

PREDICTIVE MODELS FOR TIME TO ACCEPTANCE

An Example Using "Hurricane" Articles in AMS Journals

BY ROBERT E. HODGES, JAMES B. ELSNER, AND THOMAS H. JAGGER

A hurricane article submitted by mid-September has a 50–50 chance of being accepted in time to meet the IPCC manuscript deadline of mid-March 2013.

Publication is central to science. Authors, not wanting to get scooped, are keen about the speed of the review process. Editors, focused on ensuring a thorough review, are aware that timeliness is important to a journal's reputation. It can also be a practical matter. For example, there is usually a cutoff date after which a manuscript will not be considered by an assessment report.

Here, we assume that information about manuscript review time might be useful to authors and editors, especially if it can say something about future published articles. This motivates us to collect data on publication times from recent journals and to model them using a Bayesian approach. A Bayesian model has the advantage that output is in the form of a predictive distribution (Elsner and Bossak 2001).

The purpose is to provide one example using data relevant to our area of expertise. Here, we use the American Meteorological Society (AMS) journals with articles published from January 2008 through December 2010 with the keyword "hurricane" appearing in published article titles. The search is done from the AMS journals website (<http://journals.ametsoc.org/action/doSearch>). Selecting "full text" on

a particular article brings up the abstract, keywords, and two dates—received and accepted—in month, day, and year format (as available).

We manually enter both dates into a spreadsheet along with the lead author's last name and the name of the journal. We find 133 articles with the word "hurricane" in the title having both received and accepted dates over the three-year span. Of these, 32, 34, 41, and 26 had accepted years of 2007, 2008, 2009, and 2010, respectively.

Journals publishing these articles are shown in Fig. 1. Ten different journals are represented. The four

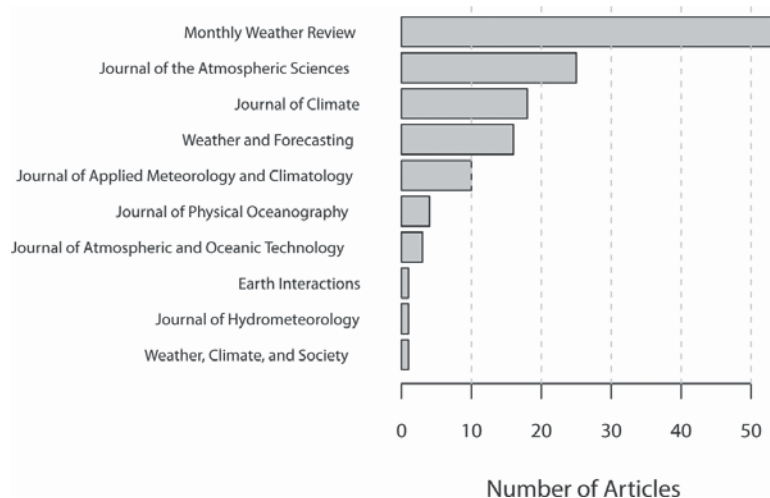


FIG. 1. AMS journals publishing articles with "hurricane" in the title.

journals with the most articles are *Monthly Weather Review* (54), *Journal of the Atmospheric Sciences* (25), *Journal of Climate* (18), and *Weather and Forecasting* (16). We compute the time period (in days) between received and accepted dates. We call this period the time to acceptance (τ), and this is the statistic of interest. By changing the keywords in the search, the data can reflect more (or less) specific topics. Here, we choose the keyword “hurricane” simply as an example and we do not consider rejected manuscripts.

Undoubtedly, there are many factors that influence the value of τ . On the editor’s side, there is prereview, review requests, and distributing the articles for review among others. On the reviewer’s side, there is workload as well as breaks for travel and vacation. On the author’s side, there is the effort needed to revise and respond to the critique. For instance, τ can be impacted by a change in editor or by improvements to editorial review processes (Jorgensen 2009). For authors, τ can be impacted by the number and length of revisions needed. For example, close to half of all submitted manuscripts to *Monthly Weather Review* require major revisions, requiring at least one additional round of reviews (Schultz 2010a). The time the manuscript spends with the author thus has considerable influence on τ during the review process.

The mean τ for all accepted manuscripts is 198 days. The mean τ is 250 in 2007, 195 in 2008, 168 in 2009, and 188 in 2010. The mean τ for each of the four journals with the most articles is 180 (*Monthly Weather Review*), 169 (*Journal of the Atmospheric Sciences*), 228 (*Journal of Climate*), and 227 (*Weather and Forecasting*). Per journal average, *Journal of Climate* is slowest and *Journal of the Atmospheric Sciences* is fastest; their difference is just less than 8 weeks. Distribution of τ in 60-day intervals is shown as a stacked bar chart in Fig. 2.

Here, the goal is a predictive distribution for τ . This will allow us to make inferences about the

time to acceptance for future manuscript published articles. We assume τ is a random variable having a gamma density given by

$$f(\tau | \alpha, \beta) = \frac{\tau^{\alpha-1} \exp\left(\frac{-\tau}{\beta}\right)}{\beta^\alpha \Gamma(\alpha)}, \quad (1)$$

where α and β are the shape and scale parameters, respectively, and $\Gamma(\alpha) = (\alpha - 1)!$. The gamma density is commonly used to model time periods (wait times, phone call lengths, etc). Now, if we place a uniform prior distribution on the parameter vector (α, β) , the posterior density is given by

$$g(\alpha, \beta | \tau) \propto g(\alpha, \beta) f(\tau | \alpha, \beta). \quad (2)$$

The uniform prior is consistent with a judgment that the parameters are the same regardless of author or journal. Random draws from this joint posterior density are summarized and also used to draw predictive samples for τ from a gamma density.

Following Albert (2009), the posterior density is reformulated in terms of $\log \alpha$ and $\log \mu$, where $\mu = \alpha \times \beta$ is the posterior mean of τ . Making use of the open-source statistical software platform **R**, we use the `gamma.sampling.post` function from Albert (2009) to compute the posterior of the reformulated parameters.

Given a pair of parameter values $(\log \alpha, \log \mu)$, the function computes the posterior probability given the data and the posterior density using

$$\log g\left(\alpha, \frac{\mu}{\alpha}\right) = \log \mu + \sum_{i=1}^n \log f\left(\tau_i | \alpha, \frac{\mu}{\alpha}\right). \quad (3)$$

It performs this computation using parameter pairs defined as a two-dimensional grid spanning the domain of the posterior. Alternatively, we could use a Markov chain Monte Carlo approach to compute the posterior density, which would allow us to use nonuniform priors for the parameters.

The average μ over the posterior draws is the mean time to acceptance, which is just less than 200 days. This might, by itself, be useful to an editor. However, the posterior gives additional information. For example, the editor can use the model to estimate the probability that the average time to acceptance will exceed 210 days (7 months) for the set of manuscripts arriving next month. Since the model provides random draws, the question is answered by finding the percentage of draws exceeding the natural loga-

AFFILIATIONS: HODGES, ELSNER, AND JAGGER—Florida State University, Tallahassee, Florida

CORRESPONDING AUTHOR: Robert E. Hodges, Department of Geography, Florida State University, Tallahassee, FL 32306
E-mail: reh3682@fsu.edu

The abstract for this article can be found in this issue, following the table of contents.

DOI:10.1175/BAMS-D-11-00133.1

In final form 20 December 2011

©2012 American Meteorological Society

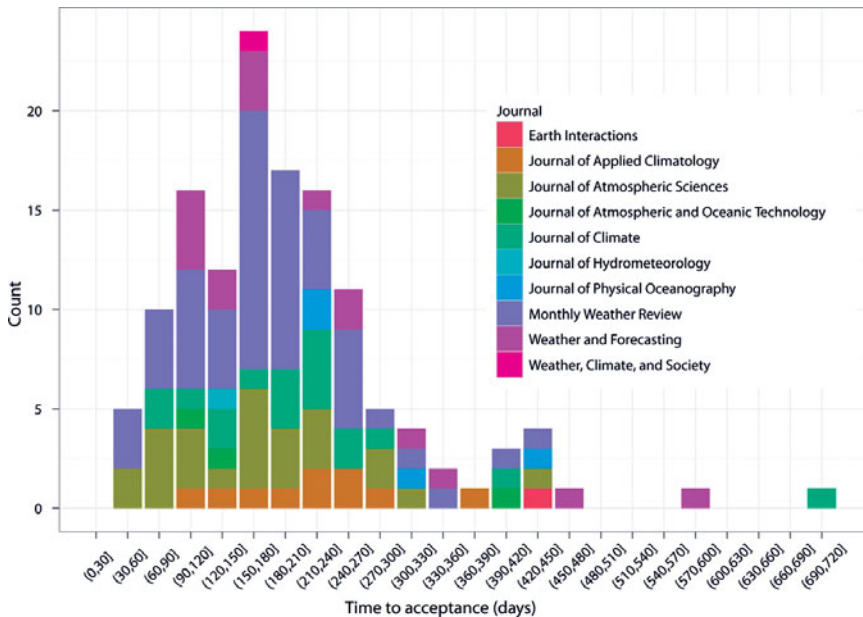


FIG. 2. Time to acceptance for AMS journals publishing articles with “hurricane” in the title.

rithm of 210. Assuming the set of manuscripts is a random sample, the model predicts 9.4%.

Averages are not as useful to an author. An author might like to know if his recently submitted manuscript, which he assumes will be accepted, will take less than 120 days (4 months) to do so. To answer this question, the author takes random draws of τ 's from a gamma density using the random draws of the parameters from the posterior distribution. This is done with the gamma function and then finding the percentage of the draws less than this many days. Here, the model predicts a probability of 21.4%. This probability is higher than the average percentage less than 120 days because it includes additional uncertainty associated with modeling an individual (his) estimate rather than a parameter estimate (the mean).

Model fit is checked by examining quantile statistics from the data against the same statistics from the posterior draws. For instance, the percentage of articles in the data with τ less than 90 days is 10.5, which compares with a percentage of 11.9 from the posterior draws. Continuing, the percentage of articles with τ longer than 360 days from the data

is 8.3, which compares with a percentage of 5.8 from the posterior draws. These differences are relatively small, indicating the model fits the data quite well.

The model is useful for authors wishing to meet a deadline such as tenure review. Or another example: For research to be considered by the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, authors must have their relevant manuscript accepted for publication by 15 March 2013.

Figure 3 shows the model's predictive probability of meeting this deadline versus submission

date. The probability is high in early 2012 when the deadline is still a year away. However, by the middle of September the probability drops below 50%, and by mid-January 2013 the probability falls below 10%.

Here, we demonstrate a statistical model that is useful to authors and editors for making decisions associated with research manuscripts. The mod-

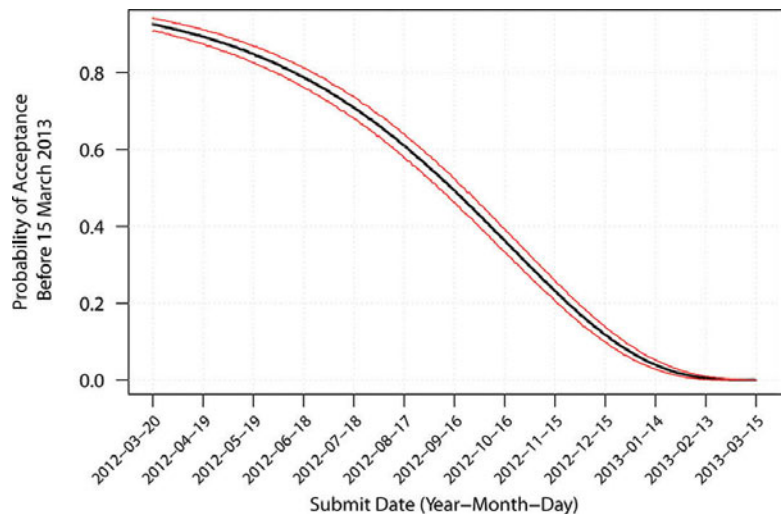


FIG. 3. Probability of acceptance as a function of submit date given the model and the data. The probability is the percentage of posterior τ draws less than τ_0 days from 15 Mar 2013. The results are based on 1,000 random draws from a gamma distribution where the scale and shape parameter are derived from the same number of draws from the joint posterior given the data. The 95% credible interval is obtained by repeating the 1,000 draws from the gamma distribution 1,000 times and taking the corresponding quantile values.

eled example is based on data collected from the AMS online publication website and R code with special functions from the LearnBayes package. The model is silent about the probability of acceptance. Schultz (2010b) finds that 138 of 409 manuscripts submitted to *Monthly Weather Review* were rejected (2006–07). Assuming the paper gets accepted, the model’s predictive probabilities reflect what can be expected if an arbitrary author submits to an arbitrary AMS journal with “hurricane” in the title. The data and code are available with the lead author upon request.

The methodology can be applied to data obtained using different search criteria. The less specific the criteria (e.g., “hurricane” or “tropical storm”), the smaller the variance (larger sample size) on answers to inferential questions but the larger the bias on those answers relative to specific interests. The more specific the search criteria (e.g., “hurricane”), the larger the variance but the smaller the bias. Thus, depending on search criteria, the methodology can produce a model for making time-to-acceptance inferences across a spectrum of topics, but it can also produce a model for making inferences for a single, more specific topic.

Nevertheless, the time it takes for a particular manuscript to be accepted might depend on process-specific to the journal and author. Certain journals and authors could have a faster or slower turnaround time. From our sample, we note that the two papers in the data with J. Kossin as lead author have τ ’s of 84 and 56 days, values that are substantially smaller than the mean. In this case, a hierarchical Bayesian model (Elsner and Jagger 2004) can be built to accommodate such differences. It would be necessary to use a Markov chain Monte Carlo procedure (e.g., Metropolis–Hasting algorithm, Gibbs sampling) to obtain the posterior probabilities (Jackman 2009, 201–202).

Finally, we note that changes to review rules, editorial staffing, and manuscript timetables and tracking will influence time to acceptance. To the extent that these changes occur during the period of data collection or subsequently, they will influence the model’s ability to accurately anticipate time to acceptance.

ACKNOWLEDGMENTS. We thank David M. Schultz for his thoughtful review of our initial manuscript. The work was supported with a contract from the Strategic Environmental Research and Development Program (SERDP SI-1700). All statistical analyses were performed using the software environment R (www.r-project.org).

REFERENCES

- Albert, J., 2009: *Bayesian Computation with R*. 2nd ed. Springer, 84 pp.
- Elsner, J., and B. Bossak, 2001: Bayesian analysis of U.S. hurricane climate. *J. Climate*, **14**, 4341–4350.
- , and T. Jagger, 2004: A hierarchical Bayesian approach to seasonal hurricane modeling. *J. Climate*, **17**, 2813–2827.
- Jackman, S., 2009: *Bayesian Analysis for the Social Sciences*. Series in Probability and Statistics, Wiley, 598 pp.
- Jorgensen, D., 2009: Editorial. *Mon. Wea. Rev.*, **137**, 2413.
- Schultz, D., 2010a: Are three heads better than two? How the number of reviewers and editor behavior affect the rejection rate. *Scientometrics*, **84**, 277–292.
- , 2010b: Rejection rates for journals publishing in the atmospheric sciences. *Bull. Amer. Meteor. Soc.*, **91**, 231–243.

A large graphic featuring a copyright symbol (©) followed by the word "copy" in a lowercase, sans-serif font.

What you're reading is more than just copy. It's also copyrighted. So before you head over to the photocopier, make sure you have permission. Contact the publisher or visit www.copyright.com.

 COPYRIGHT CLEARANCE CENTER