Forecast Verification in R

Andrea Ficchì
Politecnico di Milano, Italy (formerly, University of Reading, UK)

www.ei.deib.polimi.it/

@andrea_ficchi

Forecast Verification in R

Why?

1. Verifying hydrological forecasts is essential for researchers, developers and forecast users, to understand the skill of the forecast system and to build confidence in its results. Applications include (but are not limited to) water management, disaster risk reduction (e.g. Forecast-based Financing), hydropower, insurance, etc.

2. R provides some packages for forecast verification, with several useful functions - more than other languages (R is one of the most used languages for this scope).

Plan of the presentation:

- brief introduction on probabilistic forecast verification
- highlights on a few R packages available on CRAN and some key functions, quite known in the hydrology community but not easily findable by newcomers (several similar packages with slight differences and a review is missing, e.g. in the Hydrology taskview)

Forecast verification: how?

The hydrological literature traditionally focused on *deterministic* forecasts but in the last three decades *ensemble* forecasts have become increasingly popular and are nowadays the standard in weather and climate prediction

Ensemble Forecasts need to be assessed using probabilistic scores

Deterministic and probabilistic forecast verification: two different problems

Deterministic verification is based on simpler accuracy scores (like MAE, NSE, KGE, etc), more well-known

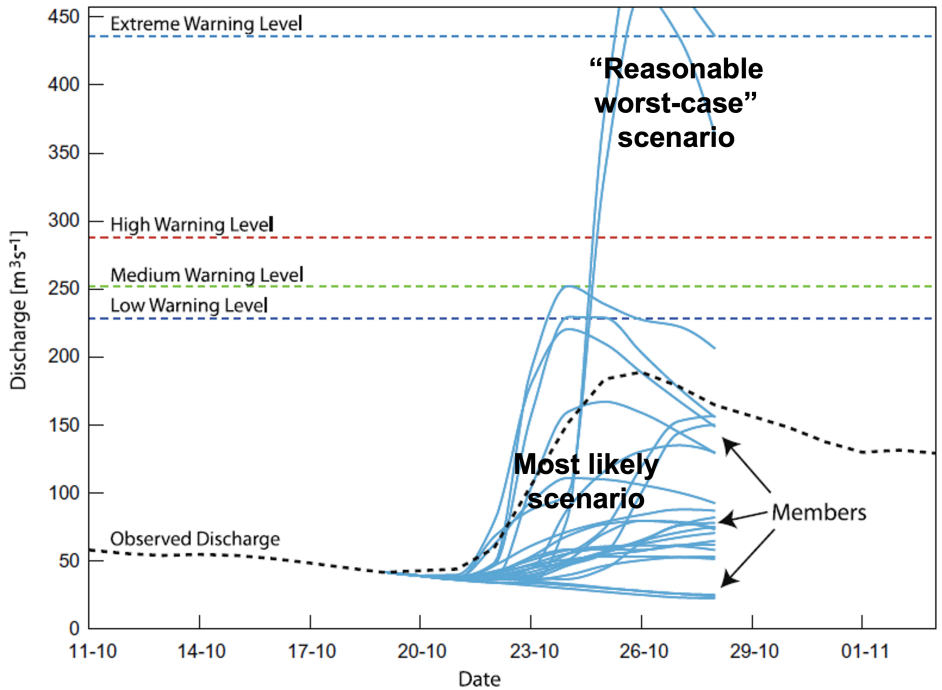
$\rightarrow$ in this presentation, we will focus on ensemble verification metrics

$\rightarrow$ First question: what is an ensemble forecast and how we can verify it?

Ensemble Forecasts

What is an Ensemble?

- An *ensemble forecast* (EF) is a set of forecasts that present the range of future weather possibilities
- Multiple simulations/forecasts are run, each with a slight variation of initial conditions and/or slightly perturbed models
- EFs show the range of possible outcomes and their likelihood over the forecast horizon



Further reading:

- Cloke, H., Pappenberger, F. (2009). "Ensemble flood forecasting: A review", Journal of Hydrology

Ensemble Forecast Quality Attributes

How can we verify an Ensemble? Need *large number of paired forecasts and observations* (Hudson, 2017)

There are several attributes of forecast quality:

- *Accuracy*
- *Reliability*
- *Discrimination*
- Resolution
- Sharpness
- Consistency in time (non-jumpiness)
- *Skill*

Forecast quality vs. value

- To assess value, it's essential to consider how forecasts are used
- User oriented skill verification --> value

R packages for Forecast Verification

List of packages for probabilistic forecast verification in R

- **verification**
- **easyVerification**
- SpecsVerification
- scoringRules
- ROCR

Other useful packages (support):

- ncdf4 (interface to process Unidata netCDF v4 Format Data Files)
- ecmwfR (downloads of forecasts and reanalysis data from ECMWF & CDS)
- hydroGOF (goodness-of-fit measures between observed & deterministic forecasts/sim.)
- other packages for hydrological analysis (see Hydrology taskview)

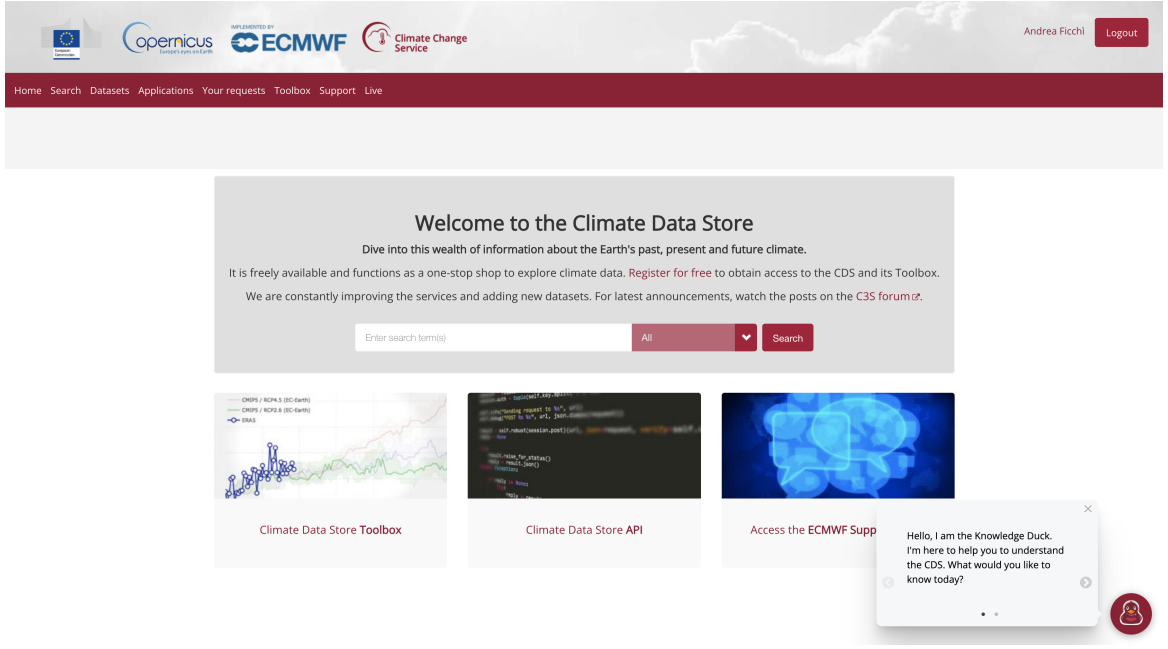
→ What can be done already by these packages? (*presentation*)

→ What cannot be done yet that would be useful? (*discussion*)

Data Availability

Global observed data can be obtained for thousands of stations from the Global Runoff Data Centre (GRDC) -- see the portal (https://www.bafg.de/GRDC/EN/) and the presentation by Louise Slater from the 2021 Short course.

Global forecasts are available from the Copernicus Climate Datastore.



'Verification' package

The 'verification' package can be used for the computation of:

- Continuous Ranked Probability Score (CRPS)
- Brier Score (BS)
- Receiver Operating Characteristics (ROC) & Area Under Curve (AUC)
- Reliability plot
- False Alarm Ratio & Probability of Detection
- and a few other scores

--> Let's install and load the package in R:

```
install.packages("verification")
library(verification)
```

--> Then, let's prepare the inputs to the functions; first, you can look at an example in the Short-course Github repository, then move to your local observations and forecast

Verification: input data preparation

Preparation of input data:

obsdata_v = Vector with observations (e.g. time series of river flow, m3/s)

ens_fcst_array = Array of forecast values, at least 2-dim (time, member)

Dimensions of *ens_fcst_array* (in the example):

Lead times x *forecast issue dates* x *ENS members*

Some functions require these derived variables (arguments):

obs = a binary observation (coded {0, 1 }) based on threshold exceedances

pred = a probability prediction on the interval [0,1].

Example: Area under ROC curve: roc.area (1/3)

Options and preparation of 'roc.area' function arguments (obs, pred):

```
lt_test <- 15 # Lead time of analysis
sel_pctl <- 90 # Percentile for flood event definition
Num_dates_2use <- dim(ens_fcst_array)[2] # Number of forecast dates to use (here all)
max_act_lt <- dim(ens_fcst_array)[1] # Max actual lead time to use
Min_N_ts_eval <- 30; min_N_obs_evts <- 10 # min number of time steps & events required
thres_frcst_ref <- "LT_depend"

thres_obs <- quantile(obs_m, probs=sel_pctl/100, na.rm=TRUE) # Compute percentile from obs
binobs <- matrix(nrow=Num_dates_2use,ncol=max_act_lt,data=0) # Create array for binary event/non-

#Loop over obs and if it exceeded the threshold, set to 1
for(i in 1:Num_dates_2use){
   for(j in 1:max_act_lt){
      if(is.na(obs_m[i,j])){
         binobs[i,j] <- NaN
      } else if(obs_m[i,j]>=thres_obs){
         binobs[i,j] <- 1
      }
   }
}
```

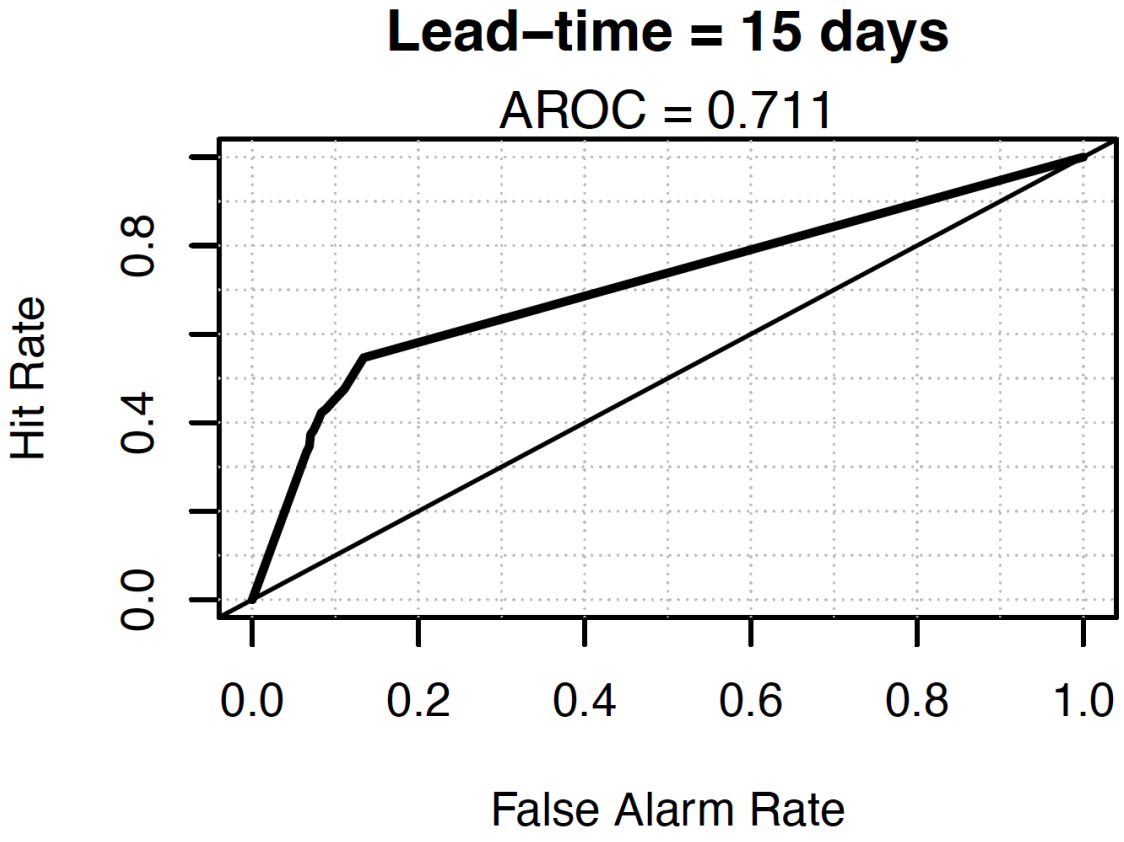
Example: Area under ROC curve: roc.area (2/3)

```
# Calculate % of forecast members exceeding the threshold
if(thres_frcst_ref == "LT_depend"){
  ## additional index j to access the specific lead-time based threshold thres_fcst[j+1]
  for(i in 1:Num_dates_2use){for(j in 1:max_act_lt){pfcst_obs[i,j] <- sum(ens_fcst_array[j, i, 1:
} else {
  for(i in 1:Num_dates_2use){for(j in 1:max_act_lt){pfcst_obs[i,j] <- sum(ens_fcst_array[j,i,1:EN
}

# Calculate of Area Under ROC (if there are enough observations)
if(sum(is.finite(binobs[,lt_test]))>Min_N_ts_eval & sum(binobs[,lt_test]==1, na.rm=T)>=min_N_obs_
   auc_obs <- roc.area(binobs[,lt_test],pfcst_obs[,lt_test])
} else {
   auc_obs <- NA
}

# Plot ROC curve (if there are enough observations)
if(sum(!is.na(pfcst_obs))>0 & sum(is.finite(binobs[,lt_test]))>Min_N_ts_eval & sum(binobs[,lt_tes
               roc.plot(binobs[,lt_test],pfcst_obs[,lt_test],thresholds=seq(0,1,0.1),show.thres=
}
```

Example: Area under ROC curve: roc.area (3/3)



False Alarm Ratio & Probability of Detection, basic

A contingency table is built from paired binary observations and binary forecast events. For the latter, one needs to fix a 'trigger' probability (that should be tailored to user preferences).

Example of calculation:

```
# Using function 'table.stats' in the 'verification' package, we can calculate
# discrimination statistics for a 2 by 2 Contingency Table
# obs and pred = vector of binary observations &  vector of binary forecasts

lt_test <- 15 # Lead time of analysis

trigger_prob <- 70  # Definition of the 'trigger probability' threshold [%]
bin_fcst_lt_j <- pfcst_obs[, j_lt_eval] >= trigger_prob/100

Stats_conting_tab <- table.stats(obs=binobs[,lt_test], pred=bin_fcst_lt_j, fudge = 0.01, silent =

FAR_Tstats_lt_j <- Stats_conting_tab$FAR
POD_Tstats_lt_j <- Stats_conting_tab$POD
```

Then, this calculation can be performed in a loop for all lead times and a standard plot of the score evolution with lead time can be generated

'easyVerification' package

The ”easyVerification" package can be used for the computation of:

- Brier Scores
- CRPS (& fair CRPS)
- Reliability Diagram
- ...

Ease of use: Absolute forecasts and observations are converted to category and probability forecasts based on the threshold or probability provided, and ouputs are reformatted to fit the input.

The CRPS function from the package “easyVerification” has the additional functionality to correct for different ensemble members, calculating the so-called **fair CRPS**, which is not possible in the 'verification' package.

--> Let's install and load the package in R:

```
install.packages("easyVerification")
library(easyVerification)
```

--> Then, let's prepare the inputs to the functions; first, you can look at an example in the Short-course Github repository, then move to your local observations and forecast

'easyVerification' package: fair CRPS

fair CRPS

Set up the forecast and observation data structure, example:

```
tm <- toyarray(c(13,5), N=15, nens=51)
fo.crpss <- veriApply("EnsCrpss", fcst=tm$fcst, obs=tm$obs)
```

Conclusions

There are some R packages providing multiple commonly used ensemble verification metrics.

It is straightforward to calculate basic skill scores for a single site.

To extend this to many site at once, one can see the presentation by Louise Slater from the 2019 Short course.

Some scores can be adapted to be more user-oriented (e.g. Coughlan de Perez et al. 2016) and adjusted to consider lead-time dependent thresholds (Zsoter et al. 2020).

I have been working towards the integration of user oriented verification metrics into a new package. Keep an eye on CRAN and the Hydrology task force for any new packages.

Please email me if you discover any other datasets or packages, so we can keep updating this community help resources in the future (andrea.ficchi@polimi.it).

Further reading

The slides and materials are available on Github: https://github.com/hydrosoc/rhydro_vEGU22/

- Coughlan de Perez, E., van den Hurk, B., van Aalst, M. K., et al.: Action-based flood forecasting for triggering humanitarian action, Hydrol. Earth Syst. Sci., 20, 3549–3560, https://doi.org/10.5194/hess-20-3549-2016, 2016.
- Hersbach, H. (2000). Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, Weather and Forecasting, 15(5), 559-570. Retrieved May 25, 2022, from https://journals.ametsoc.org/view/journals/wefo/15/5/1520-0434_2000_015_0559_dotcrp_2_0_co_2.xml
- Hudson, D. (2017), Ensemble Verification Metrics, ECMWF Annual Seminar 2017.
- MeteoSwiss: easyVerification: Ensemble Forecast Verification for Large Data Sets, R package version 0.4.2, available at: http://CRAN.R-project.org/package=easyVerification
- Wilks, D. S.: Statistical methods in the atmospheric sciences, Academic press, Oxford, 2011.
- Zsoter, E, Prudhomme, C, Stephens, E, Pappenberger, F, Cloke, H. Using ensemble reforecasts to generate flood thresholds for improved global flood forecasting. J Flood Risk Management. 2020; 13:e12658. https://doi.org/10.1111/jfr3.12658

Thank you for listening!

If you have any question or comment, please do not hesitate to contact me.

andrea.ficchi@polimi.it