Environment Canada Environnement Canada

What Is A Good Forecast: The importance of user-orientation in verification

Laurence Wilson
Atmospheric Science and Technology Branch
Montreal, Quebec

June 8, 2009

Outline

- Role(s) of verification
- Current approaches to verification systems
  - Limitations vis-à-vis users
- Increasing user relevance
- Examples
  - Ensemble forecasts of extreme precipitation
  - Spatial object–based verification
- Summary – a blueprint for relevance

Role of Verification

- Evaluation of forecast “goodness” (ref: Murphy)
  - CONSISTENCY: forecasts agree with forecaster’s true belief about the future weather [*strictly proper*]
  - QUALITY: correspondence between observations and forecasts [*verification*]
  - VALUE: increase or decrease in economic or other kind of value to someone as a result of using the forecast [*decision theory*]
- Evaluation of delivery system (“Performance measurement”)
  - timeliness (are forecasts issued in time to be useful?)
  - relevance (are forecasts delivered to intended users in a form they can understand and use?)
  - robustness (level of errors or failures in the delivery of forecasts)

Principles of (Objective) Verification

- Verification activity has value only if the information generated leads to a decision about the forecast or system being verified
  - User of the information must be identified
  - Purpose of the verification must be known in advance
- No single verification measure provides complete information about the quality of a forecast product.
- Forecast must be stated in such a way that it can be verified

Desirable characteristics of a User-relevant verification system

- Provides information that is relevant to a wide spectrum of users (Flexibility in design)
  - *<u>Ex</u>*: Multiple (user-selectable) thresholds
- Use diagnostic techniques
  - *<u>Ex</u>*: Distributions of statistics rather than (or in addition to) summary scores
- Provide uncertainty information about verification measures (e.g., confidence intervals)
- *<u>Ideally</u>* – strong interaction with users
  - Understand applications of verification information
  - Requires engagement of social science community and understanding of communication and decision-making aspects
  - Useful to actually pose a specific verification question to be answered by the system

Examples of current systems which are not really user-oriented

- Many or most NMS operational model verification systems
  - Primary user is probably modelers, because the system is usually designed by and for modelers, BUT
  - A “large” user community is assumed
  - User usually not specified or consulted
  - Mainly focused on upper air parameters
  - Verification against analysis common (incestuous)
  - QC of observations using data assimilation system
- Example: WMO “standard” NWP verification system
  - Fits all the above points
  - Goal is to compare accuracy of NWP models, BUT
  - Each center uses its own analysis, AND,
  - Each center QC’s observations against its own data assimilation system – verification dataset used varies from one center to another
- The good news is that the WMO standard system is being redesigned

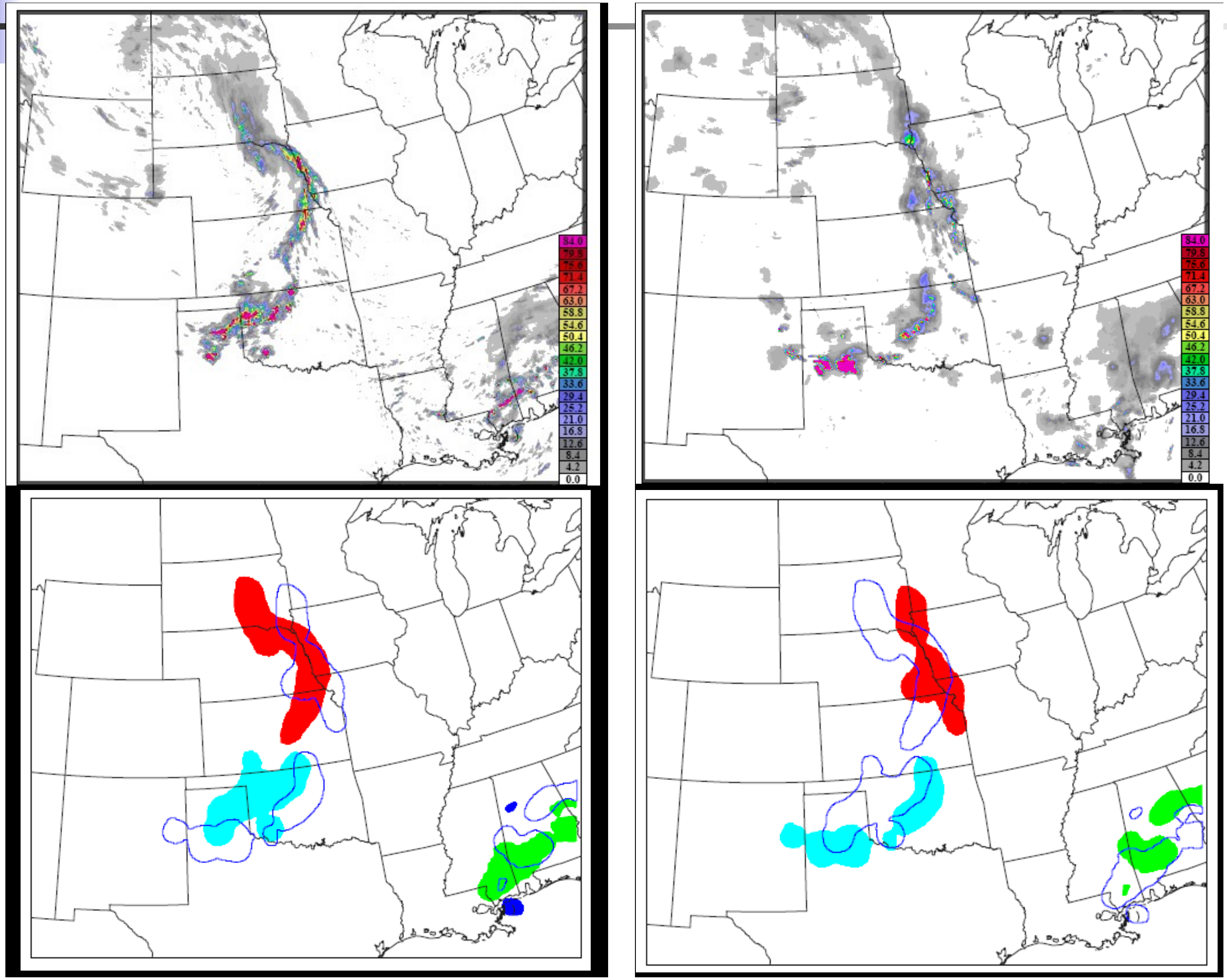
Example
1. Spatial verification of precipitation areas - MODE

- *Target users*: Forecasters who use modeled precipitation fields as guidance
- *Question*: What are the errors in WRF prediction of precipitation areas over the contiguous US?
- *Predictand*: User-relevant attributes of precipitation
- *Verification measures*
  - Ratios of fcst/obs attributes (object size, median and extreme intensities)
  - Magnitude of displacement error
- *Data*
  - 4-km Weather Research and Forecasting (WRF) model 1995; 24-h forecasts
  - 4-km Stage 2 precipitation analysis
  - Single cases and distributions over 2 years

MODE = "Method for Object Based Diagnostic Evaluation"
See Davis et al. 2006, MWR

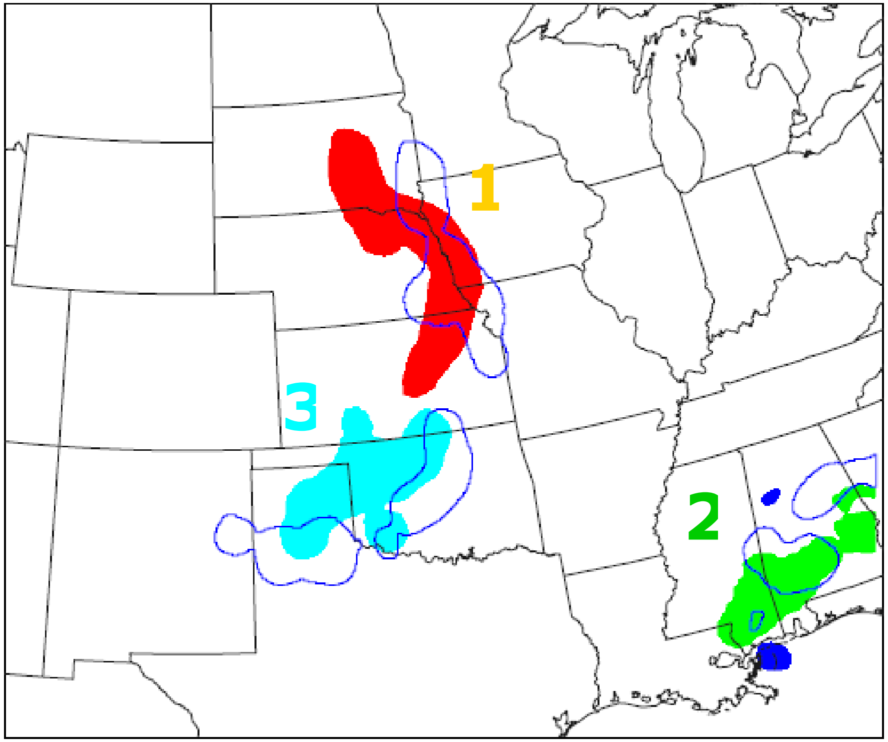
Object-based example: 1 June 2005

24-h precip forecast (4 km)



Precip analysis (4 km)

Object-based example: 1 June 2005



WRF ARW Objects with Stage II Objects overlaid

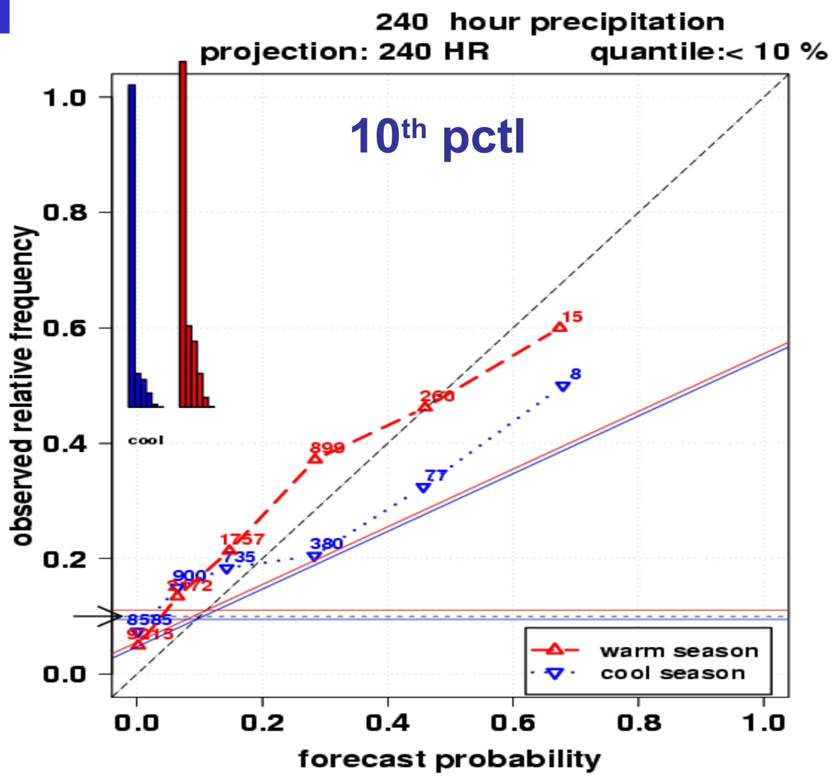
- MODE quantitative results indicate
  - Most forecast areas too large
  - Forecast areas slightly displaced
  - Median and extreme intensities too large
  - BUT – overall – forecast is pretty good
- In contrast:
  - POD = 0.40
  - FAR = 0.56
  - CSI = 0.27

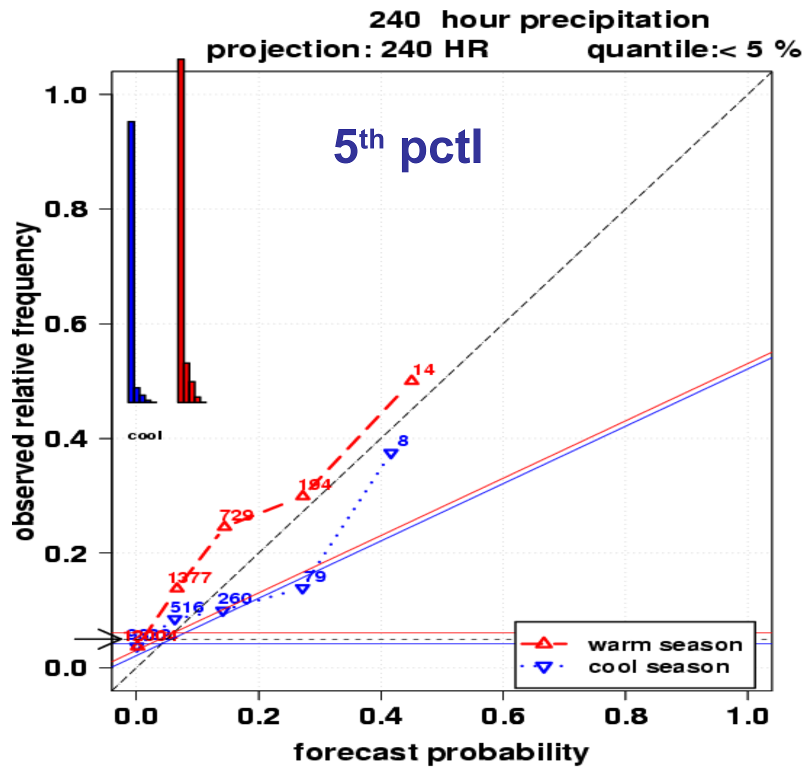
Tuning verification to the needs of users 2. Verification of unusual precipitation events

- *Target users*: Forecasters who use the Canadian ensemble system to prepare precipitation probability forecasts
- *Question*: Can the 16-member ensemble system reliably predict probabilities of unusual events?
- *Predictands*:
  - 90th, 95th and 99th percentiles of the climatological (30-40 year) precipitation distribution for 36 Canadian stations
    - 1- and 3-day accumulations
  - 5th and 10th percentiles over 10 days (dry periods)
- *Verification measures*: Reliability, ROC, Brier Skill score. 3.5 years of data available for each station

See: Peel and Wilson, 2008: Diagnostic verification of the precipitation forecasts produced by the Canadian ensemble prediction system, *Weather and Forecasting*, **23**, 596-616.

10 day dry period – 10% and 5%





Forecasts of "*probability of 10-day dry period*" are skillful and reliable, but especially in summer

Summary

- “User-oriented verification” means knowing who the users are, talking to them, and designing verification systems to answer their questions about the quality of forecasts.
- Steps in the right direction:
  - Diagnostic verification
  - Addition of spatial verification tools to “standard” point-wise tools
  - Inclusion of uncertainty estimates in verification results
  - Design of flexible systems, with user-controllable parameters.
- Long way to go
- Do you know who your users are and what they want?