CO-FPS: Metrics and Verification

Colorado Fire Prediction System March Meeting

March 29, 2016

bgb@ucar.edu

Part 1

METHODS AND APPROACHES FOR TESTING AND VERIFYING CO-FPS PRODUCTS

Outline

- What is a metric? What is forecast evaluation?
  - Difference between quality and value
- Need to include the user in determining metrics and what defines success
  - User-relevant verification
- General types of verification
  - Subjective
  - Grid-to-grid
  - Shape/spatial

What is verification?

Verify: ver·i·fy
Pronunciation: 'ver-&-"fI
1 : to confirm or substantiate in law by oath
2 : to establish the truth, accuracy, or reality of <verify the claim>
synonym see CONFIRM

- 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"

Metrics and Verification

- Metric:
  - A *standard for measuring or evaluating something*, especially one that uses figures or statistics: new metrics for gauging an organization's diversity
- Verification:
  - The process of research, examination, etc., required to prove or establish authenticity or validity
- Thus, metrics need to be carefully defined to do meaningful verification

Forecast value and user-relevant metrics

NCAR

Forecast Value (or "Goodness")

Depends on the *quality* of the forecast

and

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

Ideal: Closely connect *quality measures* to *value measures*

This concept is fundamental to selecting metrics for CO-FPS

Types of forecasts and dimensions

NCAR

Variables

- Fire extent
- Rate of spread
- Heat release
- Smoke concentration
- Significant fire phenomena
- Turbulence intensity
- Downdraft/updraft regions
- Wind shear regions
- Wind speed and directions
- Wind speed gustiness
- Surface air temperature
- Surface relative humidity
- Other?

Dimensions / Attributes

- Size
- Shape
- Location
- Timing
- Intensity
- Other?

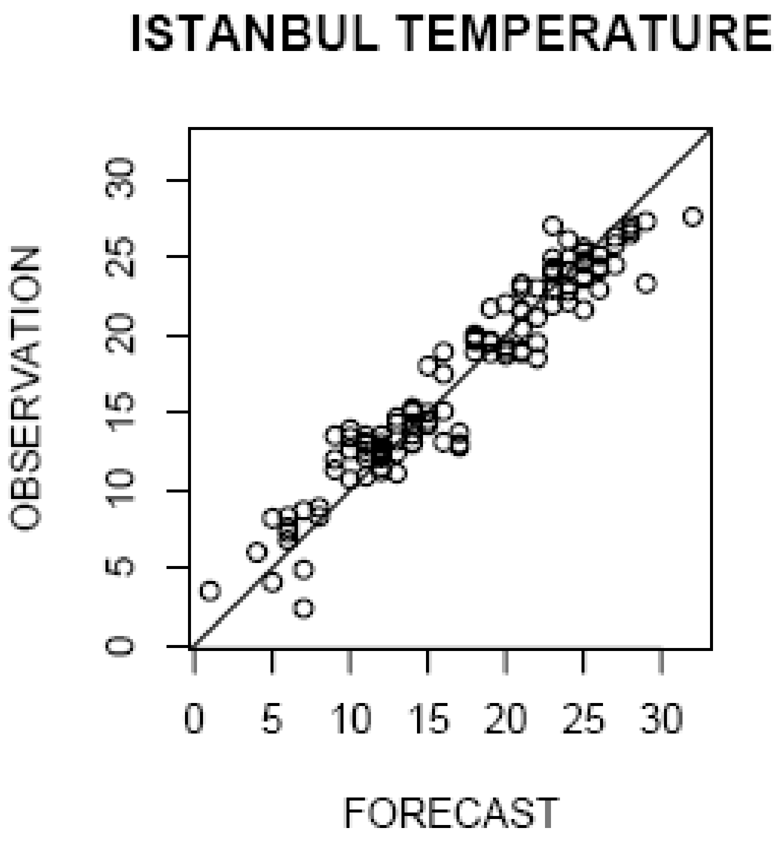
Identifying (a) **Characteristics** of the forecasts and observations and (b) which **attributes** are most important are the first steps in defining *metrics* and a *verification strategy*

Types of forecasts, observations

NCAR

- Continuous
  - Humidity at points in space and time
  - Fire intensity at points
- Categorical
  - Dichotomous
    - ❖ Fire vs. no fire
    - ❖ Strong winds vs. no strong wind
    - ❖ Low humidity vs. high humidity
    - ❖ Often formulated as Yes/No
  - Multi-category
    - ❖ Fire intensity category
    - ❖ Wind speed category
  - May result from *subsetting* continuous variables into categories

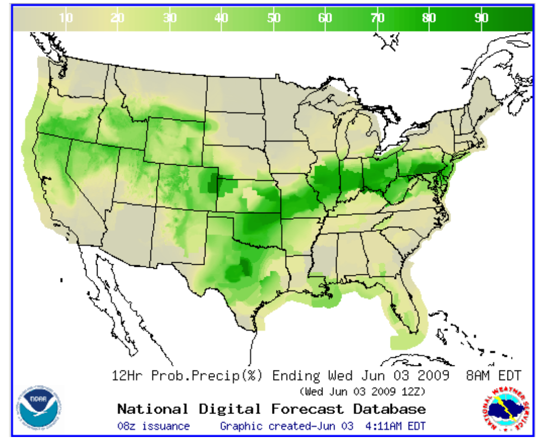
    Ex: Temperature categories of 0-10, 11-20, 21-30, etc.



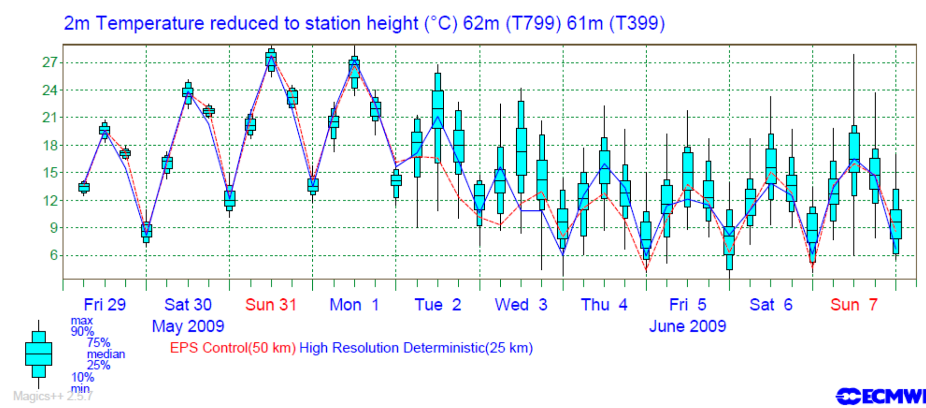
Types of forecasts, observations

- Probabilistic
  - Examples: Precipitation occurrence; wind speed category
  - Probability values may be limited to certain values (e.g., multiples of 0.1)
- Ensemble
  - Multiple iterations of a forecast (e.g., multiple model runs with different initial conditions)
    - May be transformed into a probability distribution using statistical methods

Each type of forecast (continuous, categorical, probabilistic) requires a different set of metrics



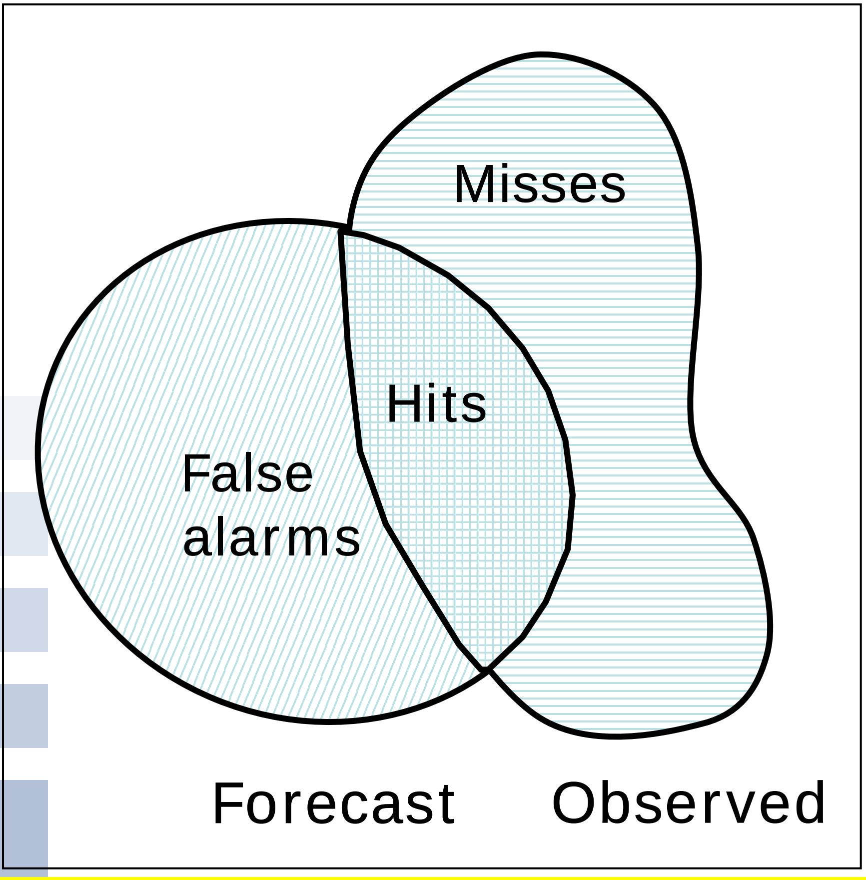
2-category precipitation forecast (PoP) for US



ECMWF 2-m temperature meteogram for Helsinki

Spatial forecasts and observations: Traditional spatial verification measures

NCAR



Could be applicable to fire spread predictions

Perfect forecast requires exact overlap!

Contingency Table

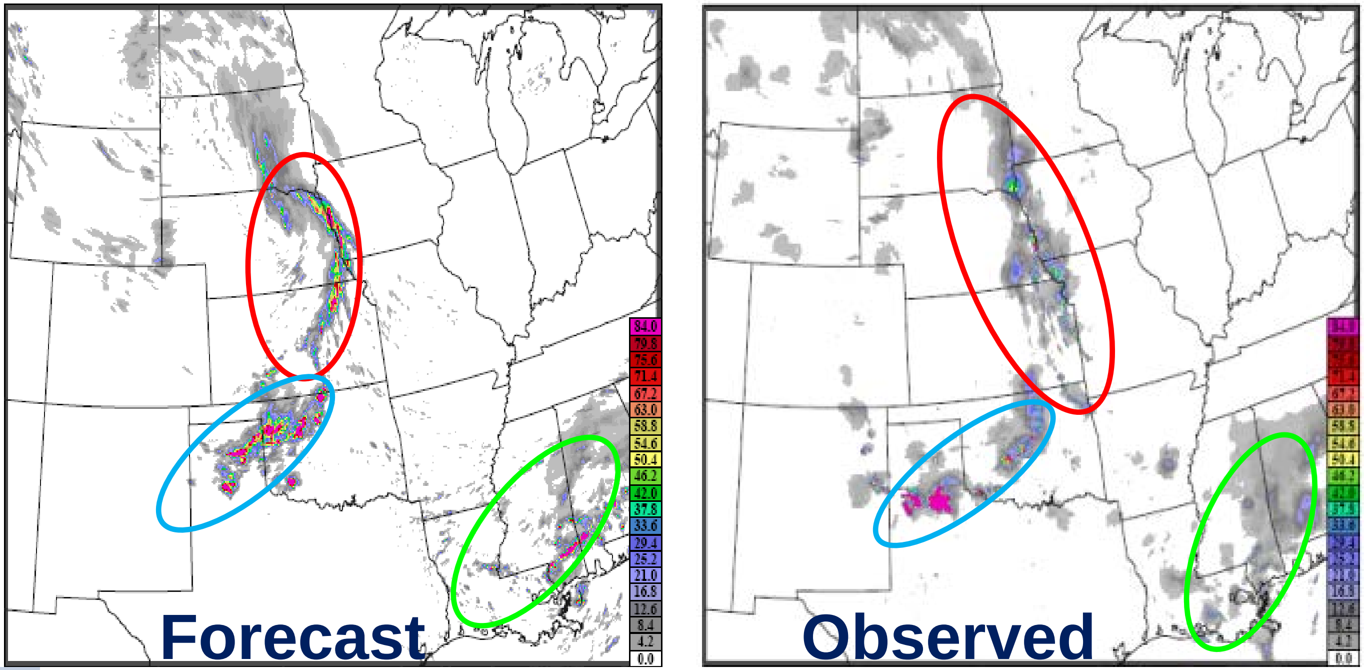
Forecast \ Observed	yes	no
yes	*hits*	*false alarms*
no	*misses*	*correct negatives*

Basic methods:

1. Create contingency table by thresholding forecast and observed values
   - *Compute traditional contingency table statistics: POD, FAR, Freq. Bias, CSI, GSS (= ETS)*
2. Directly compute errors in predictions
   - *Compute measures for continuous variables*: MSE, MAE, ME

Subjective and Spatial Approaches

NCAR



Good forecast or bad forecast?

Traditional results:
POD = 0.40 (best = 1)
FAR = 0.56 (best = 0)
CSI = 0.27 (best = 1)

- Traditional approaches indicate it is not a very good forecast
- Small errors in *location* or *magnitude* lead to poor scores. Methods for evaluation are not diagnostic – don't tell us what was good or bad

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 = 0.56
Rotated 15° (angle diff = 15%)	CSI = 0.27
Peak Rain 1/2" too much (diff in 90th percentile of intensities = 0.5)	

Selecting the best verification approach and metric depends on what we want to learn about the forecasts and how we use them to make decisions

New Spatial Verification Approaches

Neighborhood

Successive smoothing of forecasts/obs

Gives credit to "close" forecasts

Scale separation

Measure scale-dependent error

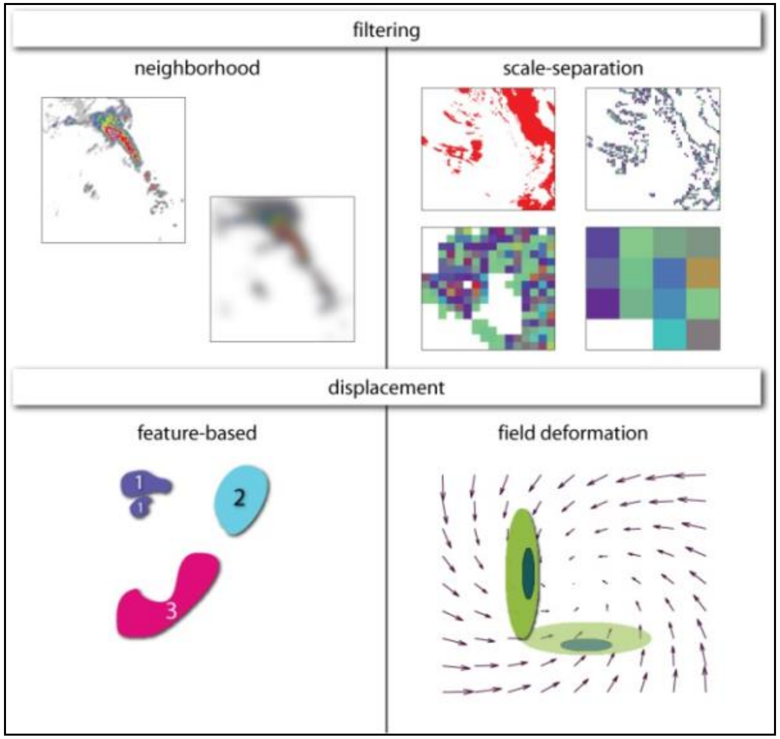
Field deformation

Measure distortion and displacement (phase error) for whole field

- *How should the forecast be adjusted to make the best match with the observed field?*

Object- and feature-based

Evaluate attributes of identifiable features



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

Object/Feature-based Approaches

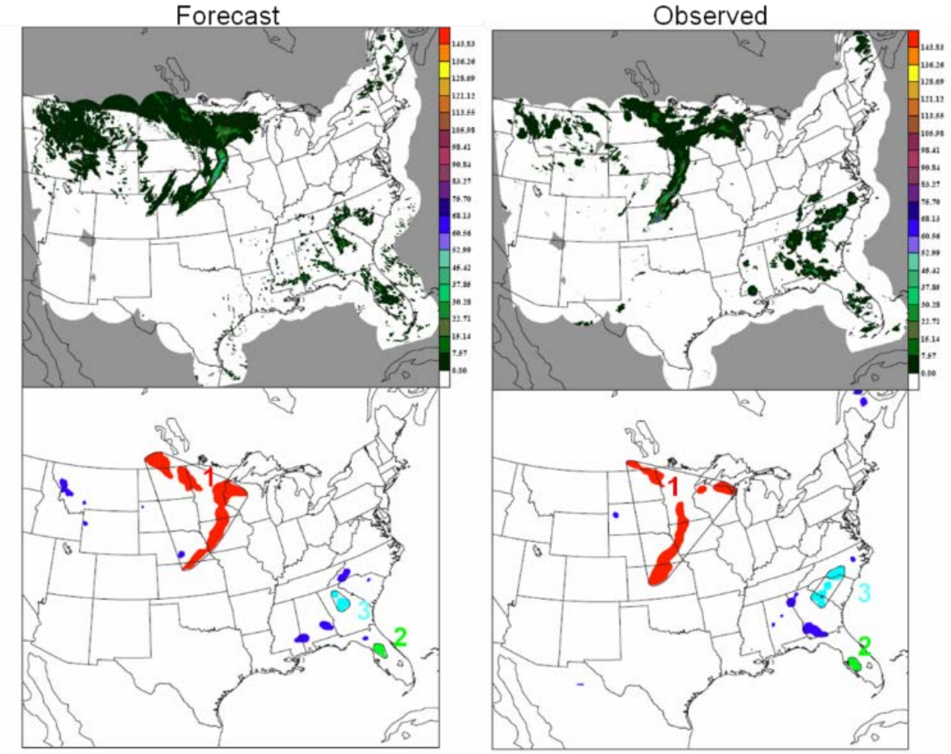
NCAR

Goals:

1. Identify relevant features in the forecast and observed fields
2. Compare *attributes of the forecast and observed features*

Examples:

- MODE
- Procrustes
- Contiguous Rain Area



MODE example (precipitation)

Part 2

PRIORITIES FOR TESTING AND VERIFICATION OF CO-FPS PRODUCTS

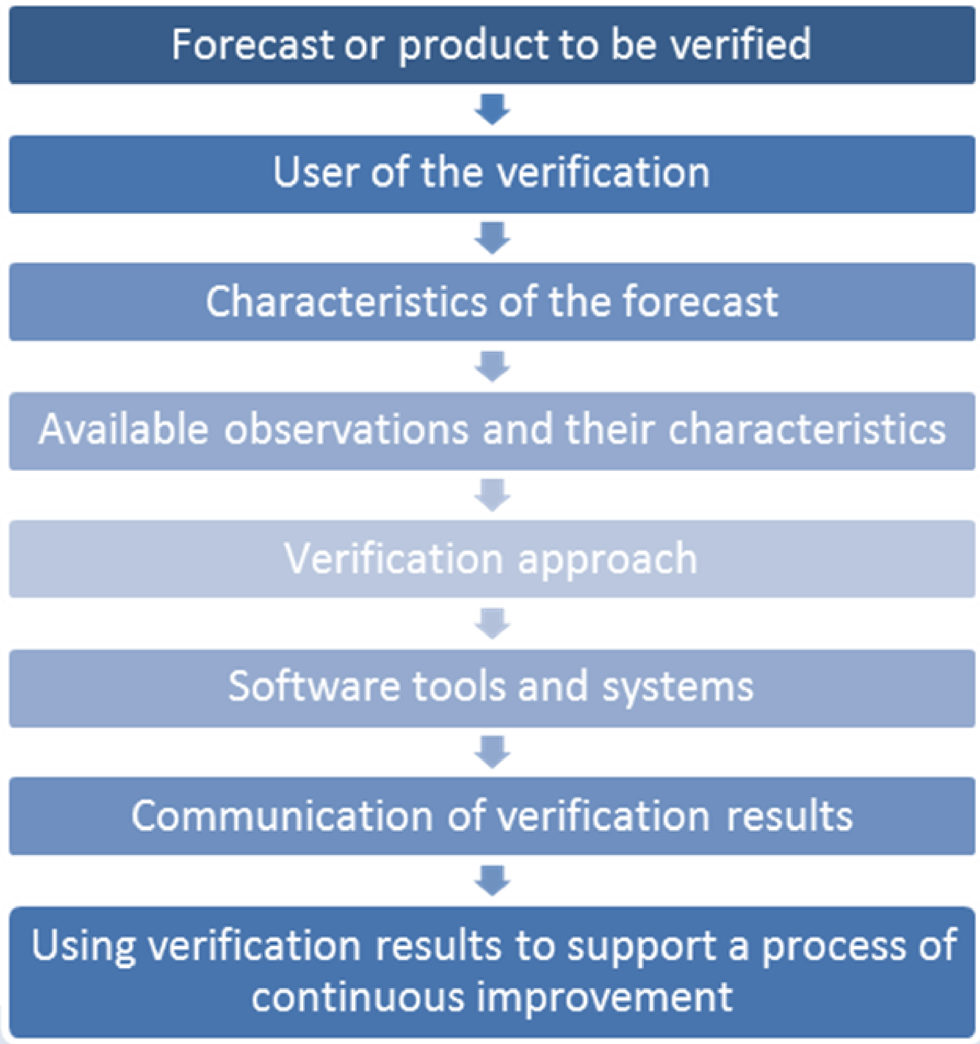
Topics

- Process for identifying metrics for
  - Supporting users
  - Determining success
  - Continuous improvement
- Examples for discussion:
  - Fire polygons
  - Fire spread – location of fire line

Australian project on evaluation of fire spread models

- Collaboration among Bureau of Meteorology (BOM) and various fire-fighting organizations (e.g., New South Wales)
- Setting out goals for metrics
  - Consider multiple aspects of forecasts, observations, and their application
  - Work closely with stakeholders
- Sharing ideas and information with us
  - We hope to leverage this collaboration

Example: Model for designing verification of routine forecasts and products



Courtesy, BOM

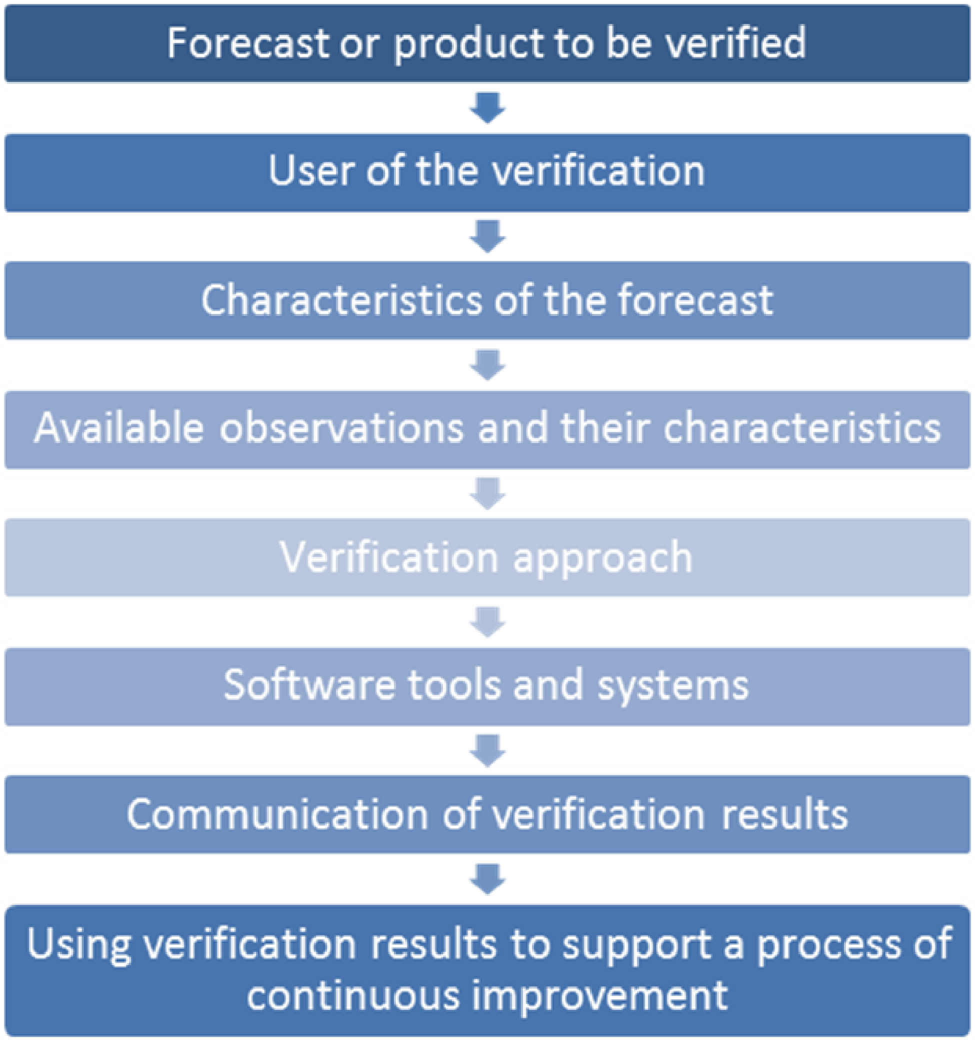
Model for designing verification of routine forecasts and products

NCAR

- Forecast or product to be verified
- ↓ User of the verification
- ↓ Characteristics of the forecast
- ↓ Available observations and their characteristics
- ↓ Verification approach
- ↓ Software tools and systems
- ↓ Communication of verification results
- ↓ Using verification results to support a process of continuous improvement

Who? How are they used? What decisions?

Model for designing verification of routine forecasts and products

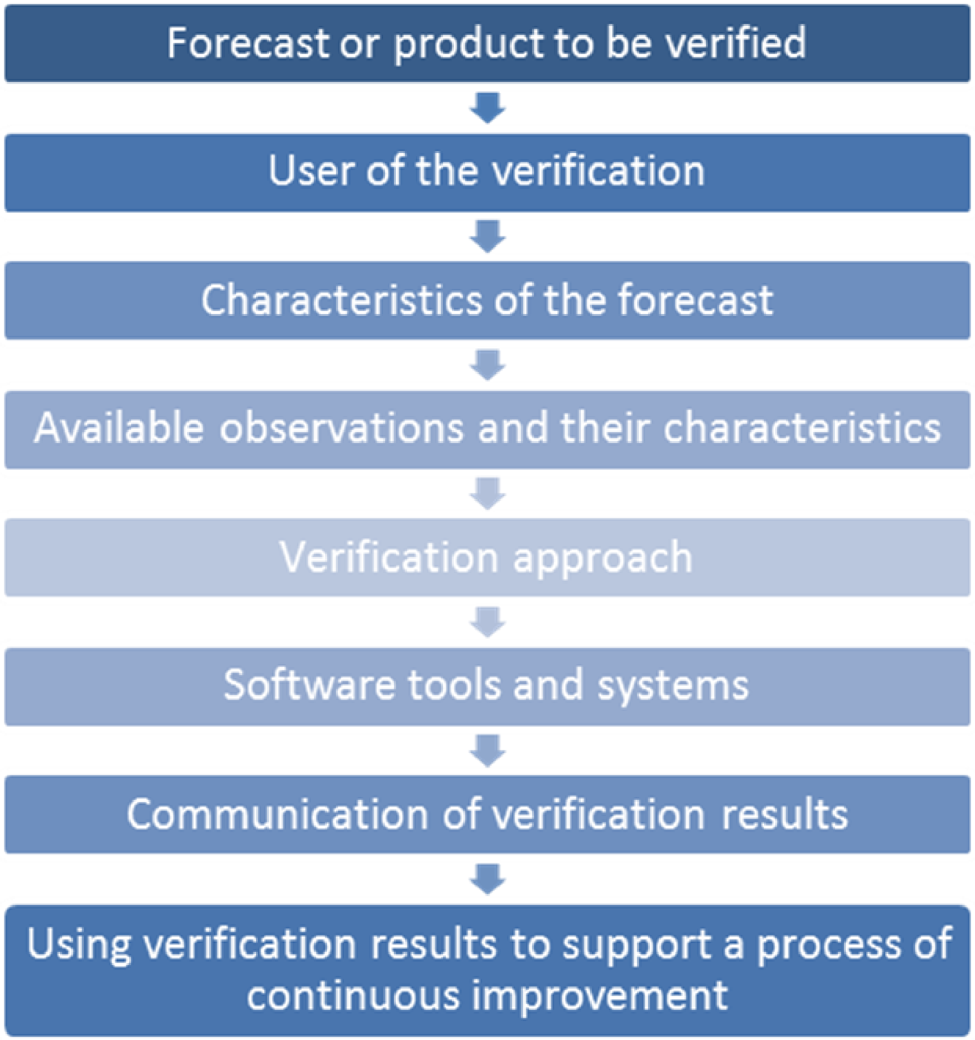


Who? How are they used? What decisions?

What characteristics are important?

Model for designing verification of routine forecasts and products

NCAR



Who? How are they used? What decisions?

What characteristics are important?

Which characteristics can be measured?

Model for designing verification of routine forecasts and products

NCAR

- Forecast or product to be verified
- User of the verification
- Characteristics of the forecast
- Available observations and their characteristics
- Verification approach
- Software tools and systems
- Communication of verification results
- Using verification results to support a process of continuous improvement

Who? How are they used? What decisions?

What characteristics are important?

Which characteristics can be measured?

What approaches can answer the questions?

Model for designing verification of routine forecasts and products

NCAR

1. Forecast or product to be verified
2. User of the verification
3. Characteristics of the forecast
4. Available observations and their characteristics
5. Verification approach
6. Software tools and systems
7. Communication of verification results
8. Using verification results to support a process of continuous improvement

Who? How are they used? What decisions?

What characteristics are important?

Which characteristics can be measured?

What approaches can answer the questions?

Identify graphics and tools to communicate to users

Model for designing verification of routine forecasts and products

Forecast or product to be verified

↓

User of the verification

↓

Characteristics of the forecast

↓

Available observations and their characteristics

↓

Verification approach

↓

Software tools and systems

↓

Communication of verification results

↓

Using verification results to support a process of continuous improvement

Who? How are they used? What decisions?

What characteristics are important?

Which characteristics can be measured?

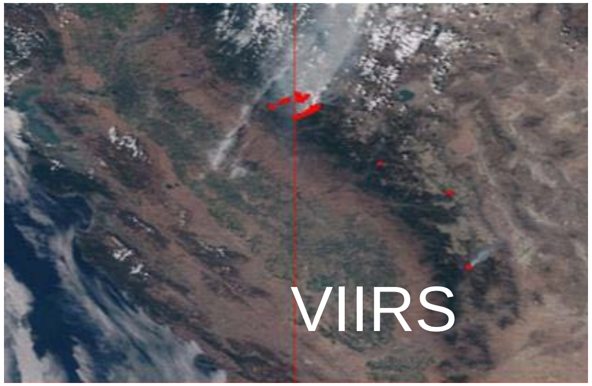
What approaches can answer the questions?

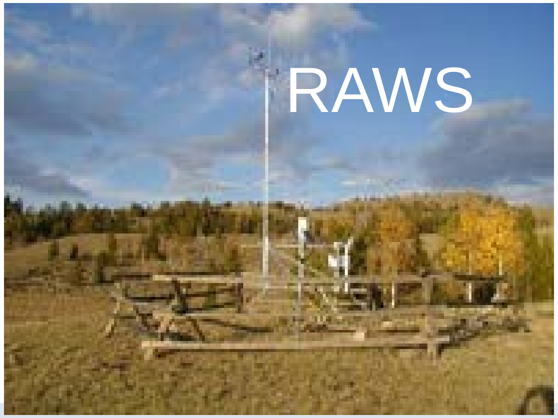
Identify graphics and tools to communicate to users

Work with team to identify needed improvements

Observations

- Appropriate observations are *key* to being able to do meaningful verification
  - Observations limit what we can verify – we can't verify things we can't observe!!
- Never forget: Observations have associated uncertainty, which impacts verification
- Fortunately, we have some pretty good obs





Examples for group discussion: Fire spread and fire location

- Users:
- Important characteristics:
- Observations:
- Verification approach:

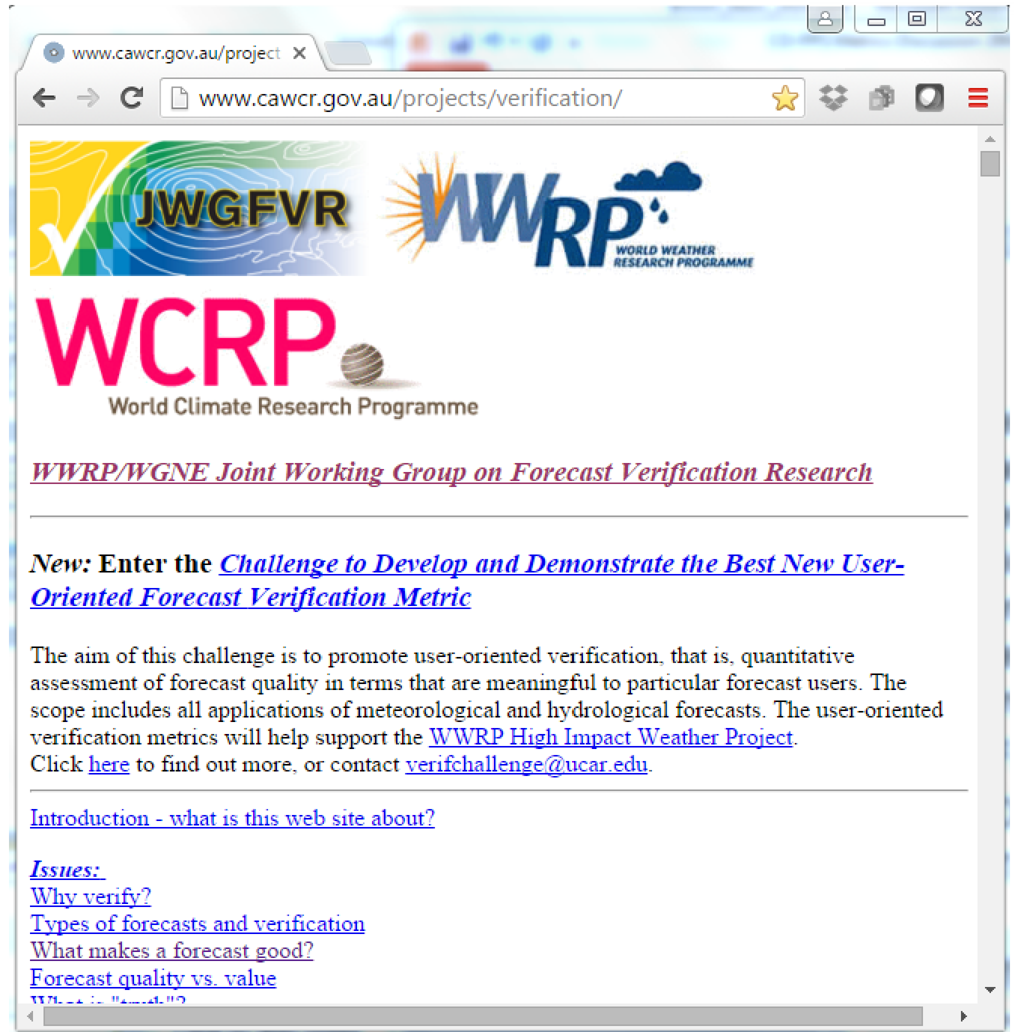
Examples for group discussion: Fire spread and fire location

- Users:
- Important characteristics:
- Observations:
- Verification approach:

Resources

Web page with many links to presentation, articles, etc. from international community

- FAQs
- Definitions
- Tools



http://www.cawcr.gov.au/projects/verification/

Summary

- Metric selection and verification planning is a collaborative process
  - Highly dependent on
    - Forecasts
    - Users
    - Applications
    - Observations
- Verification is an ongoing process, not a single step at the end of development
  - Start early and follow *through* forecast development, providing intermediate feedback
  - Continue through lifetime of forecast system