P1.48

AN ASSESSMENT OF A WEATHER FORECASTING CONTEST IN MULTI-LEVELED METEOROLOGY CLASSES

Joby Hilliker, Department of Geology and Astronomy, West Chester University, West Chester, PA *

1. INTRODUCTION

Meteorology can be divided into two main aspects – science and operations. The scientific aspect, for example, answers the question, "How does it rain?" while the operational component answers the question, "Is it going to rain?" Most introductory meteorology courses, including mine, emphasize the scientific aspect.

However, building a bridge to operational meteorology is instructive since it emphasizes application of the science. Students are enriched when they are able to incorporate their classroom studies in explaining the ambient weather or understanding a weather broadcast. These applications motive me to start each class with a map discussion, where I may show satellite imagery or the jet stream configuration.

I also display computer model output because weather prediction is a significant component of operational meteorology. The inclusion of forecasting in the course reveals to students the tools they are acquiring to construct their own weather predictions and to justifiably agree or disagree with media forecasts.

Several articles have highlighted fellow instructors' incorporation of forecast activities, modules, and contests in meteorology courses. At Iowa State University, Gallus et al. (2000) developed an interactive, Web-based severe weather activity, while Yarger et al. (2000) summarized a forecast activity for large introductory meteorology courses. The latter study was encouraging in that student forecast skill increased with time, and by extension, experience. Similarly, Kahl (2001) showed that student forecast justification of wind speed also increased during the semester.

To complement these studies, I also offer an extra-credit forecast contest similar to the National Collegiate Weather Forecast Contest (NCWFC) -- a competition offered annually to colleges and universities nationwide (Vislocky and Fritsch, 1998).

For the student, the forecasting contest encourages friendly competition, breaks up the class week with something fun, and allows extra credit to be earned. For me, the ulterior motivation is that students can relate to the forecast challenges that operational forecasters confront on a daily basis. By offering this experience, my ultimate objective is that students gain a greater appreciation of weather forecasting.

After the contest's introduction several semesters ago, a sizable data sample now exists to assess the forecasting contest's impact in the classroom. Because I have taught meteorology classes at three different levels -- representing different student populations – a unique opportunity arises to also contrast the contest's impact on student learning with each audience.

Corresponding author's address: Joby L. Hilliker, 223 Boucher Building, West Chester University, West Chester, PA 19382; e-mail: jhilliker@wcupa.edu

2. CONTEST OVERVIEW

The forecasting contest offered at West Chester University is a simplified version of the NCWFC and that described in Kahl et al. (2004). Each Monday, students forecast, via e-mail, the high temperature for the following day at the Philadelphia Airport (PHL), 20 miles southeast of the University. Deadline for forecast submission is 11:59 PM on Monday.

To assist students in their forecasting, I integrate model temperature forecasts into Monday's map discussion, as well as highlight a potential forecast "issue." I emphasize there are a plethora of meteorological factors a forecaster must consider, including advection, frontal timing, and cloud cover. For the latter, I remind students, using radiation arguments, that additional cloud cover during the afternoon may bias the high temperature lower than forecast.

I encourage students, particularly at the beginning of the semester, to consult as many forecasting sources as possible (e.g., National Weather Service, media broadcasts). Although students do not have the necessary skills and/or confidence to construct their own predictions, through this they become exposed to the myriad of weather websites and forecasts available.

On Tuesday evenings, I compile the student forecasts and note the observed Tuesday high temperature at PHL. With these data, I then calculate each student's absolute forecast error, i.e., the number of degrees, either warmer or colder, the forecast departs from the actual (Aguado and Burt, 2004).

Wednesday's class begins with showing the distribution of the class's forecast high temperatures. An example forecast distribution is depicted in Fig. 1.

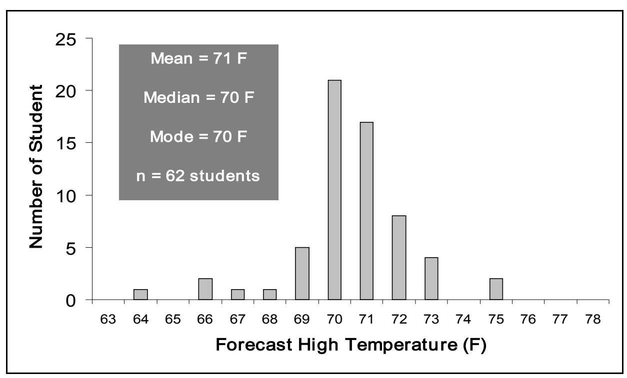

Fig. 1. Example student temperature forecast distribution with statistical measures. The actual high temperature is superimposed.

Here, students can compare their forecast to their peers, as well as to the mean, median, and mode of the class. Showing the distribution each week is pedagogically instructive since it reveals that: a) forecasts are inherently uncertain, and that b) this uncertainty changes on a daily basis. Some distributions are flat (i.e., having high kurtosis), revealing significant forecast uncertainty, while other times there is low kurtosis, indicating greater certainty (Wilks, 1995).

Next, I superimpose the actual PHL high temperature on the forecast distribution (Fig. 1). With that, a table similar to Table 1 is presented showing current student rankings based on cumulative absolute error. Also shown via the table are the students' previous weekly rank, and forecasts with absolute errors for that week. Students who fail to submit a forecast by 11:59 PM Monday receive the maximum absolute error (i.e., the worst forecast) in their class section. Thus, a single missed forecast penalizes the student, but not to an unrecoverable degree.

RANK THIS WEEK	RANK LAST WEEK	NAME	WEEK #10 FORE-CAST	WEEK #10 ERROR	TOTAL ERROR
1	2	Student A	72	0	16
2	4	Student B	70	2	19
3	4	Student C	69	3	20
4	8	Student D	71	1	21
--	--	DR. HILLIKER	71	1	22
5	1	Student E	NA	8	22
6	9	Student F	70	2	23
6	7	Student G	68	4	23
8	12	Student H	73	1	24
8	2	Student I	NA	8	24
10	13	Student J	70	2	26
10	6	Student K	64	8	26

Table 1. Example contest weekly summary depicting student temperature forecasts, errors, and ranking based on cumulative error. "NA" indicates no forecast was submitted by that student.

There are typically 9 to 11 forecast weeks during a semester. At the end of the semester, extra percentage points are added to their final grade based on final contest rankings. For added incentive, I also participate in the contest and grant an extra point if a student beats the professor.

3. DATA COLLECTION

Table 2 shows a summary of course descriptions and sample sizes spanning five semesters (August 2004 to December 2006). Included in the study were 15 sections of an introductory meteorology course designed for undergraduates who need to fulfill the University's science elective requirement. This "Gen. Ed." group consisted of 423 students and comprised the largest of the three student samples. Also taught during the study period were three sections of a meteorology course required for students in the Bachelor's of Science program in geology. A sample of 66, mainly upper classmen, was collected in this "majors" group. Finally, one graduate meteorology class ("grad") was taught, with 16 mostly in-service teachers present.

COURSE NAME / DESCRIPTION	ABBREVI-ATION	# SECTIONS	# STUDENTS
"Our Atmosphere" / open to any student	"Gen. Ed."	15	423
"Introduction to Meteorology" / B. S. Geology students, mainly upperclassmen	"Majors"	3	66
"Principles of Meteorology" / M. A. Geology graduate students	"Grad"	1	16

Table 2. Summary of meteorology course descriptions, abbreviations, and number of sections and students used in study

4. CONTEST PERFORMANCE

a. Relationship to participation

Since the forecast contest is optional, it was useful pedagogically to assess the percentage of students – and type of student -- who strived for extra credit.

Participation percentages revealed two trends. The first was that participation was greatest the first week of the contest and trended downward through the semester. The more revealing trend was that participation increased with course level. Average participation in the contest's final weeks was 40% for Gen. Ed. students, 50% for majors, and is highest (90%) for graduate students.

To gain insight as to why students chose not to participate, I invited students at all levels to reveal their reason(s) for not participating. To elicit their most honest response, I incorporated the question as part of the final exam.

Table 3 is a ranking of the most common responses from those students who participated in the contest less than half the semester weeks. By far, the most common response, regardless of course level, was forgetfulness. This often led to discouragement (ranked second) and eventual discontinuation if the student was not ranking highly in the contest. The third most common reason revolved around logistical issues (e.g., lack of a home computer, sending forecast to incorrect e-mail address) in submitting forecasts.

REASON FOR NOT PARTICIPATING IN CONTEST	GEN. ED.	MAJORS	GRAD
Forgetfulness	75%	79%	0%
Discouragement	31%	7%	0%
No Home Computer / Internet Access	19%	14%	0%
Apathy	6%	0%	0%
Overestimated Extra-Credit Benefit	0%	14%	0%

Table 3. Reasons (ranked by frequency) students cited for not participating in the forecasting contest as a function of course level. Totals exceed 100% since students occasionally offered multiple reasons for not participating.

b. Relationship to course grade

To further explore this relationship between academic ethic and contest performance, scatterplots were constructed correlating each student's final contest rank to his/her final course grade. Any student who withdrew from the class was omitted from the sample.

Figs. 2a-c reveal the plots for the Gen. Ed., majors, and graduate student populations, respectively. Simple linear regression was then applied on each dataset to extract a relationship.

A negative trend is most evident (R^2 = 0.11) in the Gen. Ed. plot. To verify its significance, a hypothesis test ("inference on two population means") was performed to determine if the mean final course grade of the Top 5 Gen. Ed. contest finishers [sample size (n) = 81; mean (μ) = 88.0%; standard deviation (σ) = 8.1%] was in fact higher than the mean final course grade of those Gen. Ed. students who finished positions 16-20 [n = 81; μ = 81.9%; σ = 9.3%]. The obtained test t-value of (88.0% – 81.9%) / 1.37, or 4.45, exceeded the critical t-value (using n = 81; α = 99.5%) of 2.64 (Neter et al., 1996).

Although the trend is not as pronounced for majors (R^2 = 0.06), hypothesis testing on the Top 5 versus Bottom 5 majors' contest finishers (n = 30 for both groups) resulted in a test t-value of (89.1% - 83.4%) / 1.97, or 2.89, again exceeding the critical t-value (n = 30; α = 99.5%) of 2.75.

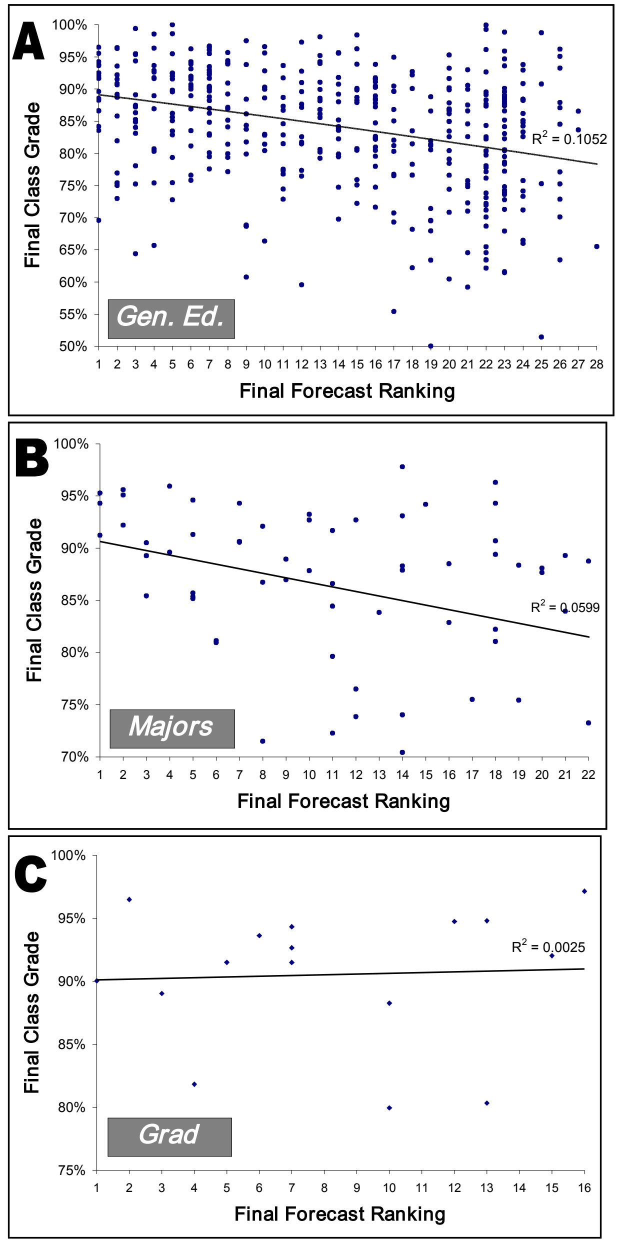

Fig. 2. Scatterplots showing final course grade versus final contest rankings for (a) Gen. Ed., (b) majors, and (c) graduate student samples. A linear trend line and R^2 value are superimposed on each plot.

These results suggest that both Gen. Ed. and majors students who did poorly in class were less likely to finish well in the contest. This is a disappointing inference since it is this type of student who may have an aversion to science and/or not be confident test-takers to whom I target the extra credit.

This conclusion, however, is not valid for graduate students, who had a non-existent relationship (R^2 ~0.00). However, it was encouraging that most graduate students participate in the contest even though extra credit points were not as critical. Their motivation for participating may solely have been to acquire forecasting experience.

5. ASSESSMENT OF LEARNING

Although it is insightful to explore the type of student who did well in the contest, the crux of this study was to assess the impact of the contest on student learning. One objective, quantitative strategy to achieve this objective was determining if student forecasts became more accurate as the semester progressed.

One simple method to accomplish this was to construct a plot of student temperature error versus contest week. Student error was computed using an average of forecast errors from those students finishing the contest in the Top 10% in their respective section.

However, as the solid, black line in Fig. 3 reveals, no meaningful trend could be gleaned since errors also varied widely from forecast to forecast -- largely a reflection of the forecast's complexity on that particular week. This natural variation disadvantageously masked any temporal trend in improvement with which this study was extracting.

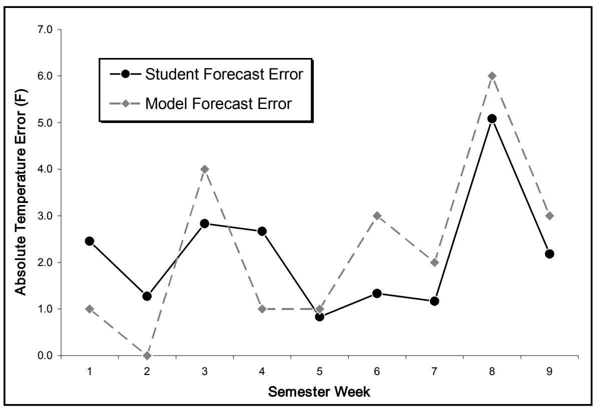

Fig. 3. Absolute temperature error (in ºF) of student (solid, black line) and model forecasts (dashed, gray line) for an example semester.

An alternative strategy was to determine if the best student forecasters were improving with time with respect to some reference (e.g., computer model output). Applying this approach circumvented the aforementioned issue since model performance (dashed, gray line in Fig. 3) also fluctuated significantly from week to week. Therefore, a new parameter -- the *difference* between the model and student error -- was computed. Here, a positive value indicates the student had less error (i.e., a more accurate forecast).

The dashed, gray line in Fig. 4 shows student/model error differences, averaged over all semesters and course levels, as a function of contest week. Linear regression on the average, weighted by the large number of Gen. Ed. sectors, revealed a positive trend (R^2 = 0.36). A hypothesis test verifying the slope's significance resulted in a test t-value of (0.149 / .074), or 2.01. This value exceeded the critical t-value of 1.90 (using 7 degrees of freedom and a 95% confidence level), suggesting that student forecasts in fact did become more accurate with respect to model output as the semester advanced.

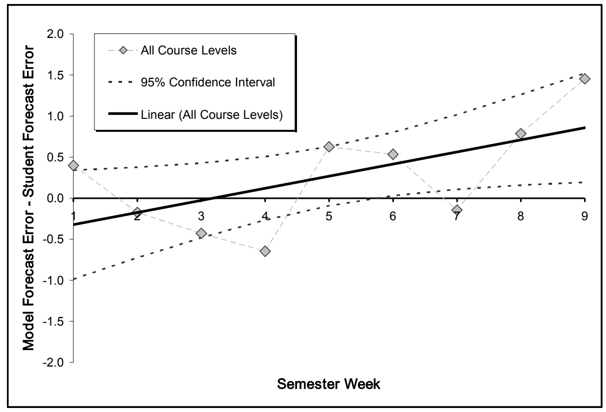

Fig. 4. Student/model forecast error differences (dashed, gray line) as a function of contest week averaged over all courses and all semesters. A linear trend line (bold, black line), with 95% confidence interval bands (dashed, black lines) have been superimposed.

In addition, because the trend line crossed "0" during the semester, student forecasts not only improved but became superior to corresponding computer model forecasts toward semester's end. To support this conclusion, 95% confidence interval bands along the trend line were calculated. As the dashed, black lines reveal in Fig. 4, the 95% confidence interval for the final forecast week was [0.2ºF, 1.5ºF], remaining above zero.

Stratification of the data indicated student forecast improvement was achieved for all course levels. Hypothesis testing revealed the trend was most significant (test t-value = 2.14) for Gen. Ed. students, and nearing the 95% significance level for the graduate class (test t-value = 1.86). An analysis on the more limited majors sample resulted in an anomalously lower (0.74) test t-value.

It is worth emphasizing that, similar to operational forecasters, students did not construct original forecasts, but adjusted model or media guidance. As a result, mean student high temperature errors of ~2ºF were consistent with those of government and private industry forecasters (National Weather Service, 2005; Estupiñan et al., 2006). However, the encouraging and more germane aspects of this study was confirming student forecast improvement, and that students were adjusting model forecasts in a favorable (i.e., more accurate) direction by semester's end.

6. CONCLUSIONS

This study assessed the impact of an optional, extra-credit weather forecasting contest on three different introductory meteorology courses representing three different student populations: general education, majors, and graduate. As such, a unique opportunity existed to contrast the contest's impact as a function of course level.

The following summarizes the main points from the study:

- There were notable differences in contest participation with respect to course level, ranging from ~40% for Gen. Ed. students to 90% for graduate students in the final contest week. The most common reason for student lack of participation was forgetfulness, with discouragement ranked second.
- A statistically significant (at the 99.5% level) trend existed between final course grade and contest ranking at the Gen. Ed. and majors levels, suggesting that student academic ethic motivated contest success. No correlation between course grade and contest ranking was evident at the graduate level – a revealing result given their 90% participation rate.
- A quantitative assessment of learning revealed that students constructed more accurate forecasts as the semester progressed. The improvement was such that, by the end of the semester, their forecasts were statistically superior (at the 95% level) to corresponding computer model forecasts.

Although it was discouraging not all students participated – particularly those for which extra credit would have been of greatest assistance – students who did participate acquired practical experience in weather forecasting. Through both the increased operational experience and knowledge of atmospheric behavior during the course, students constructed increasingly more accurate next-day high temperature forecasts.

Maintaining the contest is time-consuming, and this study provided encouraging feedback on its positive impact in the classroom. It is my goal that other educators who incorporate this contest in their classes will also find it rewarding.

7. REFERENCES

Aguado, E., and Burt, J. E., 2004: *Understanding Weather and Climate.* Prentice-Hall, 474 pp.

Estupiñan, J.G., et al., 2006: Real-time forecast verification tools at The Weather Channel, Atlanta, GA. *Proceedings from the 18th conference on Probability and Statistics*, Atlanta, GA, Amer. Meteor. Soc., 7 p.

Gallus, Jr., W. A., Yarger, D. N., and Herzmann, D. E., 2000, An interactive severe weather activity to motivate student learning, *Bull. Amer. Meteor. Soc.*, **81**, 2205-2212.

Kahl, J. D. W., 2001, Meteorology online: Weather forecasting using the Internet. The Science Teacher, **68**, 22-25.

Kahl, J., Horwitz, K., Berg, C., and Gruhl, M., 2004, The quest for the perfect weather forecaster, Science Scope, **27**, 24-27.

National Weather Service, 2005, Shareholders' Report 2004, *http://www.srh.weather.gov/ffc/html/anreport04.pdf*

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W., 1996: *Applied Linear Statistical Models*. Fourth edition. McGraw-Hill, 1408 pp.

Vislocky, R. L., and Fritsch, J. M., 1998, Performance of an advanced MOS system in the 1996–97 National Collegiate Weather Forecasting Contest, *Bull. Amer. Meteor. Soc.*, **78**, 2851–2857.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. First edition. Academic Press, 464 pp.

Yarger, D. N., Gallus, Jr., W. A., Taber, M., Boysen, J. P., and Castleberry, P., 2000, A forecasting activity for a large introductory meteorology course, *Bull. Amer. Meteor. Soc.*, **8**1, 31-39.