

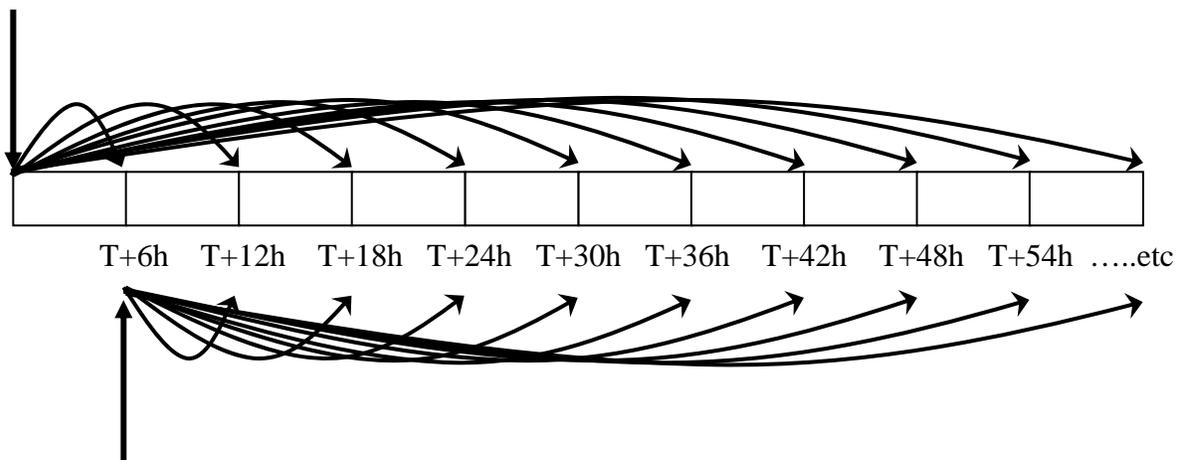
CBRFC Verification Workshop: ensemble verification lab sessions.

Session 1: exploring errors in synthetic temperature forecasts (20 mins)

This example deals with some synthetic temperature forecasts for summer 2006 from Las Vegas, Nevada. The forecasts begin on 1st May 2006 at 12:00am Pacific Standard Time (with daylight saving) and conclude on 30th September 2006 at 12:00am PDT. A forecast is made every 6 hours and an observation is available for every forecast time. At each forecast time, a forecast is produced for 10 days into the future at the 6-hourly timestep (i.e. a 10 'lead day', 6-hourly, forecast). Thus, a forecast produced on 08/02/06: 12:00am will conclude on 08/12/06: 12:00am. This is repeated for each new forecast timestep:

START (T): 05/01/06: 12:00am PDT

Issue 6-hourly forecasts for 10 days into the future.



START+6h: 05/01/06: 18:00am PDT

Issue 6-hourly forecasts for 10 days into the future. ETC....

The forecasts and observations are generated from a simple regression model, where we assume:

1. The forecasts and observations are normally distributed;
2. The forecasts are *unbiased in the mean and in the spread*;
3. The forecasts are *highly correlated* with the observations ($\rho = 0.95$);
4. There is no loss of skill with increasing lead time.

In other words, the forecasts are unusually good. You are now going to look at how these properties are reflected in some verification results. To begin with, you are going to work with an existing project in EVS for which pairs of forecasts and observations have already been generated.

Tasks to complete:

1. Each group has a specific working directory for the lab sessions under “/porky/verif_wkshop/EVS_lab/”. The EVS software and lab data are stored in a common directory. The working directory for your group corresponds to the number of your workstation (e.g. GROUP_1 for workstation 1). Navigate to the common software directory under the root directory “/porky/verif_wkshop/EVS_lab/COMMON/Software/”. Open EVS by typing **java -jar EVS.jar**.
2. To open an existing project, navigate to **File > Open project...** or use the shortcut in the menu bar. Navigate to the working directory of your group. Locate the project file for the first session “/Session_1/Vegas_temps.evs” and open it.
3. The EVS windows are separated into three sections: 1) Verification (for verifying one or more variables and at one or more locations); 2) Aggregation (for averaging verification results across multiple basins); and 3) Output (for generating plots etc.). You may now spend a few minutes looking through the first window in the **Verification** tab. This window shows all of the verification units currently loaded (i.e. one). A verification unit comprises a single time-series at a single forecast point (i.e. temperature in Las Vegas). A unit is selected in the table. When a unit is selected, its properties are shown. These include the source data, the verification period etc. Take a few minutes to browse through the various inputs. If you make a mistake and accidentally delete or modify some data *and save the changes*, back-up projects are located in the “/BACKUP/” directory underneath your group directory.
4. Navigate to the second window of the **Verification** tab by clicking **Next** in the lower right corner. The second window defines which metrics will be computed for the verification unit selected in the first window. In the table (top left), you will see two major groups of metrics: 1) metrics that test the ensemble mean forecast (single-valued metrics) and; 2) metrics that test the full probability distribution (ensemble metrics). When you highlight a metric in the table, some information about that metric is displayed in an adjacent window. When you select a metric (check the tick box), some parameters also appear. Default values are given. Spend a few minutes exploring the options.
5. We are going to generate three metrics for inspection: 1) The “Mean continuous ranked probability score”, the “Modified box plot” and the “Cumulative Talagrand diagram”. Select these metrics. Click the **Run** button. This will generate the numerical results from the metrics. These results will be used to generate verification products, such as written numerical output or some default plots to write or display.

6. Navigate to the **Output** tab by selecting this tab or clicking **Next** twice. You will see that a verification unit with results has been listed in the table. Highlighting a unit in the table will show a list of available results. One result exists for each metric computed. Highlighting the results of one metric will show the lead times (currently in hours) for which those results are available. Take a few minutes to explore these features (try right-clicking on the results tables also).
7. A product is generated by selecting one or more lead times for one or more metrics (checking the tick boxes). Once some metrics and lead times have been selected, there are two options for generating products. Clicking **Run** will generate products with default options (don't do this yet). The default options are to write graphics in PNG format and to write numerical results in XML format (to the output folder in the first window). Clicking **Run...** will provide some more advanced options and allow you to use an internal plotting tool within EVS. You will generate some products next.
8. Select the **Continuous ranked probability score**. Now, right click on the lead time table and **Select all times**. Recall that this measures the average squared deviation between the forecast, which is a probability distribution, and a single-valued observation. The size of the CRPS is determined by the ensemble spread and the distance of the forecast from the observation, increasing with both. In addition, select the box plot and again specify all lead times. The box plots will show the average distribution of errors from each lead time in the form of a box.

Finally, select the **Talagrand diagram** and select ONE lead time to plot (any one will do). Unlike the single-valued metrics (CRPS, mean error, correlation, and Brier score), the Talagrand diagram, among others, has one diagram per lead time. We are going to use the non-default **Run...** option. BEFORE clicking this, spend a few minutes thinking about what the plots might look like given the forecast scenario described above. The first two plots (mean error and correlation) will display those quantities with increasing lead time. The Talagrand plot will show one plot for the selected lead time. What would you expect?

9. Click **Run....** This will open a window of options. In the **Write** tab, select the **Write graphical output** option. Navigate to the **Display** tab and check the **Display graphical output** option and then click **Run**, which will show the plots in an internal viewer (and write them to file also).
 - a. Do the results appear as expected (tip: what would the errors of a good forecast look like)?
 - b. Why are the results not even better? For example, why is the line in the Talagrand plot not perfectly diagonal? Please ask questions if you have them....

Session 2: introducing some bias into the temperature forecasts (20 mins)

Using the same temperature scenario from session 1, you are now going to explore the impacts of two types of bias in the temperature forecasts, namely:

1. A bias in the ensemble mean forecast. This bias will increase with forecast lead time. This can sometimes happen in ensemble systems when forecasting biases compound and propagate through time (e.g. due to poor initial conditions in a model). Here, a bias will be added to the forecast mean, $E[Y]$, such that $E[Y] = E[Y] + 0.05 * lead_hours$ (i.e. a bias equal to 5% of the lead time).
2. A bias in the spread of the ensemble forecasts. Specifically, the forecast spread will be too small. This is common in ensemble systems because it is difficult to account for all sources of uncertainty. This will be a constant bias, whereby the forecast spread is 50% smaller than the spread of the observations (i.e. the climatological distribution).

Note that these two types of bias will be explored separately. In reality, it is common for multiple sources of bias to interact.

Again, you will work with a predefined project in EVS. In the final exercise, you will set-up your own project.

Tasks to complete:

1. Once in EVS, navigate to the “/Session_2” directory underneath your working directory and open the EVS project named “Vegas_temps_biased.evs”. You will notice that this project contains two ‘verification units’. One unit contains the ensemble forecasts with a biased mean, and one contains the ensemble forecasts with a biased spread. Spend a few minutes looking through the inputs by selecting each verification unit in turn.
2. Select the first verification unit (biased mean) in the table and then move to the second window in the **Verification** tab. Select the required metrics for generation (see Session 1). Repeat this for the second verification unit (biased spread), selecting the unit in the first window and then checking all required metrics in the second window. Now click *******All******* to compute the required metrics for all units in the project.
3. Navigate to the **Output** tab where you should find two verification units with results. Now, generate the same plots you generated in the first Session. You can do this quickly for each unit by right-clicking on the product table and selecting **All lead times and products**. Once you have selected the output, you may use the default **Run** option to generate the products. These plots should appear in the output directory specified in the first window.

4. Open some of the plots from the previous three exercises (unbiased mean, biased mean and biased spread).
 - a. Do the results appear as expected? For example, can you tell that the forecasts with biased mean deteriorate in quality with lead time?
 - b. Why are the values of CRPS much larger for the scenario with the biased mean (the box plot may help you understand this)?
 - c. Why do the Cumulative Talagrand diagrams for the longer lead times contain no data points at all?

Session 3: verifying some real temperature forecasts (30 mins):

This session will build upon the previous two exercises by verifying some real temperature forecasts from Newport, PA in the MARFC. In general, temperature forecasts are much better behaved than precipitation or streamflow forecasts, as hydrological models include uncertainties from a wide range of sources (including meteorological inputs). The final exercise will focus on streamflow forecasts from the same basin.

Again, an EVS project has been partially defined for you.

Tasks to complete:

1. Once in EVS, navigate to the “/Session_3” directory and open the EVS project named “Newport_temps.evs”. Notice that the timezone for the observations is CST (GMT-6hrs) rather than the anticipated EST (GMT-5hrs) for Newport, PA. This is because the observations were actually collected one hour after the forecasts. This is an approximation necessary for pairing the forecasts and observations (i.e. an assumption that a one hour delay is not too important).
2. EVS allows you to target specific sets of forecasts and observations for verification purposes. This is known as conditional verification. Two types of conditions are possible (they may be combined), namely: 1) date conditions; and 2) conditions based on variable value. The latter include conditions based on forecast values (e.g. select all temperatures where the ensemble mean > 90° F) and conditions based on observed values (e.g. select all temperatures where the observed value < 32°F, 0°C). The date conditions currently allow a subset of days, weeks, months or years to be selected in the overall verification period.
3. In this example, you are going to verify ‘cold-season’ temperatures from 2005, which are taken to be November – April inclusive. Note that you might have decided to select forecasts with low temperatures rather than specific months, but there are some interesting patterns to explore in the full daily temperature cycle. In the first window of the **Verification** tab, click the **More** button in the bottom right corner of the **Verification period** section. The resulting dialog provides options for conditional verification. The first window allows conditioning based on dates. Select the **Months of year** category in the left table. Now, deselect any months that do not fall within the ‘cold period’ of November-April inclusive. Click **OK** to accept these changes.
4. Generate some verification results and inspect the plots:
 - a. Are the temperature forecasts biased in any way, and if so, how?
 - b. What in particular do you notice about the plots showing a range of lead times? What might cause these error patterns?

Session 4: verifying some real streamflow forecasts (30 mins):

In this exercise, you will verify streamflow forecasts from Newport, PA, for the period 01/01/2005 to 12/31/2005. Unlike the previous exercises, you will define your own project by specifying all of the input parameters required to verify the forecasts. In addition, you will aggregate the forecasts and observations from the actual 6-hourly times to daily values. Aggregation is particularly useful for verifying long-term forecasts, where hourly or even daily fluctuations are not important for decision making purposes (indeed, you don't want them to influence the verification results). In EVS, you may aggregate forecasts and observations from any forecast frequency to any longer period, providing there are sufficient data to do so.

Tasks to complete:

1. If you have an earlier project open in EVS, create a new project for this exercise (**File > New project...**). Navigate to the first window in the **Verification** tab. You can add a new verification unit with some default parameters by clicking **Add** in the lower left corner of this window. Rename the unit to something more appropriate by changing the component identifiers (e.g. "Newport" for both the basin and time-series IDs and "Streamflow" for the variable ID: note that default variable IDs are provided by right-clicking on the variable ID box). You may save changes to a project at any time.
2. You will need to specify the source of the forecasts and observations and the time systems in which they are stored. All data currently originate from file. The forecast files are located in "/COMMON/Data/Session_3/Newport_ESP". You can specify the location manually (being sure to include the full path from the root directory) or you can use the file system browser, which is accessed through the adjacent button. Select either the folder or all of the forecast files within the folder. Specify **UTC - 12 hours** for the forecast time system. Then specify the path to the observations, namely: "/COMMON/Data/Session_3/Newport_Obs/Newport.MAP06.obs". Specify **UTC + 6 hours** for the observed time system.
3. Enter the start and end dates for the verification. Note that a Calendar dialog is provided in the buttons adjacent to the date fields. The dates need not match the required verification period exactly, but should enclose this period. Enter 14 days for the forecast lead period.
4. As indicated above, the forecasts and observations will be aggregated from 6-hourly values to daily values. Enter 1 day for the aggregation period.
5. Select all of the available metrics and generate the verification results. You will receive a warning message here because no scale information has been associated with the forecasts and observations (this can be done in the first

window, but we did not do it). For example, if the forecasts and observations were measured in different units, we could not conduct the verification without changing the units first. Click OK because they are in the same units (you simply did not enter them).

Note that most of the processing time is spent reading the forecasts from file. One advantage of starting EVS from a command prompt is that you can view the stage of processing via the messages written to standard out. Once the forecasts have been read and paired with the observations, the time required to generate the verification results is greatly reduced (try by clicking **Run** again if you're interested).

6. Explore the verification results:
 - a. Notice that there are now 24 hours between forecast lead times instead of 6 hours. Why is the first lead time 15 hours?
 - b. Examine the box plot of forecasting errors for all forecast lead times. What type of forecasting bias is present?
 - c. Notice from the box plots that the ensemble spread is very low at small forecast lead times. Why does this impact the reliability of the forecasts as shown in the Talagrand diagram (e.g. compare the Talagrand diagrams at lead time 15 and lead time 327)?