo ded fields (objects)
- netCDF
  - gridded object fields
  - view with ncview



copyright 2011, UCAR, all rights reserved

MODE Analysis Tool

mode_analysis



copyright 2011, UCAR, all rights reserved

MODE Analysis Tool
-*summary* Example

Command Line

```
mode analysis  summary \
     lookin mode output/wrf4ncep/40km/ge03 \
     fcst   cluster \
     area min 100 \
     column centroid lat  column centroid lon \
     column area \
     column axis ang \
     column length
```

Output

```
Total mode lines read =  393
Total mode lines kept =   17
```

Field	N	Min	Max	Mean	StdDev	P10	P25	P50	P75	P90	Sum
centroid lat	17	31 97	46 24	38 65	3 81	33 89	36 13	38 54	40 12	43 99	657 00
centroid lon	17	103 89	85 20	96 32	5 91	103 15	102 65	96 26	93 95	86 78	1637 49
area	17	180 00	8393 00	2955 06	2246 49	624 80	1206 00	2662 00	3958 00	5732 20	50236 00

copyright 2011, UCAR, all rights reserved

MODE Analysis Tool

Command Line

```
mode analysis -bycase -lookin mode output/wrf4ncep/40km/ge03. -single -simple
```

Output

```
Total mode lines read =  393
Total mode lines kept =  141
```

Fcst Valid Time	Area Matched	Area Unmatched	# Fcst Matched	# Fcst Unmatched	# Obs Matched	# Obs Unmatched
Apr 26, 2005 00:00:00	3210	1046	2	4	1	1
May 13, 2005 00:00:00	8892	9320	2	19	1	2
May 14, 2005 00:00:00	16994	4534	7	4	5	3
May 18, 2005 00:00:00	6057	852	3	2	2	1
May 19, 2005 00:00:00	1777	1624	1	5	2	1
May 25, 2005 00:00:00	8583	928	4	2	4	2
Jun 1, 2005 00:00:00	12456	2657	5	6	6	2
Jun 3, 2005 00:00:00	7561	102	11	1	5	0

copyright 2011, UCAR, all rights reserved

Scoring MODE Object Forecasts

use total interest threshold to separate matched objects, or "hits" from false alarms and misses

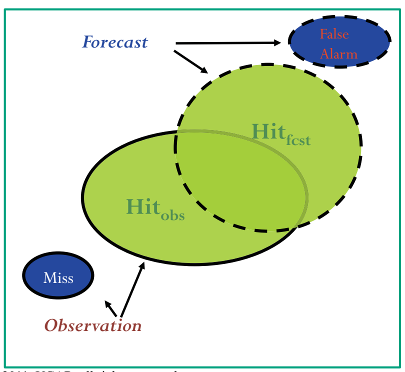
Traditional Categorical Statistics

critical success index (CSI)

$$\frac{\text{Hit}}{\text{Hit} + \text{Miss} + \text{False Alarm}}$$

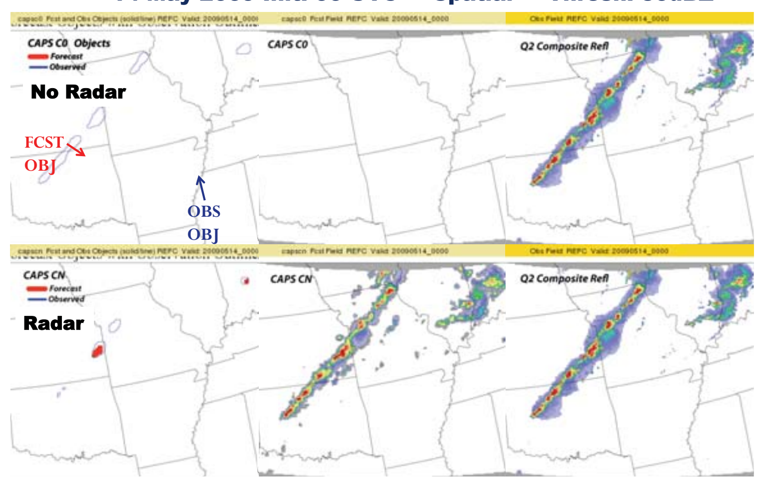
bias $\frac{\text{Hit} + \text{False Alarm}}{\text{Hit} + \text{Miss}}$

sometimes *area weighted*



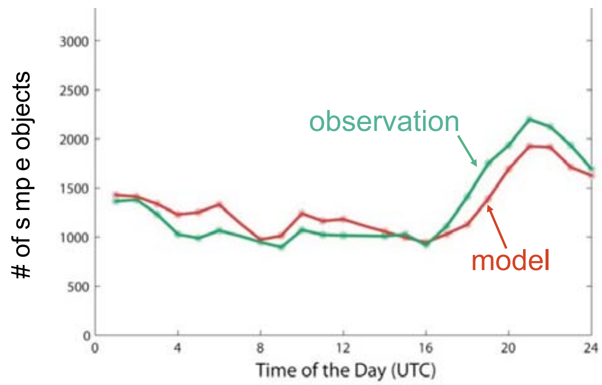
copyright 2011, UCAR, all rights reserved

14 May 2009 Init: 00 UTC   Spatial   Thresh: 30dBZ



Objects   Forecast Field   Observed Field

Diurnal Cycle of the Number of Storms



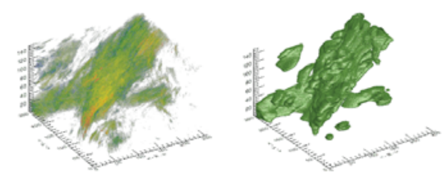
copyright 2011, UCAR, all rights reserved

Interpreting MODE Output

PostScript, ASCII, and netCDF

mode analysis Tool

Examples of Diagnostic Evaluations



Spatial Scale:

Neighborhood and Scale-separation Approaches to Verification

Eric Gilleland

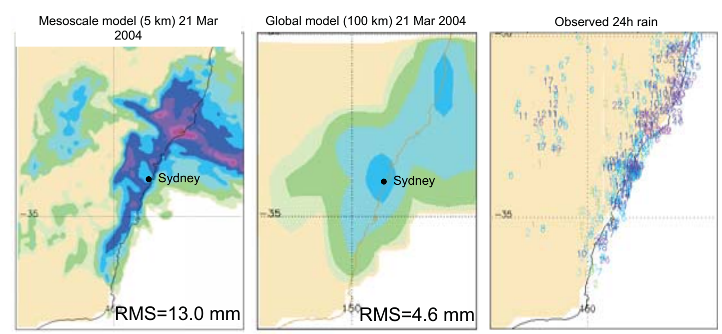
National Center for Atmospheric Research

Boulder, CO Email: ahijevyc@ucar.edu

slides from **D. Ahijevych** and others

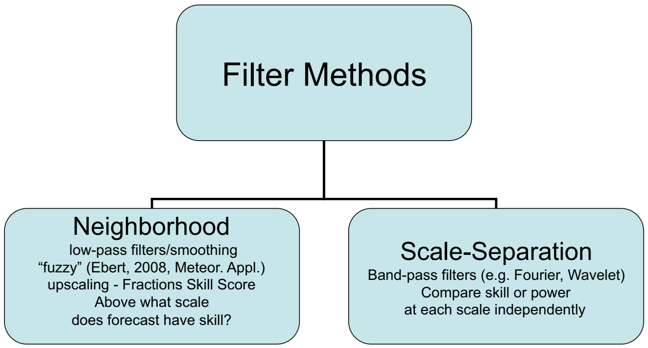
Impact of high vs. low resolution (or smoothing)

Which forecast would you rather use?



copyright 2010 UCAR, a l rights reserved

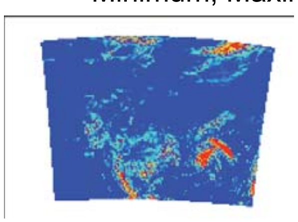
Giving Credit for a Close Forecast



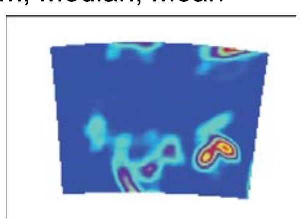
copyright 2010 UCAR, all rights reserved

Neighborhood Methods Smoothing

- Smoothing Filters in MET
  - Minimum, Maximum, Median, Mean



original

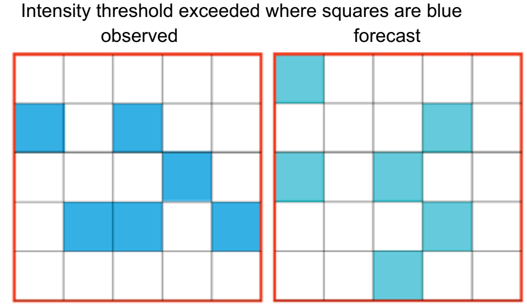


mean

copyright 2010 UCAR, all rights reserved

Neighborhood Methods: Fractional coverage of events

Intensity threshold exceeded where squares are blue



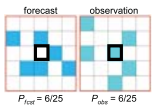
Fraction = 6/25 = 0.24   Fraction = 6/25 = 0.24

P is the fractional event frequency within the neighborhood.
This is calculated for all *n* grid points in the domain

slide from Mittermaier

copyright 2010 UCAR, a l rights reserved

Neighborhood Methods



n = grid points in the domain

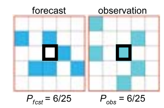
$$\frac{1}{n}\sum_{i=1}^{n}\left(P_{fcst,\,i}-P_{obs,\,i}\right)^2$$

Fractions Brier Score

Roberts and Lean (2008)

copyright 2010 UCAR, all rights reserved

Neighborhood Methods



$$1-\frac{\text{Fractions Brier Score}}{\text{(worst possible) Fractions Brier Score}}$$

Fractions Skill Score (FSS) of Roberts and Lean (2008)

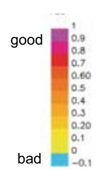
copyright 2010 UCAR, a l rights reserved

Neighborhood Methods

Table courtesy of E Ebert.

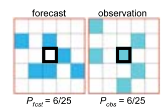
Fractions skill score – FSS

Spatial scale (pts) \ Threshold (mm)	1	2	5	10	20	50	100	200
129	**0.96**	**0.95**	**0.93**	**0.87**	**0.80**	0.28	0.00	
65	**0.91**	**0.88**	**0.83**	**0.75**	0.50	0.27	0.00	
33	**0.78**	**0.72**	**0.60**	**0.51**	0.29	0.11	0.00	
17	**0.62**	**0.53**	0.35	0.23	0.08	0.00	0.00	
9	0.51	0.42	0.24	0.13	0.03	0.00	0.00	
5	0.44	0.36	0.19	0.09	0.02	0.00	0.00	
3	0.40	0.32	0.16	0.08	0.02	0.00	0.00	
1	0.33	0.25	0.13	0.05	0.01	0.00	0.00	

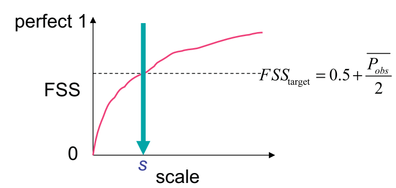


copyright 2010 UCAR, all rights reserved

Neighborhood Methods



- FSS improves with neighborhood size, or *scale*.
- To find a “useful” scale, define a target FSS.
  usually a random forecast with same frequency of events as observations
- Find scale *(s)* where FSS exceeds FSS_{target}.



copyright 2010 UCAR, a l rights reserved

Neighborhood Methods

- In MET, Neighborhood methods are in grid_stat tool.
- Smoothing filters in MET
  - Minimum
  - Maximum
  - Median
  - Mean
- Fractional coverage
  - Fractions Brier Score
  - Fractions Skill Score
- See Ebert (*Meteorol. Appl*. 2008) for a summary and comparison of these techniques.

copyright 2010 UCAR, all rights reserved

Giving Credit for a Close Forecast

Filter Methods

Neighborhood
low-pass filters/smoothing
"fuzzy" (Ebert, 2008, Meteor. Appl.)
upscaling - Fractions Skill Score
Above what scale
does forecast have skill?

Scale-Separation
Band-pass filters (e.g. Fourier, Wavelet)
Compare skill or power
at each scale independently

copyright 2010 UCAR, a l rights reserved

Scale-Separation Methods

- Fourier
  - Skamarock (2004), MWR **132**:3019-3032
  - Harris *et al*. (2001), J Hydrometeorol. **2**:406-418
  - Tustison *et al*. (2001), JGR **106**(D11): 11775-11784
  - *and many more…*
- Wavelet
  - Briggs and Levine (1997), MWR **125**:1329-1341
  - Casati *et al*. (2004). [In MET wavelet_stat tool]

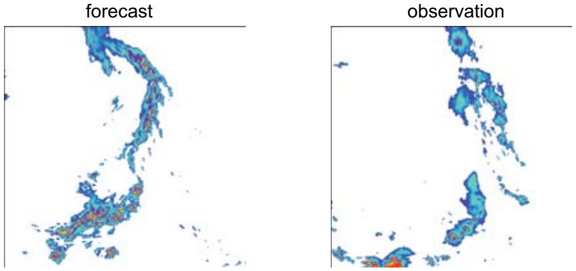
copyright 2010 UCAR, all rights reserved

Scale Separation Methods

Intensity Scale (IS) (Casati *et al*., 2004)

1. Create binary fields for a threshold

forecast observation



copyright 2010 UCAR, a l rights reserved

Scale Separation Methods

Intensity Scale (IS) (Casati *et al.*, 2004)

1. Create binary fields for a threshold

forecast

observation

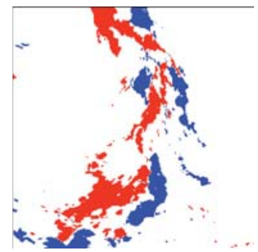
copyright 2010 UCAR, all rights reserved

Scale Separation Methods

Intensity Scale (IS) (Casati *et al.*, 2004)

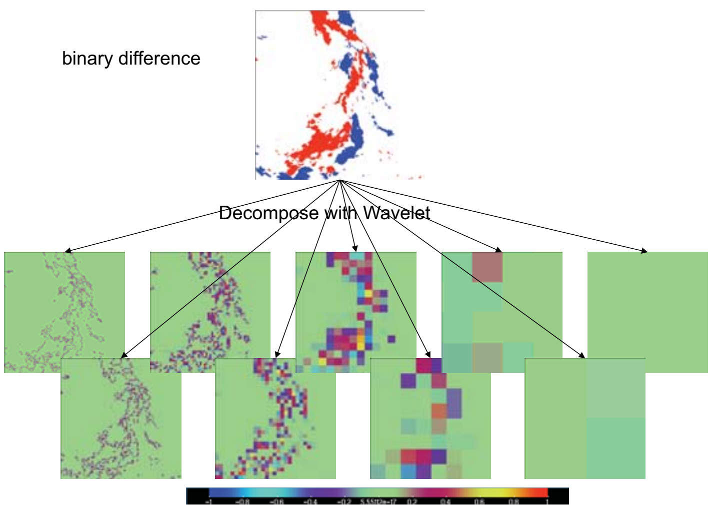
Subtract binary fields for a threshold

binary difference



copyright 2010 UCAR, a l rights reserved

binary difference



copyright 2010 UCAR, all rights reserved

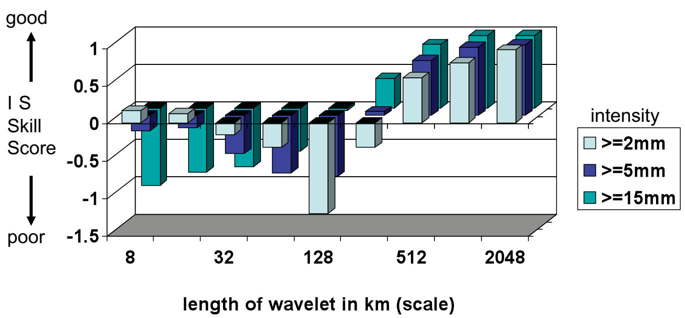
Scale Separation Methods

Intensity Scale (IS) (Casati *et al.*, 2004)

1. Create binary fields for a threshold
2. Apply wavelet decomposition to binary difference
3. Calculate mean squared error MSE for each scale j
4. Calculate MSE for a random forecast based on the sample climatology
5. Intensity-scale Skill Score $IS\ skill\ score_j = 1 - \frac{MSE_j}{\frac{MSE_{random}}{n+1}}$
6. Repeat for multiple thresholds

copyright 2010 UCAR, a l rights reserved

Intensity-Scale Skill Score



copyright 2010 UCAR, all rights reserved

Thank you…Questions?

Spatial Methods Intercomparison Project

http://www.ral.ucar.edu/projects/icp

copyright 2010 UCAR, a l rights reserved

Customizing and Understanding Wavelet Stats

Tara Jensen

copyright 2011, UCAR, all rights reserved.

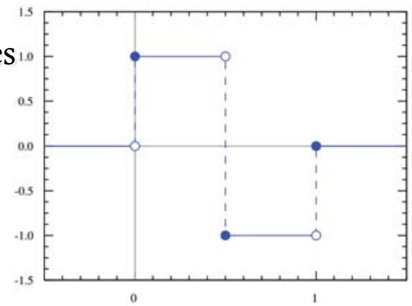
Tiles – Grid must be 2^n x 2^n

- Cut down – user selected subset (square)
- Tiles – automated selection of subset(s)
- Pad with zeros – not recommended unless adds very small number of points.

copyright 2011, UCAR, all rights reserved

Wavelets

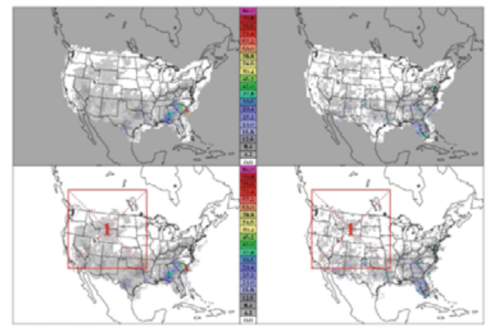
- Haar
- Centered Haar
- Daubechies
- Centered Daubechies
- B spline
- Centered B spline



copyright 2011, UCAR, all rights reserved

Wavelet-Stat: APCP at A24

Forecast Observation



Model Name: WRF

Init Time: Aug 7, 2005 00:00:00
Valid Time: Aug 8, 2005 00:00:00
Lead Time: 24:00:00
Accum Time: 24:00:00

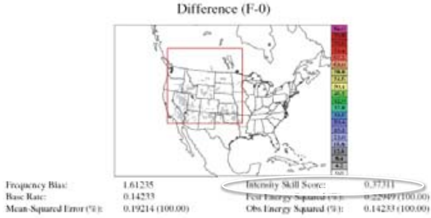
Tile Method: User-Defined
Tile Count: 1
Tile Dim: 64 x 64
Tile Corner: (45, 45)

Mask Missing: Fcst/Obs
Wavelet(k): Haar (2)

copyright 2011, UCAR, all rights reserved

Wavelet-Stat: APCP at A24, Tile 1, >1.000, Binary

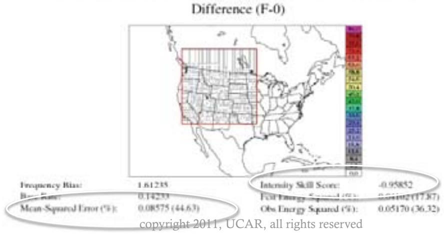
Difference (F-0)



Overall forecast has skill (ISS > 0)

Wavelet-Stat: APCP at A24, Tile 1, >1.000, Scale 1

Difference (F-0)



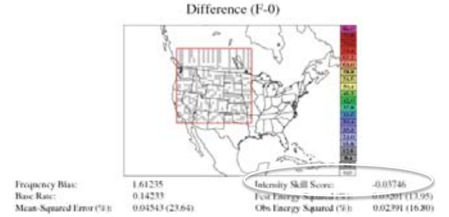
Errors at this scale account for nearly half of the MSE.

At this scale, forecast does not have skill (ISS < 0)

copyright 2011, UCAR, all rights reserved

Wavelet-Stat: APCP at A24, Tile 1, >1.000, Scale 2

Difference (F-0)



Forecast transitions from no skill to skill at scale 3.

Wavelet-Stat: APCP at A24, Tile 1, >1.000, Scale 3

Difference (F-0)

copyright 2011, UCAR, all rights reserved

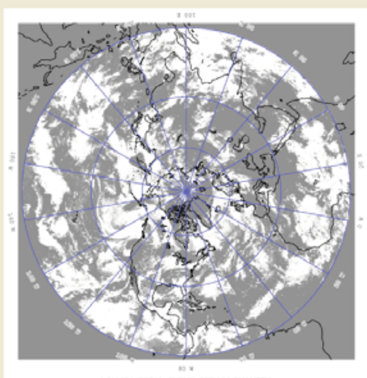
Summary

- Wavelet tool provides a flexible method for decomposing spatial fields into different scales.
- Once decomposed, verification measures at each *physical* scale can be examined and compared.

copyright 2011, UCAR, all rights reserved

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Example of WWMCA Plot Tool Output




```
usage:  wwmca_plot [ -outdir path ] wwmca_cloud_pct_file_list
```

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Regridder

- Uses Config File
- Specify "To" Grid by name or parameter list
- Specify Interpolation Parameters
- Specify NetCDF Output Parameters

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

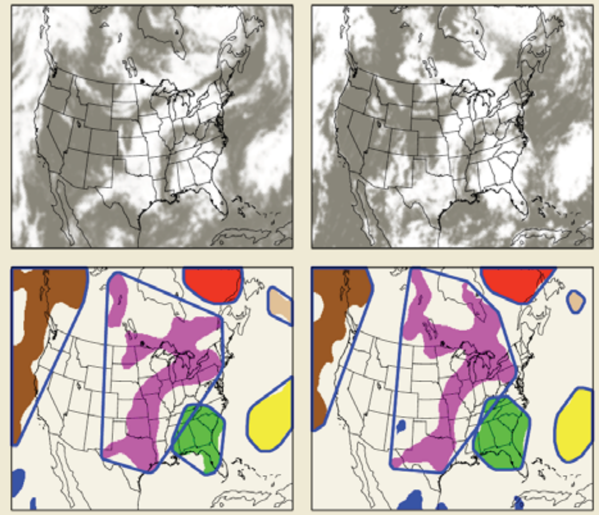
Regridder

- Only need 1 input file unless the "To" grid comes close to the equator.

```
usage:  wwmca_regrid
            -out filename
            -config filename
            [ -nh wwmca_filename ]
            [ -sh wwmca_filename ]
```

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

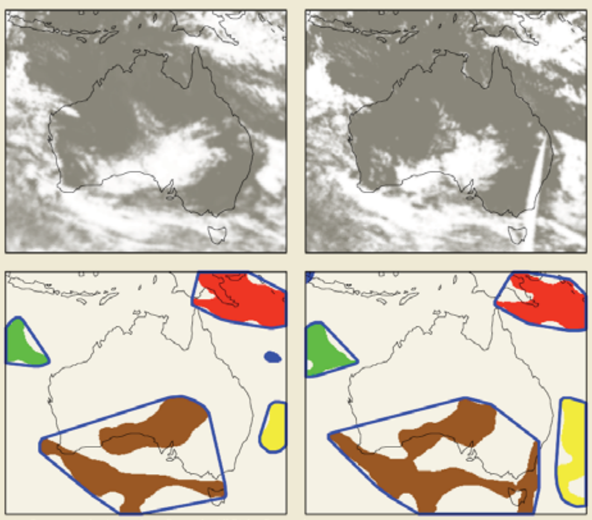
Forecast



Observed

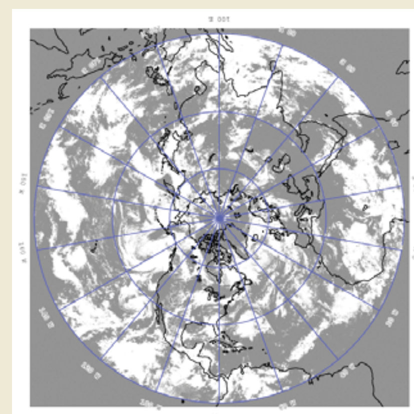
© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Forecast



Observed

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.



© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

METViewer: MET Data Visualization

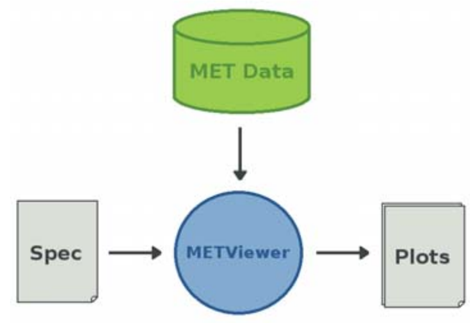
Paul Oldenburg
NCAR - RAL

Developmental Testbed Center

NCAR

What is METViewer?

METViewer is a data visualization (plotting) system



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Demand Arises...

- MET outputs statistics in data tables
- Users frequently request a system for visualizing MET output data
- Scientists want plots showing statistics for variables, cases and relationships
- Often, many plots are reviewed to find a small number of illustrative cases

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

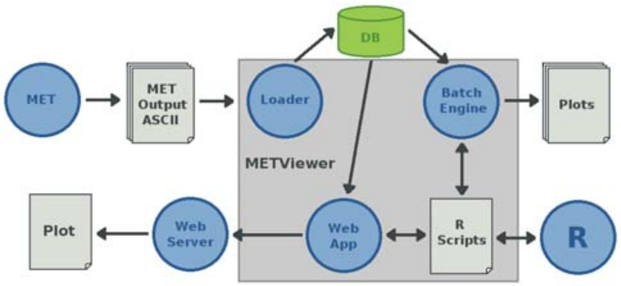
Design Principles

- **Relational database:** for searching, sorting and grouping MET output data - MySQL
- **Plotting system:** for consistent and robust graphics – R
- **Web application:** for accessibility, portability and ease of use – java & apache tomcat
- **Batch processing:** for generating many plots with only a few simple commands - java
- **Distributable**

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

METViewer Design

METViewer consists of three components which implement the solution ideas



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Plot Specification

- Data type:
  - Traditional Statistics
  - MODE Statistics
- Plot type / Template:
  - Series plot
  - Box plot
  - Bar plot
- Independent variable and values:
  - Lead time (FCST_LEAD)
  - Valid / Initialization time (FCST_VALID_BEG)
  - Threshold (FCST_THRESH)
  - Vertical Level (FCST_LEV)

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Plot Specification Cont'd

- Dependent variable(s)
  - Forecast Variable (e.g. APCP_06, TMP, WIND)
  - Statistic (e.g. CSI, RMSE, BRIER, AREA, ANGLE_DIFF)
- Series variables and values: lines on plot
- Fixed variables and values: case information
- Aggregation CTCs or SL1L2
- Plot formatting
- Series formatting & confidence intervals:
  - Normal
  - Bootstrap
  - Standard error (median or mean) configurable a

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Plot Types

Currently Implemented:

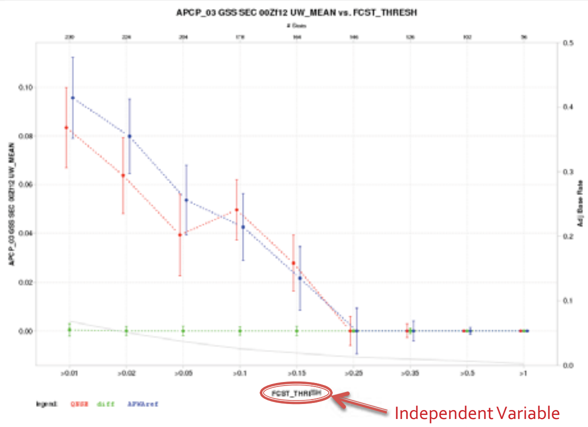
- Time series
- Threshold series
- Vertical levels (independent variable on vertical axis)
- Box plot / Points Distribution plot
- Bar plot

Forthcoming:

- Scatter plot (statistic 1 vs. statistic 2)
- Frequency Histogram
- Rank Histogram
- ROC

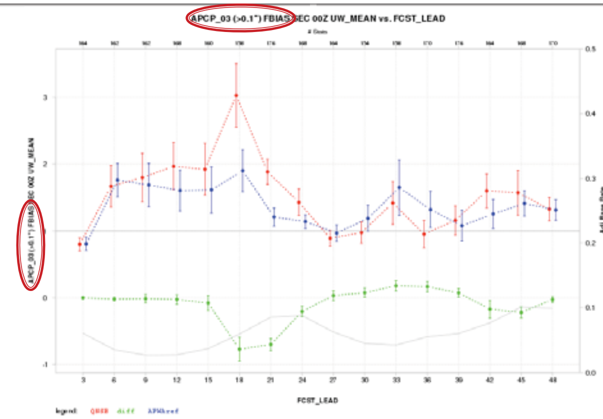
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Independent Variable



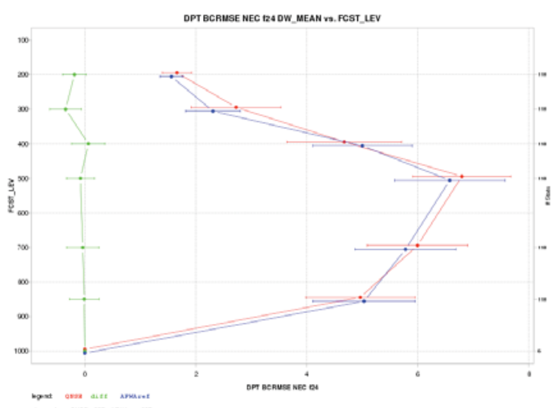
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Dependent Variable



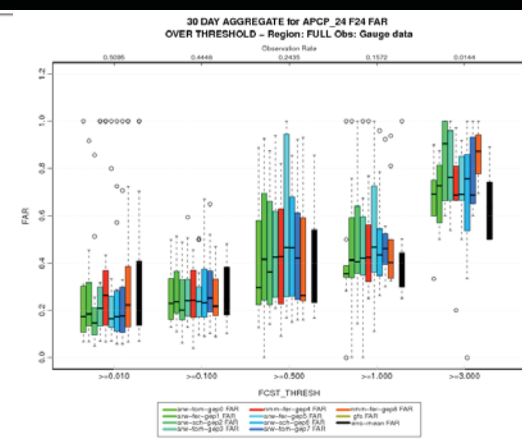
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Vertical Levels



Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Box Plot



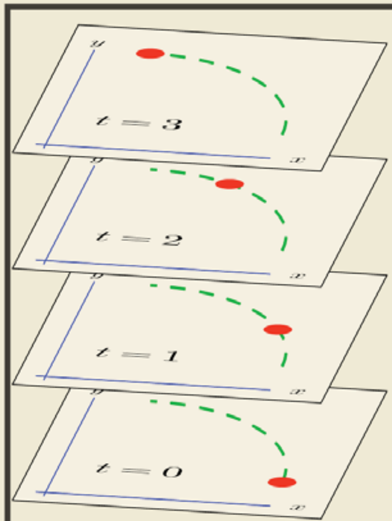
Copyright 2011, University Corporation for Atmospheric Research, all rights reserved

Web App Demo

… and questions

http://www.dtcenter.org/met/metviewer/db/qnse

Copyright 2011, University Corporation for Atmospheric Research, all rights reserved



Time Slices

Stacked Vertically

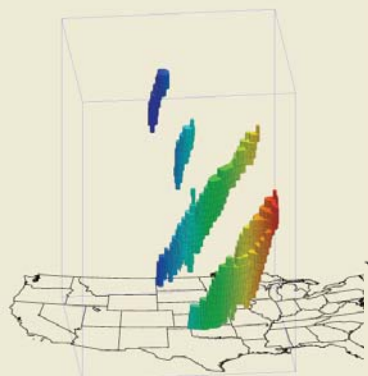
Moving 2D Object

Sweeps Out

3D Spacetime Object

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

3D Objects



June 13, 2002
IHOP Precip Data

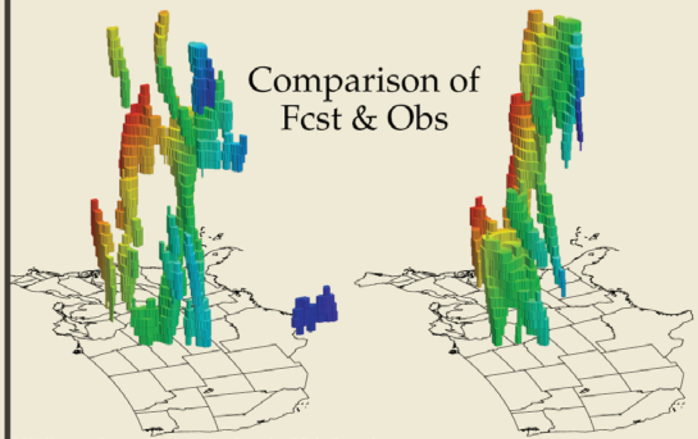
Vertical Dimension
is Time

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

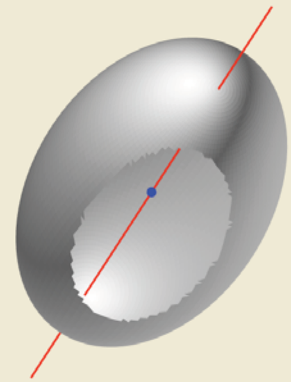
Forecast

Stage IV

Comparison of
Fcst & Obs



© 2011 University Corporation for Atmospheric Research. All Rights Reserved.



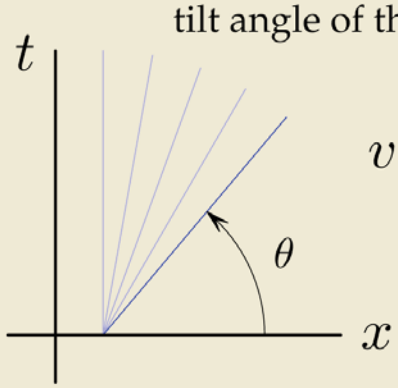
Most attributes
for 2D objects
generalize easily
to the 3D case

Here we see the centroid
and axis calculated for this
3D ellipsoidal object

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.

Velocity is related to the

tilt angle of the axis



$$v \propto \cot\theta$$

© 2011 University Corporation for Atmospheric Research. All Rights Reserved.