P2.36 USING BOI_VERIFY TO IDENTIFY MODEL PREFERENCES IN LARGE TEMPERATURE-CHANGE EVENTS

Leslie R. Colin *
NOAA/NWS, WFO Boise, ID

## 1. INTRODUCTION

Several years ago the National Weather Service (NWS) converted from text-based to graphical-based forecasting. NWS field offices now routinely prepare forecasts for their domains in gridded form and these grids can be fitted together to make a national mosaic. The gridded approach allows forecasts to be verified numerically. In 2006, Tim Barker (NWS WFO Boise, ID) created a framework for archiving model and human gridded forecasts of temperature, probability of precipitation (PoP), wind, and other elements, as well as observed grids of those elements into NWS' GFE (Graphical Forecast Editor) environment. This framework is called BOI_Verify and it is used at other NWS field offices besides WFO Boise.

The NWS is especially interested in verifying high-impact events. One type of high-impact event is large day-to-day temperature change. This article describes a way that BOI_Verify has been used to compare the forecast accuracy of various model grids, with the results sorted according to the magnitude of the day-to-day temperature changes.

## 2. VERIFICATION METHODOLOGY

The GUI shown in Fig 1 accesses the BOI_Verify archives to verify the accuracy of various models in forecasting Maximum temperature (MaxT) and Minimum temperature (MinT) for various forecast periods. NWS field offices produce two forecast cycles per day, one during the night shift (00Z cycle) and the other during the day shift (12Z cycle). The night shift begins with MaxT for the coming day, which makes it a period 1 MaxT. Period 2 is the MinT for the next day, period 3 is MaxT for that day, and so on. The day shift begins with MinT for the next day, making it a period 1 MinT. This is followed by a period 2 MaxT for that day, then a period 3 MinT for the following day, and so on.

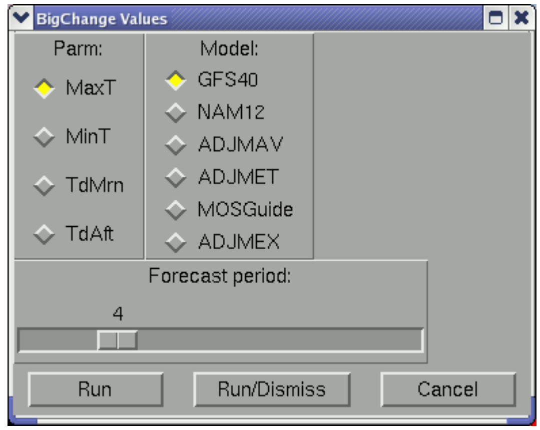


Fig. 1. Tool GUI

Here we pick the GFS40 model MaxT for period 4, i.e., produced on day shift (12Z cycle) for two days hence. The verification domain is the area served by WFO Boise, ID, which includes part of southwestern Idaho and southeastern Oregon (Fig. 2). The grid mesh is 2.5 km creating a total of 22199 points in the domain. (Actually, the hatched area can be changed to any area, but for this article it will remain the WFO domain)

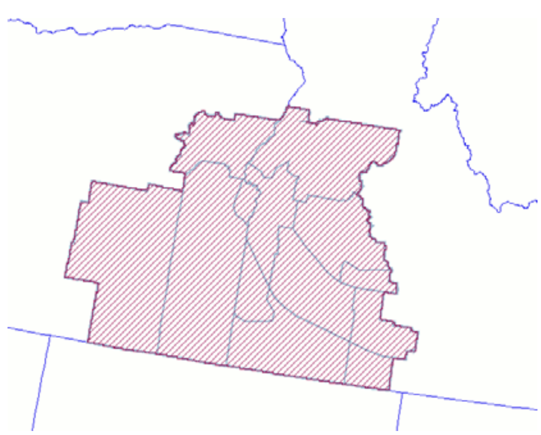

Fig. 2. Geographical domain of WFO Boise, ID.

---

* author e-mail address: les.colin@noaa.gov

When the GUI is activated it retrieves the past 40 days of archived grids for GFS40 4th period MaxT, together with observed MaxT for the same dates. The observed MaxT grids are then sorted according to the largest overall MaxT change from the previous day. The date with largest cooling from the day before is ranked #1, and the date with largest warming from the day before is ranked #40.

Then, for each date in the ranked list, the GFS40 4th period MaxT is verified (forecast minus observed) at all 22199 points of the domain, and the mean error for each date is recorded.

The results are plotted in Fig 3, below.

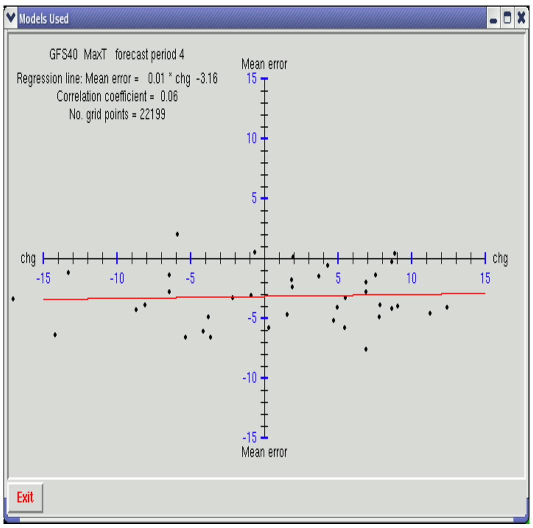


Fig. 3. Mean GFS40 4th period MaxT forecast-error vs observed day-to-day MaxT change.

The x-axis represents the average change in MaxT from the day before, so a dot on the left side of Fig. 3 means cooling from the day before. The y-axis shows the mean forecast error made by 4th period GFS40 MaxT for each date. In this case the GFS40 shows a -3.16 F degree bias, i.e., it usually forecasts too cold. But the near-zero slope of the regression line shows that it doesn't matter whether a given day has cooled or warmed from the day before; the GFS40 still averages about 3.16 F degrees too cold for 4th period MaxT.

Fig 4 shows how the GFS40 performs on 12th period MaxT:

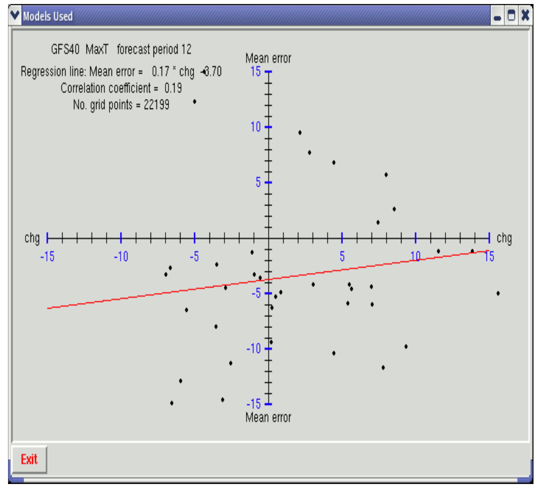


Fig. 4. Same as Fig 3 but for 12th Period GFS40 MaxT.

Again the GFS has a cold bias, but the regression line has a positive slope, which means that after removing the bias, the GFS40 over-forecasts the magnitude of the changes for 12th period MaxT.

The NWS produces MOS statistical forecast data for both the GFS and the NAM models at many locations throughout the U.S. The GFS MOS is called MAV, while the NAM MOS is called MET. These data are analyzed into the GFE by a process called smart-initialization. On the GFE domain the MAV is called ADJMAV, while the MET is called ADJMET. The ADJMAV runs through period 5, but the ADJMEX (which is also based on the GFS) runs through period 14. Due to model inconsistency at later time periods all the MOS data tend increasingly toward climatology, which means they are less responsive to large day-to-day changes than in earlier forecast periods.

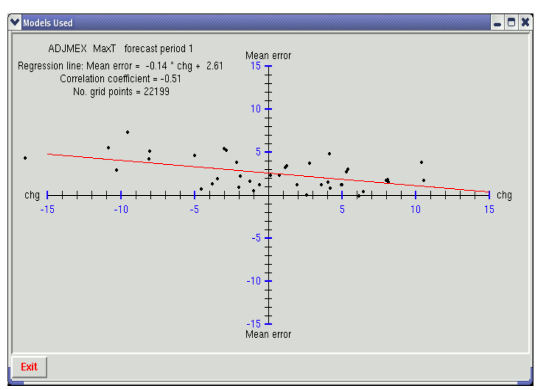


Fig 5. ADJMEX for period 1 MaxT.

Fig 5 shows that the ADJMEX has a warm bias for period 1 MaxT, and also a slight negative slope in the regression line, meaning that it slightly under-forecasts the magnitude of the day-to-day change for both warming and cooling.

Fig 6 shows ADJMEX for period 13 MaxT. Note the larger negative slope of the regression line, which means it more seriously under-forecasts the magnitude of the changes. Note also, that at period 13 the ADJMEX almost never forecasts a change larger than +/- 5 degrees from the day before, but it makes larger errors.

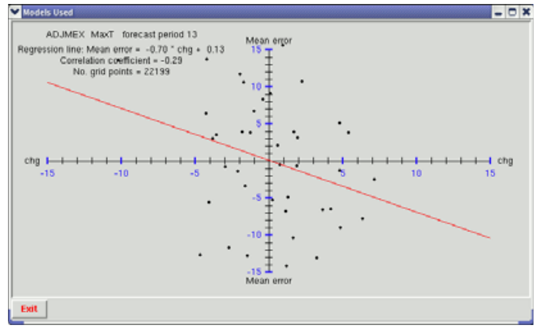


Fig. 6. ADJMEX for period 13 MaxT.

### 3. OVERALL RESULTS

Tables I and II summarize mean forecast error vs day-to-day change for various forecast periods. Each box contains the slope of the regression line of forecast error vs day-to-day change, correlation coefficient, and mean forecast bias.

| MaxT | Pd 1 | Pd 2 | Pd 3 | Pd 4 | Pd 13 | Pd 14 |
|---|---|---|---|---|---|---|
| GFS40 | --- | .00, .03<br>-2.50 | .01,.03<br>-2.92 | .04,.12<br>-2.94 | .19,.20<br>-3.73 | .27,.23<br>-3.39 |
| NAM12 | -.08,-.29<br>-2.01 | -.09,-.29<br>-2.24 | -.09,-.28<br>-3.06 | --- | | |
| ADJMAV | -.12,-.44<br>2.45 | -.13,-.48<br>2.51 | -.12,-.37<br>2.32 | -.20,-.47<br>2.40 | | |
| ADJMET | -.08,-.30<br>-.74 | -.04,-.10<br>-.34 | -.07,-.25<br>-.1.30 | -.12,-.22<br>-1.50 | | |
| ADJMEX | -.16,-.52<br>2.56 | -.18,-.57<br>2.52 | -.12,-.35<br>2.37 | -.17,-.46<br>2.43 | -.64,-.27<br>.27 | -.58,-.26<br>1.01 |

Table I. MaxT regression slope, correlation coefficient, and bias.

| MinT | Pd 1 | Pd 2 | Pd 3 | Pd 4 | Pd 13 | Pd 14 |
|---|---|---|---|---|---|---|
| GFS40 | .02,.04<br>.94 | .04,.14<br>.26 | .05,.09<br>.52 | .13,.28<br>-.07 | .33,.37<br>-.31 | .32,.33<br>-.02 |
| NAM12 | .35,.61<br>5.01 | .29,.61<br>4.03 | --- | .26,.48<br>3.86 | | |
| ADJMAV | -.23,-.55<br>-.82 | -.17,-.39<br>-.97 | -.20,-.41<br>-.95 | -.15,-.26<br>-1.12 | | |
| ADJMET | .01,-.01<br>-1.24 | .02,.01<br>-2.40 | .03,.03<br>-2.04 | -.02,-.12<br>-2.20 | | |
| ADJMEX | -.26,-.53<br>-.76 | -.23,-.44<br>-.83 | -.28,-.47<br>-.92 | -.16,-.26<br>-.99 | -.11,-.09<br>-1.25 | -.42,-.20<br>-2.28 |

Table II. MinT regression slope, correlation coefficient, and bias.

Only the GFS40 and ADJMEX have data for periods 13, 14.

From the tables we observe the following:

1. The GFS40 regression slope increases with time, indicating the GFS40 "senses" more of the change in later periods. The NAM12 maintains relatively constant regression slope with time.

2. The NAM12 has a large negative bias in MinT and a large positive bias in MaxT for all periods.

3. For MaxT the regression slopes of all the MOS-based (i.e., ADJ) forecasts become increasingly negative with time, indicating under-forecasting i.e., reluctance to forecast the magnitude of the day-to-day change in MaxT. This trend is not present, however, for MinT.

4. Correlation coefficients are small whenever regression slopes are small. In the extreme case when regression slope is zero, mean forecast error has no correlation with day-to-day change.

## 4. A FORECASTING SCHEME

BOI_Verify can be used to retrieve past analogs of any given model forecast. For example, it can retrieve all the GFS40 first-period MaxT in the past 50 days. This means a current GFS40 first-period MaxT can be compared pattern-wise with each of them. The results can be sorted and the best analogs can be selected from the top of the sorted list. Now the analogs also have accompanying observed MaxT grids, so they can be verified and their errors can be used to correct the current forecast right away.

For example, the GFS40 expected significant (average 11 F degrees) first-period MaxT warming for May 16 over May 15 in the WFO Boise domain. Fig 7 shows the error histogram made by that GFS40. The 89 score represents 100 minus the mean-squared forecast error. Note the 1.50 degree cold bias, as well as the wide spread of the histogram curve.

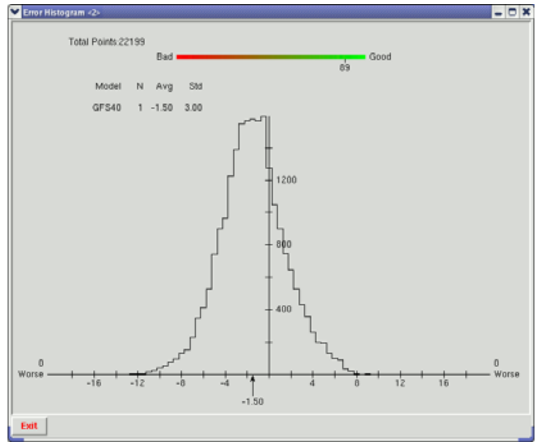


Fig. 7. Error histogram of GFS40 Pd-1 MaxT.

We now access the BOI_Verify archives and retrieve the 12 best analogs to the current GFS40. Verification of the 12 analogs produces 12 error grids, which are then averaged together and applied to the current forecast. All this can be done in real time. Fig 8 shows the error histogram of the modified GFS40 forecast.

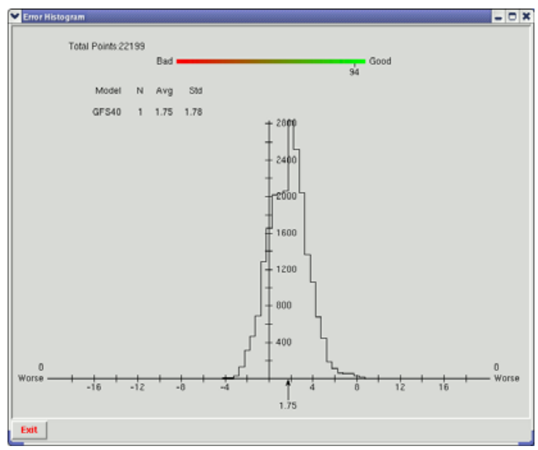


Fig. 8. Error histogram of GFS40 Pd-1 MaxT adjusted by 12 previous analogs.

The 1.75 degree warm bias shows that in this strong warming event the GFS40 is adjusted even warmer by the analogs. Note the 94 score, as well as the narrower curve. This is a much better forecast than the GFS40 forecast of Fig. 7.

## 5. REFERENCES

Barker, T.W., 2006: BOIVerify 1.0. http://www.mdl.nws.noaa.gov/~applications/STR/generalappinfoout.php3?appnum=1089