Verification package for “R”

Matt Pocernich

pocernic@ucar.edu

Research Application Program
verify: Verification Toolkit

Purpose:

- To verify a variety of types of forecasts with different kinds of observations. These use standard as well as new verification techniques.
- Forecast formats differ. There are continuous, probabilistic, binary, ordinal and nominal.
- Observations differ as well.
- Created initially for internal audience, hopefully useful to a larger audience. We’ve avoided weather specific references.

Features:

- Calculates a range of common skill scores.
- Diagnostic plots - attribute diagrams, receiver operating characteristic plots (ROC), conditional quantile plots.
- `verify` creates a `verify` class object, which utilizes methods.
**Functions**

attribute  Attribute diagram
brier  Brier score
crps  Continuous ranked probability score
int.scale.verify  Barbara Casati’s intensity scale verification model
leps  Linear error in probability space.
observation.error, measurement.error  Briggs’ measurement error score
reliability.plot  Reliability plot
roc.plot  Relative operating characteristic curve plot - with bi-normal distribution assumption and empirical distribution assumption.
verify  Principle package function.
verify Function

- `verify(obs, pred, tseries= NULL, baseline = NULL, 
  frcst.type = "prob", obs.type = "binary", 
  thresholds = seq(0,1,0.1))`

- Types of forecasts and observations include binary, probabilistic, 
  continuous, spatial?, distribution?.

- Depending on the type of forecast/obs, this function yields the 
  following information (Brier Score, Brier Skill Score, Skill Score, 
  Hit rate, False alarm rate, Threat Score or Critical Success Index, 
  Equitable Threat Score, Bias, Contingency Table, Heidke Skill 
  Score, Kuniper Skill Score, Mean error, Mean-squared error, Mean 
  absolute error, ...
Probabilistic Forecast Example

A <- verify(prob.frscs.dat$obs, prob.frscs.dat$frcst/100)
If baseline is not included, baseline values
will be calculated from the sample obs.
> summary(A)

The forecasts are probabilistic, the observations
are binary.
Sample baseline calculated from observations.
Brier Score (BS) = 0.1615
Brier Score - Baseline = 0.2251
Skill Score = 0.2824
Reliability = 0.004194
Resolution = 0.06528
Uncertainty = 0.2251
Forecast probability, $y_i$

Observed relative frequency, $\bar{o}_i$

No skill

No resolution

AWG Forecast

Attribute Plot
**ROC Plot**

**AWG Forecast**

Black lines are the empirical ROC. Red lines and symbols are the bi-normal ROC. The area under the binormal curve is in parenthesis.

Model A: 0.817 (0.809)

Verification package for “R” – p.7/18
Reliability Plot

AWG reliability plot

- Forecast probability, $y_i$
- Observed relative frequency, $\hat{o}_i$

Model 1
Comparing Multiple Models

- ROC plot will accept multiple forecasts.
- Attribute plot will accept multiple forecasts.
  - With two models, sharpness diagrams will be presented.
  - With more than two models, extra diagrams will not be shown.
- Output from `verify`, `roc.plot`, `attribute`, `int.scale.verify` can be directed to another object or output.
Conditional Quantile Plot for Continuous forecast/obs

Sample Conditional Quantile Plot

- Median
- 25th/75th Quantiles
- 10th/90th Quantiles

Forecast Value vs. Observed Value
Other contributed functions

- Measurement Errors
  - Developed by Matt Briggs - Cornell Medical School.
  - Used to calculate results presented earlier today.
- Continuous ranked probability scores
  - Algorithm contributed by Tilman Gneitting
Spatial Intensity Scale

• Developed by Barbara Casati
• Produces both scale intensity plots and mean square error plots. (Without Contours).
Scale-Intensity Plot

NIMROD example

Threshold

Spatial scale: $2^n$

Bias

Verification package for "R" – p.13/18
The use of R in the software development process

- Using packages increase the robustness of analysis. Major code reuse. Object oriented style.
- Simple to use the cvs version control system.
- Package updates go into the nightly build.
- Platform independent
Next Steps

- Incorporate a spatial class of verification routines.
- Improve methods for verifying forecasts expressed as distributions – particularly those expressed in forms other than the normal.
- Continue to solicit contributions from others.
- Write article in R Notes. Explore working in other topics in verification.
Other R packages developed at NCAR

- **fields** - spatial statistics tools - by Geophysical Statistics Project (Doug Nychka)
- **RadioSonde** - displays data from weather instruments - Tim Hoar and Eric Gilleland
- **extRemes** - graphical user interface for extreme value statistics - by Eric Gilleland (ismev by Stuart Coles)
- **rapRtools** - Internal set of functions containing functions such as wind roses and run computation tools.
For more information

www.r-project.org (See CRAN for package)
www.ucar.edu
www.cgd.ucar.edu/stats/
www.rap.ucar.edu/

Eric Gilleland ericg@ucar.edu
Doug Nychka nychka@ucar.edu
Matt Pocernich pocernic@ucar.edu
Statistics at the NCAR

NCAR Mission Statement:

“to support, enhance, and extend the capabilities of the university community, nationally and internationally; to understand the behavior of the atmospheric and related systems and the global environment; and to foster the transfer of knowledge and technology for the betterment of life on Earth.”