THE USE OF REGRESSION DATA IN WEATHER PREDICTION: AN EXERCISE FOR HIGH SCHOOL INTERNS IN A COLLEGE SUMMER ENVIRONMENTAL WORKSHOP AT A MINORITY-SERVING INSTITUTION (MSI)

Jeremy R. Montague, Ph.D.

SNHS-Biology 305, School of Natural and Health Sciences, Barry University, Miami Shores, FL 33161

305-899-3218 • *jmontague@mail.barry.edu* • *http://snhs-jmontague.barry.edu*

Abstract

Barry University, a Catholic, coeducational, Minority-Serving Institution in South Florida, developed in 2004 a summer environmental workshop course (BIO 295) that enrolled high school students traditionally underrepresented in the NOAA-related sciences. Ten participants were selected from Dade County high schools for this first-time offering. These students earned two credit hours of college lab science by completing a variety of biological, chemical and physical lab exercises over a three week period (14 June – 1 July). These included collection of field data on incident solar radiation and ground temperature. These data were used in regression analyses similar to those developed for simple hurricane models. The students learned theoretical approaches to weather forecasting, and methods in data measurement, website data downloads, spreadsheet calculations, and Microsoft PowerPoint" file designs. The students made formal presentations to the faculty. The post-program assessment data indicated that the students responded positively to the exercise.

"In a very crude sense, [weather] forecasting simply means calling up someone to the west of you and asking them to look out their window." (Junger, 1997, p. 99).

If only things were that simple. Forecasting events such as stock market fluctuations, elections results, sports championships, or (more dreadfully) land-falling hurricanes, has never been any sort of exact science. Yet the costs of a bad prediction can be high. I write these very words today with apprehension, as Hurricane Frances moves just to our north and Hurricane Ivan approaches from the southeast.

Hurricane modeling began in earnest during the 1960's (Saffir, 1973). The techniques have become increasingly powerful and the forecasts increasingly accurate. But surprises do occur, and mistaken forecasts can lead to catastrophic damage and loss of life for those areas left unprepared (Powell and Houston, 1996; Powell, et al., 1996; Rappaport, 1994; Vickery, et al., 2000; Willoughby and Black, 1996).

Hurricane modeling remains in its infancy, particularly in terms of our understanding of causes and effects of broader global influences such as greenhouse warming or El nino weather events (Landsea and Knaff, 2000).

Further complicating matters, some experts believe we may be entering a decadal period of increased tropical storm activity during which hurricanes may become not only more frequent, but more powerful and more unpredictable as well (Goldenberg, et al., 2001).

Can we minimize the errors? That depends partly on the students we train today. How best to prepare the next generation of hurricane forecasters? This is a concern for educators in the NOAA-related sciences, particularly at colleges and universities that enroll large numbers of students traditionally left out of disciplines such as oceanography or meteorology. We need to tap the largest available talent pools. Here I report on one particular exercise developed for a three-week summer college workshop for high school students interested in the NOAA-related sciences. The exercise involved measurement of ground surface temperature and incident solar energy (measured in lux units); these data were then incorporated into a simple linear regression using temperature and incident radiation as the dependent and independent variables.

Training New Forecasters

Barry University is a Catholic, coeducational, and Minority-serving institution in South Florida. Of the undergraduate biology majors in the School of Natural and Health Sciences, roughly 75% are female, and roughly 54% are minority (mainly Hispanic, Caribbean and African-American); these are students who have been traditionally underrepresented in the NOAA-related sciences.

To better serve our students we designed our BIO 295 Summer Environmental Sciences Workshop, and we recruited ten talented teenagers from local high schools to fill our inaugural class in 2004. We focused on developing hands-on lab and field projects covering topics in biodiversity, ecosystem restoration, technological applications in sciences, enhancement of computer skills, etc. This design attempts to integrate computer-based exercises (so-called "virtual labs:" Carnevale, 2003) into a traditional lab science curriculum, and it addresses recent concerns about the levels of science and math skills for entering college students (Anonymous, 2004).

Hurricane Models

Atlantic hurricane forecasting is mainly the responsibility for the Tropical Prediction Center of the National Hurricane Center (NHC). NHC today relies on nearly twenty different operational and experimental computer models, each using hundreds of complicated mathematical equations requiring enormous memory storage and very high calculating speeds (DeMaria, 1997; Elsner, 2003; Elsner and Jagger, 2003, Weber, 2003; Zorita and von Storch, 1999). Some of the models, e.g., CLIPER (the CLImatology and PERsistence model), and GFDL (from Princeton University's General Fluid Dynamics Lab) have been refined and upgraded over many years. Others, e.g., UKMET (from U.K.'s Meteorological Office) or FSU (Florida State University's super-ensemble) are much more recent packages.

Each model requires numerical input in the form of oceanographic and meteorological variables, such as sea-surface temperature (SST), wind velocity, degrees latitude/longitude, etc.

There are two basic types of hurricane models used by NHC: (1) statistical (those that predict a current storm's future behavior based upon historical data from previous storms) and (2) dynamical (those that use current global atmospheric and oceanographic conditions to forecast the behavior of the current storm). In practice, forecasters use these in combinations. This paper focuses mainly on the statistical models that rely on regression equations (for example, see Aberson, 1998).

Regression Using Historical Data

Suppose we want to predict the sea-surface temperature (SST) for a particular storm 24 hours ahead of time. We begin by selecting a random sampling of past storm data, say, SST at 11 a.m. on day 5 during ten different storm histories (storm-A, storm-B, etc.). Each SST is then paired with the SST 24 hours later for the same storm (Figure 1).

A simple linear regression equation may be generated in the form of T°C at t0+24 = slope*(T°C at t0) + y-intercept. The scatter of data points creates a predicted range of statistical error.

Such a hurricane model, then, incorporates many thousands of such regression calculations per model run.

In general, a model is initialized with current (state) parameters at the start of a six or twelve-hour run, and its results are compared later with the updated storm condi-

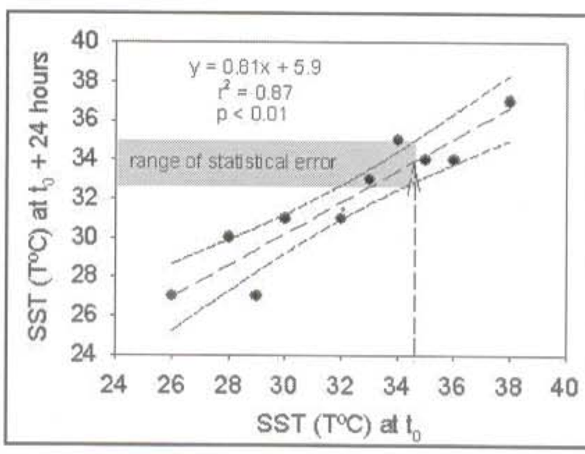

Figure 1. Linear regression (with 95%-CIE boundaries) using hypothetical SST data (n = 10 randomly selected storm days).

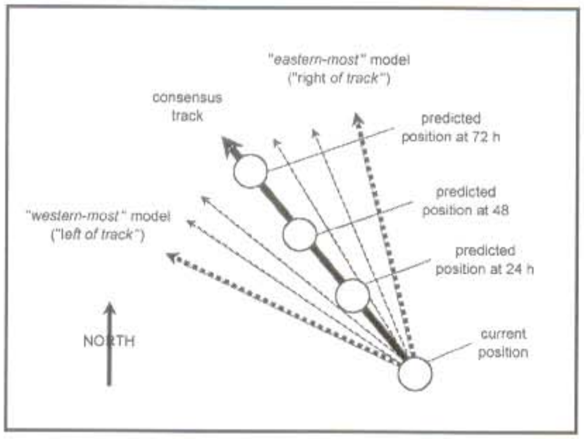

Figure 2. Hypothetical example of a consensus storm track.

tions. The future conditions are then predicted by forecasters who use their experience and intuition to choose among the better-performing models from run to run.

The result is a consensus track based on a compromise solution among all the competing models (Figure 2). Current state-of-the-art modeling produces 24-hour, 48-hour and 72-hour predicted positions with up to hundreds of miles of potential error (in a so-called "increasing cone-width") at the end of the forecast period. NHC seeks to minimize the cone-width of uncertainty.

The hurricane models may also be subdivided into two functional types: (1) track guidance models that forecast direction of movement, and (2) intensity models that forecast strength and wind velocities. In theory, each model predicts both track and intensity, but in practice, the track models tend to be poorer at predicting intensity, and vice versa. No one model has an outstanding performance record; each has its limitations. A model's success varies season to season. In addition, models performing well early in a storm's history have been known to become quirky and unpredictable later in the same storm, behaving erratically over short periods of time (Weber, 2003).

To assess the output of any one model requires an understanding of the statistical nature of data sampling and its use of regression equations. It is this understanding we'd like to pass onto our students.

Workshop Exercise

Rather than focusing on prediction of future values of a single variable (as in our previous SST example), we designed a regression exercise that would explore the statistical relationship between two variables. This approach allowed us to retain the advantages of correlation and regression while giving the students an opportunity to examine the interaction of actual physical and meteorological variables.

To stimulate student curiosity about the relationship between thermal heat absorbed by the ocean surface and subsequent change in SST, we devised an outdoor exercise using a thermometer and a hand-held digital photometer.

We assigned pairs of students to three 20-m linear transects per team stretched across our main campus quadrangle. Each transect included mainly grassy surface with an occasional concrete sidewalk surface; we did not quantify the percent of transect length intersected by concrete. At every 5-m intercept, the students recorded ground temperature to the nearest degree Celsius, and incident solar radiation in relative lux units using a digital photometer (Model DLM2, J.J. Electronics). We made no allowances or adjustments for cloud cover, and so we collected a variety of points in both shade and direct sunlight (Figure 3; n = 90 x,y data points recorded).

The students downloaded the data from the course website and used simple linear regression to calculate a predicted temperature from any given lux value (Figure 4).

Figure 3. Collecting the field data.

Note that the regression analysis revealed a non-significant relation between the variables ($p = 0.16$), an important lesson for our students. Our data showed that incident solar radiation would not be a reliable predictor for ground temperature. Given the theoretical expectation that increased solar flux would lead to increased ground temperature, our results seem at first glance surprising. But the students soon recognized the problems in not accounting for variability in surface textures (grass vs. concrete) or in cloud cover. Thus, our results focused their attention on re-designing the methods.

The Presentations

Each student was assigned to describe and present the results of an exercise in an oral format to the SNHS faculty on the last day of the program (to be accompanied by Microsoft PowerPoint" slides of their own designs). We spent two weeks on training the students with spreadsheets and PowerPoint, techniques, and they relished an opportunity to get up and perform. Each chose his or her own color schemes, graphic designs, etc. This was a process they seemed to really enjoy.

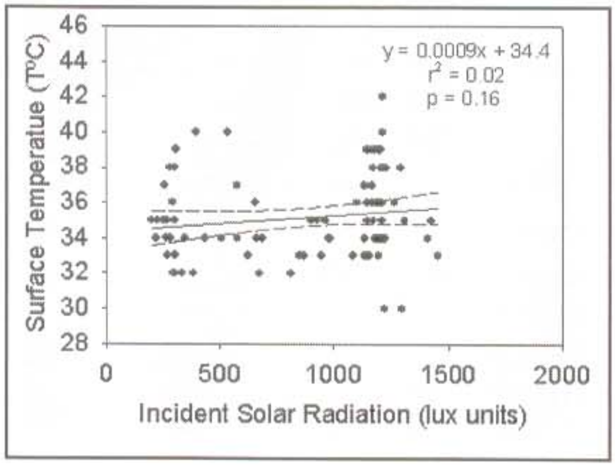

Figure 4. Regression (with 95%-CIE boundaries) collected from 20-m transects (n = 90).

The Post-Workshop Assessment

The students completed an anonymous, post-workshop assessment instrument concerning their likes, dislikes, and expectations for their summer experience; this was included in our Six-Month Period Progress Report to NOAA (July 2004). Our major findings were that the students were very enthusiastic about the introduction to computer software packages such as Microsoft Excel, and PowerPoint,,. Though some of the students had some experience with the packages, all reported learning many new techniques they planned to use in their future studies.

During informal discussions, several students reported being pleasantly surprised with the basics of statistical inference and regression modeling, which I believe will lead them into detailed explorations in this field.

Perhaps 20 years from now, we'll discover that one of the great hurricane modelers of the 21st-Century had actually sat in our classroom this past summer. I certainly hope so.

Conclusions

The NOAA/EPP Award gave us the opportunity to recruit and train a hitherto underrepresented cohort of young high school students. The summer workshop gave the students a taste of the college experience, as well as two hours of college credit. The particular exercise using regression statistics and predictive modeling allowed them to consider the career possibilities and opportunities in the NOAA-related sciences.

ACKNOWLEDGMENTS

Significant assistance on this project came from Sister John Karen Frei, Ph.D., Gerhild Packert, Ph.D., Silvia Macia, Ph.D., David Molnar, Ph.D., and Linda Johnson, B.S. This project was funded by NOAA Award No: NA03OAR4810129 (2003-2006), Barry University EPP/MSI Environmental Entrepreneurship Program.

CITED LITERATURE

Aberson, S.D., 1998: Five-day tropical cyclone track forecasts in the North Atlantic basin. *Wea. Forecast.*, **13**, 1005-1015.

Anonymous, 2004: Average national ACT score rises for first time since 1997, but many students still not ready for college science and math courses. Online National Data Release from: http://www.act.org/news/releases/2004/8-18-04.html (viewed 8-18-2004).

Carnevale, D., 2003: The virtual lab experiment. *Chronicle Higher Ed.*, **49**, 30-32.

DeMaria, M., 1997: Summary of the NHC/TPC tropical cyclone track and intensity guidance models. *http://www.nhc.noaa.gov/aboutmodels.shtml* (viewed 9-1-2004).

Elsner, J.B., 2003: Tracking hurricanes. *Bull. Amer. Meteor. Soc.*, **84**, 353-356.

Elsner, J.B. and T.H. Jagger, 2004. A hierarchical Bayesian approach to seasonal hurricane modeling. *J. Climate*, **14**, 2813-2827.

Goldenberg, S.B., C.W. Landsea, A.M. Mestas-Nuñez, and W.M. Gray, 2001: The recent increase in Atlantic hurricane activity: causes and implications. *Science*, **293**, 474-479.

Junger, S., 1997: *The perfect storm.* HarperPerennial, NY (236 pages).

Landsea, C.W., and J.A. Knaff, 2000: How much skill was there in forecasting the very strong 1997-98 El Niño? *Bull. Amer. Meteor. Soc.*, **81**, 2107-2120.

Powell, M.D., and S.H. Houston, 1996: Hurricane Andrew's landfall in South Florida. Part II: Surface wind fields and potential real-time applications. *Wea. Forecast.*, **11**, 329-349.

Powell, M.D., S.H. Houston, and T.A. Reinhold, 1996: Hurricane Andrew's landfall in South Florida. Part I: Standardizing measurements for documentation of surface wind fields. *Wea. Forecast.*, **11**, 304 - 328.

Rappaport, E.N., 1994: Hurricane Andrew. *Weather*, **49**, 51-61,

Saffir, H.S., 1973: Hurricane wind and storm surge. *The Military Engineer*, **423**, 4-5.

Vickery, P.J., P.F. Skerlj, and L.A. Twisdale, 2000: Simulation of hurricane risk in the U. S. using empirical track model technique. *J. Struct. Engineering*, **126**, 1222-1237.

Weber, H.C., 2003: Hurricane track prediction using a statistical ensemble of numerical models. Month. *Wea. Rev.* **131**, 749-770.

Willoughby, H.E. and P.G. Black, 1996: Hurricane Andrew in Florida: dynamics of a disaster. *Bull. Amer. Meteror. Soc.*, **77**, 543-549.

Zorita, E., H. von Storch, 1999: The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *J. Climate*, **12**, 2474-2489.